THE ROYAL SOCIETY

doi 10.1098/rspb.2000.1306

Common structure in panels of short ecological time-series

Qiwei Yao[1], Howell Tong[1,2], Bärbel Finkenstädt[3] and Nils Chr. Stenseth[4*]

[1]*Department of Statistics, London School of Economics, London WC2A 2AE, UK*
[2]*Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam, Hong Kong*
[3]*Department of Statistics, University of Warwick, Coventry CV4 7AL, UK*
[4]*Division of Zoology, Department of Biology, University of Oslo, N-0316 Oslo, Norway*

Typically, in many studies in ecology, epidemiology, biomedicine and others, we are confronted with panels of short time-series of which we are interested in obtaining a biologically meaningful grouping. Here, we propose a bootstrap approach to test whether the regression functions or the variances of the error terms in a family of stochastic regression models are the same. Our general setting includes panels of time-series models as a special case. We rigorously justify the use of the test by investigating its asymptotic properties, both theoretically and through simulations. The latter confirm that for finite sample size, bootstrap provides a better approximation than classical asymptotic theory. We then apply the proposed tests to the mink–muskrat data across 81 trapping regions in Canada. Ecologically interpretable groupings are obtained, which serve as a necessary first step before a fuller biological and statistical analysis of the food chain interaction.

Keywords: bootstrap; Canadian mink–muskrat data; nonlinear time-series; predator–prey interactions; similarity measure; threshold modelling

1. INTRODUCTION

Following Elton's (1924) pioneering work, one of the key issues in ecology has been to understand the mechanisms underlying the periodic population fluctuations of northern regions. For terrestrial vertebrates such as microtine rodents and the Canadian snowshoe hare, attention has recently been focused on the food chain interaction (e.g. plant–herbivore and predator–prey interactions) as a possible explanatory mechanism (see Hanski *et al.* 1993; Krebs *et al.* 1995; Stenseth *et al.* 1996*a*, 1997, 1998*a*). Such food chain interactions are typically nonlinear (May 1981, 1986), which may reflect the so-called phase dependence due to, among other things, the different hunting (or escaping) behaviour of the predator (or the prey) at different stages of the population cycle (Framstad *et al.* 1997; Stenseth *et al.* 1998*b*). It is therefore an important question whether the food chain interactions may be grouped according to exogenous factors such as habitat (see Stenseth *et al.* 1999). Often we have only partial information on the food chain interactions. For example, there is a general lack of data on both predator and prey from the same area and over the same time-period. As a result, many statistical analyses of food chain dynamics rely on formulating models in delay coordinates based on either the predator or the prey data only (see, for example, Stenseth 1999). The time-series modelling of Canadian lynx data is a typical case in point (see, for example, Tong 1990, § 7.2).

Here we study the annual numbers of muskrats (*Ondatra zibethicus*) and minks (*Mustela vison*) caught over 81 trapping regions in Canada for a period of 25 years (see Erb *et al.* 2000). The data are extracted from the records compiled by the Hudson Bay Company on fur sales at auction in the period 1925–1949. Such data are typically referred to as panel time-series since we have parallel series during the same time-period over 81 posts (see, for example, Baltagi 1995). Any approach to modelling the mink–muskrat time-series inevitably faces the difficulty of having only 25 points *vis-à-vis* a cycle length of, for example, around ten years. To perform a more powerful analysis based on a larger sample size, the first important step is to pool the data from those regions that share a common or similar structure. In fact, this is a typical problem within the field of ecology because population abundance data are often available only for a relatively short time-span. However, sometimes, several such time-series of data referring to more or less the same ecological process do exist. Another similar example is the Hokkaido grey-sided vole data analysed by Hjellvik & Tjøstheim (1999) (see also Stenseth *et al.* 1996*b*; Bjørnstad *et al.* 1996, 1998; Lindström *et al.* 1998). To perform a more powerful analysis based on a larger sample size, it is essential to pool the time-series data from those regions that share a similar dynamic structure.

The mink and muskrat data constitute an exceptional set of field data, in that we have information on both prey (the muskrat) and its key predator (the mink) (see Errington 1961, 1963). These data sets therefore offer a unique opportunity to carry out systematic statistical analyses aimed at a deeper understanding of the ecological interaction from a quantitative point of view. Figure 1*a* shows the locations of the 81 posts, most of which are located in the so-called boreal forest. Figure 1*b* depicts the time-series plots of the mink and muskrat data from eight randomly selected posts. Most series exhibit cycles with a period of around ten years (see, for example, Erb *et al.* 2000). There exists a clear synchrony between the fluctuations of the two species with a delay of about one or two years. The ecological interaction between the two species may differ from one region to another, not least in

*Author for correspondence (n.c.stenseth@bio.uio.no).

Received 12 June 2000 *Accepted* 10 August 2000

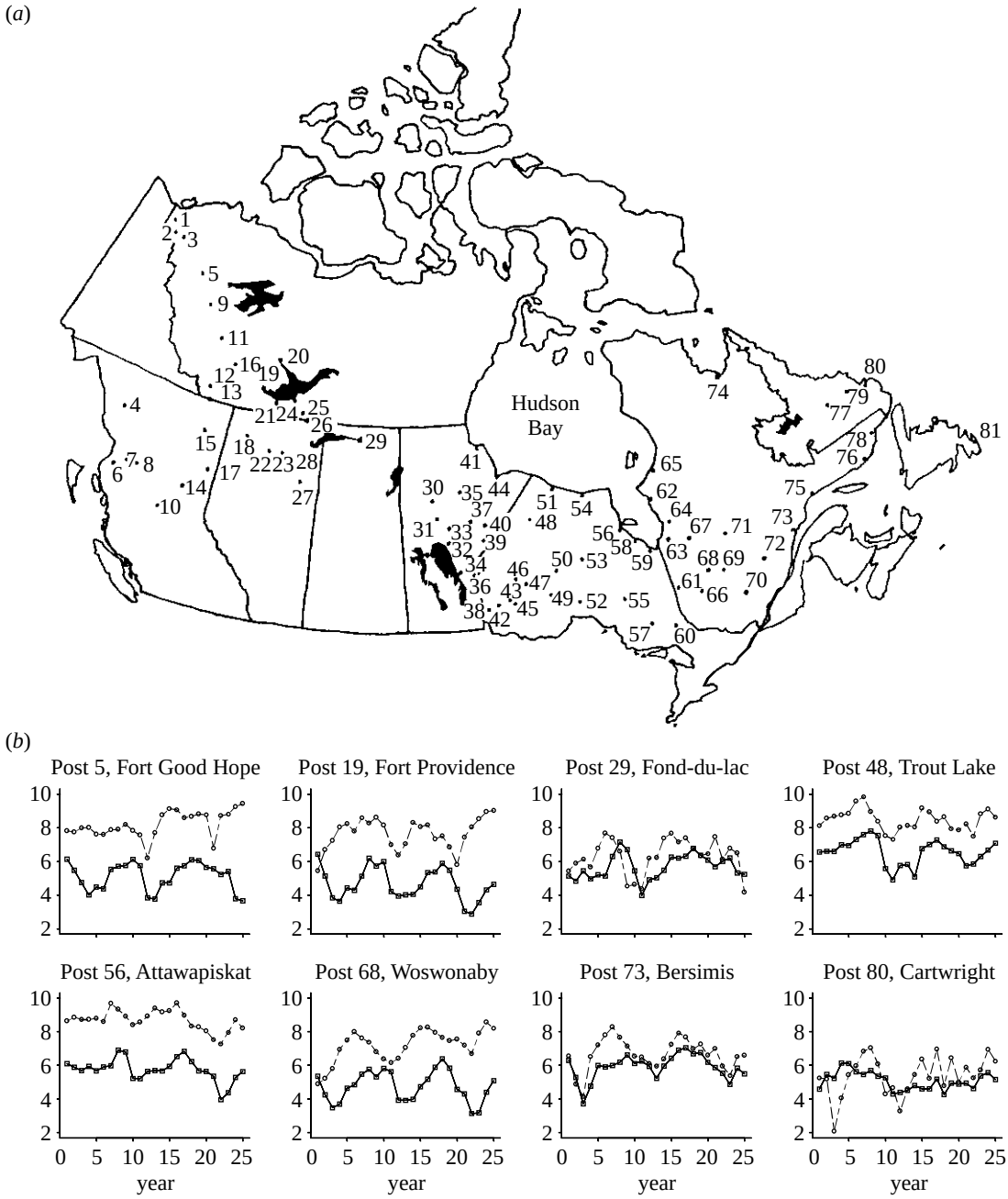

Figure 1. (*a*) A map of 81 trapping posts for the mink and the muskrat in Canada in 1925–1949. (*b*) Time-series plots of mink and muskrat data (on the natural logarithmic scale) from eight randomly selected posts. Solid lines and squares, mink; dashed lines and circles, muskrats.

opportunities for the muskrat to hide and thereby avoid the chance of being preyed upon by the mink. Hence, the mink–muskrat data provides us with both the motive and a testing ground for developing tests for common structure. Whereas asymptotic results for large data sets are available, the problem of testing for common structure in small data sets has received little attention. In this paper, we develop a new statistical test based on a bootstrap approach that serves exactly this purpose. The bootstrap is known to be a particularly powerful tool for data analysis. Its good performance in many important statistical problems such as estimation and model fitting has been established by theoretical analyses, by simulation studies and by applications to real data (see, for example, Efron & Tibshirani 1993; Shao & Tu 1995).

Within a general setting of regression models, we will perform tests of homogeneity in the regression functions and the variances. We justify the applicability of the bootstrap methods by showing that the asymptotic bootstrap distributions are indeed the asymptotic distributions of the test statistics under the null hypotheses. We conduct simulations for a set of linear models, which show that for finite sample sizes the bootstrap provides better approximation to the distributions of the test statistics than classical asymptotic theory.

The paper is organized as follows. In § 2 we introduce the statistical hypotheses of common structure and the bootstrap tests. The application to the Canadian mink–muskrat data is presented in § 3, where we also touch on the identification of regions for which the mink and the

muskrat interact with each other in a similar manner, although this is not the main focus of this paper. The asymptotic results, which rigorously justify the use of the test, are given in two appendices. Numerical examples using simulated models are reported in Appendix C.

2. TESTS FOR COMMON STRUCTURE

(a) *Test statistics*

Consider p regression models

$$Y_{ik} = m(X_{ik}, \theta_k) + \varepsilon_{ik}, \qquad i = 1, \ldots, n_k;\ k = 1, \ldots, p, \tag{1}$$

where Y_{ik} is a random variable, X_{ik} is a $q \times 1$ random vector, the functional form of m is known from biological theory, θ_k is an unknown $d \times 1$ parameter vector (θ_k stands for the autoregressive parameter in the mink–muskrat dynamic model in § 3 below), $\{\{\varepsilon_{ik}\}, k = 1, \ldots, p\}$ are p independent stochastic sequences with mean zero and variance

$$\mathrm{Var}(\varepsilon_{ik}) = \sigma_k^2. \tag{2}$$

Note that for an autoregressive time-series model the set of covariates consists of lagged values of the response variable, i.e. $X_{ik} = (Y_{i-1,k}, \ldots, Y_{i-q,k})^\tau$.

A common structure implies that the parameters in the regression functions and the error variances are the same. Thus, we formulate two null hypotheses:

$$H_0 : \theta_1 = \ldots = \theta_p, \tag{3}$$

$$\mathcal{J}_0 : \sigma_1^2 = \ldots = \sigma_p^2. \tag{4}$$

Under the assumption of normality of ε_{ik}, classical statistical theory provides us with the log-likelihood ratio statistics for testing H_0 and $\mathcal{J}_0$. These are, respectively, given by

$$T_H = \sum_{k=1}^{p} n_k \{\log S_k(\hat{\theta}) - \log S_k(\hat{\theta}_k)\}, \tag{5}$$

$$T_{\mathcal{J}} = n \log\left\{ n^{-1} \sum_{k=1}^{p} S_k(\hat{\theta}_k) \right\} - \sum_{k=1}^{p} n_k \log\{n_k^{-1} S_k(\hat{\theta}_k)\}, \tag{6}$$

where $n = n_1 + \ldots + n_p$. S_k stands for the sums of squares of error

$$S_k(\theta) = \sum_{i=1}^{n_k} \{Y_{ik} - m(X_{ik}, \theta)\}^2, \tag{7}$$

$$\hat{\theta}_k = \arg\min_{\theta} S_k(\theta), \tag{8a}$$

$$\hat{\theta} = \arg\min_{\theta} \sum_{k=1}^{p} n_k \log\{S_k(\theta)\}, \tag{8b}$$

where $\hat{\theta}_k$ is the least-squares estimator of the parameters of the kth individual model and $\hat{\theta}$ is the weighted least-squares estimator based on an overall model.

The standard theory of likelihood ratio tests entails that T_H and $T_{\mathcal{J}}$ are both asymptotically χ^2-distributed under H_0 and $\mathcal{J}_0$, respectively. In fact, this asymptotic property still holds under a more general condition (theorem B1 in Appendix B). In most cases, the least-squares estimation in equations (8) involves solving a nonlinear optimization problem. We adopt the downhill simplex method (Press *et al.* 1992, chapter 10.4) for the simulations and applications reported in this paper.

We now develop the bootstrap testing approach for both testing problems where we make use of the likelihood ratio test statistics evolving from classical normal theory.

(b) *Bootstrapping*

(i) *Bootstrapping for testing H_0*

Let

$$Y_{ik}^* = m(X_{ik}, \hat{\theta}) + \varepsilon_{ik}^*, \qquad i = 1, \ldots, n_k;\ k = 1, \ldots, p, \tag{9}$$

where $\hat{\theta}$ is the estimate given in equation (8*b*) from the pooled data from all p models. The $\{\varepsilon_{ik}^*\}$ are independent samples from the empirical distribution of the centred residuals $\{\hat{\varepsilon}_{ik} - \bar{\varepsilon}_k,\ i = 1, \ldots, n_k\}$, where

$$\hat{\varepsilon}_{ik} = Y_{ik} - m(X_{ik}, \hat{\theta}_k), \tag{10a}$$

$$\bar{\varepsilon}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \hat{\varepsilon}_{ik}, \tag{10b}$$

and $\hat{\theta}_k$ is the least-squares estimator in equation (8*a*). It can be seen that $E(\varepsilon_{1k}^*|\{X_{ij}, Y_{ij}\}) = 0$.

We define a bootstrap statistic T_H^* in the same way as T_H in equation (5) with $\{X_{ik}, Y_{ik}\}$ replaced by $\{X_{ik}, Y_{ik}^*\}$. We reject H_0 if T_H is greater than the upper α-point of the conditional distribution of T_H^* given $\{X_{ik}, Y_{ik}\}$. The latter can be evaluated via repeated samplings from equation (9). In fact, the p-value of the test is the relative frequency of the event $\{T_H^* \geqslant T_H\}$ in the bootstrap replications.

(ii) *Bootstrapping for testing $\mathcal{J}_0$*

We generate bootstrap samples from the equation

$$Y_{ik}^* = m(X_{ik}, \hat{\theta}_k) + \varepsilon_{ik}^\dagger, \qquad i = 1, \ldots, n_k;\ k = 1, \ldots, p, \tag{11}$$

where $\hat{\theta}_k$ is given as in equation (8*a*), and $\{\varepsilon_{ik}^\dagger\}$ are independent samples from the empirical distribution of $\{\tilde{\varepsilon}_{ik}, i = 1, \ldots, n_k\}$, and

$$\tilde{\varepsilon}_{ik} = \hat{\sigma}_k^{-1} \left\{ \sum_{l=1}^{p} \hat{\sigma}_l^2 / p \right\}^{1/2} (\hat{\varepsilon}_{ik} - \bar{\varepsilon}_k), \tag{12a}$$

$$\hat{\sigma}_k^2 = \frac{1}{n_k} \sum_{i=1}^{n_k} (\hat{\varepsilon}_{ik} - \bar{\varepsilon}_k)^2, \tag{12b}$$

where $\hat{\varepsilon}_{ik}$ and $\bar{\varepsilon}_k$ are given as in equation (10). We define $\{\tilde{\varepsilon}_{ik}\}$ in such a way that the null hypothesis $\mathcal{J}_0$ holds under model (11). It can be seen that

$$E(\varepsilon_{1k}^\dagger|\{X_{ij}, Y_{ij}\}) = 0, \tag{13a}$$

$$\mathrm{Var}(\varepsilon^{\dagger}_{1k}|\{X_{ij}, Y_{ij}\}) = \sigma^2_* \equiv \frac{1}{p}\sum_{l=1}^{p}\hat{\sigma}^2_l. \qquad (13b)$$

We define a bootstrap statistic $T^*_{\mathcal{J}}$ in the same way as $T_{\mathcal{J}}$ in equation (6) with $\{X_{ik}, Y_{ik}\}$ replaced by $\{X_{ik}, Y^*_{ik}\}$. We reject $\mathcal{J}_0$ if $T_{\mathcal{J}}$ is greater than the upper α-point of the conditional distribution of $T^*_{\mathcal{J}}$ given $\{X_{ik}, Y_{ik}\}$.

The use of the bootstrap tests is justified rigorously by theorems B1 and B2 (Appendix B), which state that the asymptotic conditional distribution of T^*_H (or $T^*_{\mathcal{J}}$) is almost surely equal to the asymptotic distribution of T_H (or $T_{\mathcal{J}}$) under hypothesis H_0 (or $\mathcal{J}_0$).

3. CANADIAN MINK–MUSKRAT DATA

(a) *Classification*

In this section, we aim at classifying the time-series from the 81 trapping posts shown in figure 1*a* into a smaller number of groups, such that within each group the mink and the muskrat interact in a similar manner. Before applying techniques of cluster analysis, we need to formulate an ecological model to describe the mink–muskrat interaction. Based on the food chain interaction model developed by May (1981), Stenseth *et al.* (1997) proposed a deterministic model to describe the predator–prey interaction, namely

$$\begin{cases} X_{t+1} - X_t = a_0(\theta_t) - a_1(\theta_t)X_t - a_2(\theta_t)Y_t, \\ Y_{t+1} - Y_t = b_0(\theta_t) - b_1(\theta_t)Y_t + b_2(\theta_t)X_t, \end{cases} \qquad (14)$$

where X_t and Y_t denote the population abundances, on a natural logarithmic scale, of a prey and the predator at time t, $a_i(\cdot)$ and $b_i(\cdot)$ are non-negative functions, and θ_t is an indicator representing the regime effect at time t, which is determined by X_t and/or Y_t. Biologically speaking, $a_1(\theta_t)$ and $b_1(\theta_t)$ reflect the within species regulation whereas $a_2(\theta_t)$ and $b_2(\theta_t)$ reflect, among others, the food chain interaction between the two species (see, for example, Stenseth *et al.* 1997, 1998*a*), and $a_0(\theta_t)$ and $b_0(\theta_t)$ are the intrinsic rates of changes. The implementation of the above food chain models for the purpose of data analysis has been facilitated by using a threshold to reflect the regime effect (see, for example, Stenseth 1999). Successful applications of this strategy include Framstad *et al.* (1997), Stenseth *et al.* (1998*a*,*b*) and Chan *et al.* (1997). Accordingly, we model the population abundance of mink at year $(t+1)$, i.e. Y_{t+1}, as a threshold regression on its lagged value Y_t and the population abundance of muskrat at year t, i.e. X_t, with a threshold on X_t. The implied model in which Y_t is moved from the left-hand side of the equation to the right-hand side, with added random noise, has the form

$$Y_{t+1} = \begin{cases} b_{10} + b_{11}Y_t + b_{12}X_t + \varepsilon_t & \text{if } X_t \leqslant r, \\ b_{20} + b_{21}Y_t + b_{22}X_t + \varepsilon_t & \text{if } X_t > r, \end{cases} \qquad (15)$$

where r is the threshold that divides the state space into two regimes called the lower and the upper regime, respectively. It is easy to see from the second equation in (14) that in the above model, both b_{12} and b_{22} should be non-negative. The following analysis is based on model (15), where we model the mink as a function of the immediate lagged values of both mink and muskrat.

Assuming a model of the form (15) for each of the 81 posts, we apply the bootstrap tests for H_0 and $\mathcal{J}_0$ to each possible pair chosen from the 81 posts. For each test, we search for the threshold r among the 60% inner sample range of X_t, and we replicate bootstrap sampling 400 times. Because the p-values across different tests are not always directly comparable in the sense that smaller p-values do not necessarily imply stronger statistical evidence against null hypotheses across different tests (see Gibbons & Pratt 1975), we introduce the following similarity measure for each pair

$$D(T, T^*) = \frac{1}{1 + \{(T - \mu(T^*))_+\}^2/\sigma^2(T^*)}, \qquad (16)$$

where T is the test statistic defined as in equation (5), T^* denotes its bootstrap counterpart, and $\mu(T^*)$ and $\sigma^2(T^*)$ are the conditional mean and the conditional variance of T^* given the original observations. In the above expression, $(\cdot)_+$ denotes the truncation at zero, i.e. $(x)_+$ equals x if $x > 0$ and zero otherwise. Without this truncation, $D(T, T^*)$ is the data depth defined by Mahalanobis (1936), which measures the depth (or the centrality) of a given point T with respect to the (conditional) distribution of T^* (see also Liu & Singh 1997). We truncate $(T - \mu_*)$ because our tests are one-sided.

Testing H_0 involves solving a nonlinear optimization problem for each replication of the bootstrap, which is time consuming, especially if we also conduct a genuine search for the threshold parameter r. To speed up the calculation, we assume in the test for H_0 that the threshold parameters are given in the bootstrap replications and are equal to their estimated values from the original data.

The next step is to perform the classification. There are several alternatives and we describe below one feasible method. A full development of optimal classification lies outside the scope of this paper. Because the data are the numbers of furs sold in the market, it is conceivable that different posts have different sampling weights. Therefore, we first standardized the mink series and the muskrat series separately for each post; i.e. we subtract the mean from original data and then divide them by the standard deviation for each series with length 25. We apply the following grouping strategy:

(i) apply a test for common structure for each pair among the 81 posts based on the threshold regression model (15),
(ii) form a similarity measure for each pair based on the above test using the modified Mahalanobis' data-depth (16),
(iii) group the posts by the complete linkage method (cluster analysis; see, for example, Sharma 1996, chapter 7.5).

We use the complete linkage method in order that all the posts in each cluster have a similar structure in the sense that the similarity measure between each pair within each cluster is as large as possible. The discrepancy in the regression functions across different posts is larger than the one in the variances. Further, the posts may be divided into three to seven clusters according to the similarity in the regression functions. Figure 2 presents the

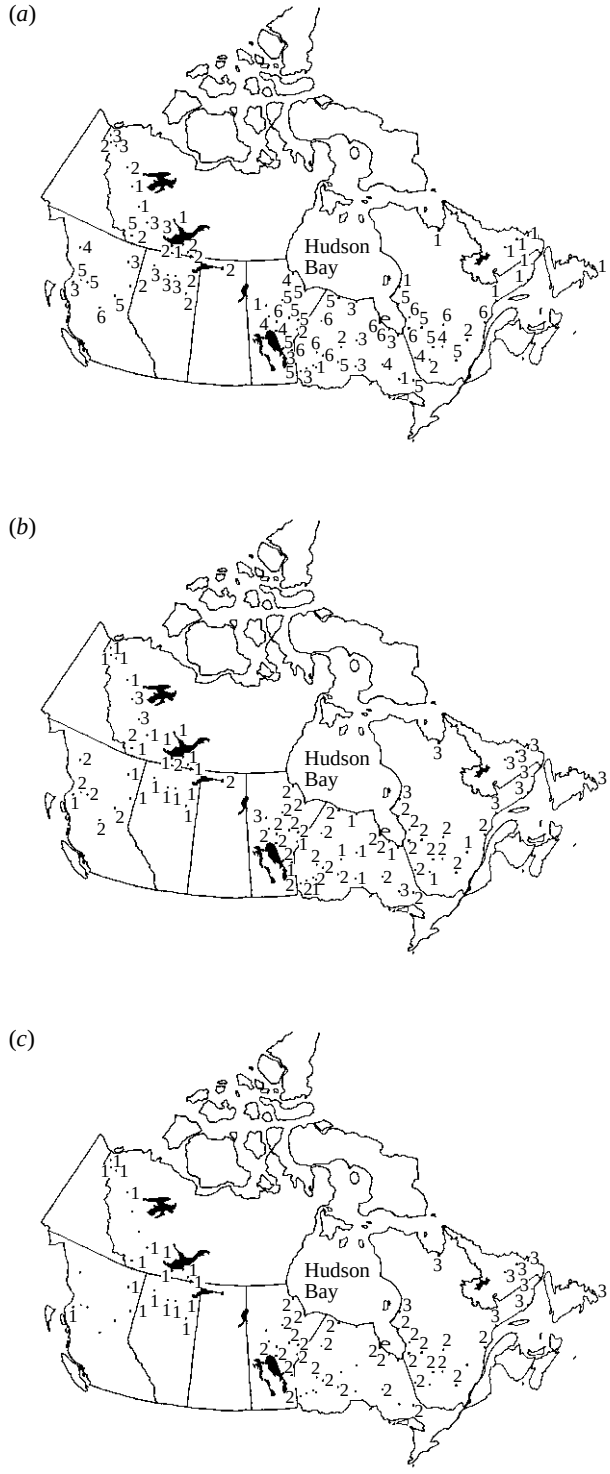

Figure 2. Classification of mink–muskrat posts based on tests for hypothesis H_0. Locations of the posts labelled according to the classification with (*a*) six clusters, and (*b*) three clusters. (*c*) Locations of the posts in the three selected groups from (*b*). Posts not in the groups are plotted without labels.

geographical map (using the longitude–latitude convention) of the classification with six clusters in figure 2*a* and three clusters in figure 2*b*. In figure 2*b*, the majorities of clusters 1, 2 and 3 are located in the western, middle and eastern area, respectively. This grouping has a clear geographic component and is therefore interesting from an ecological point of view (see Stenseth *et al*. 1999).

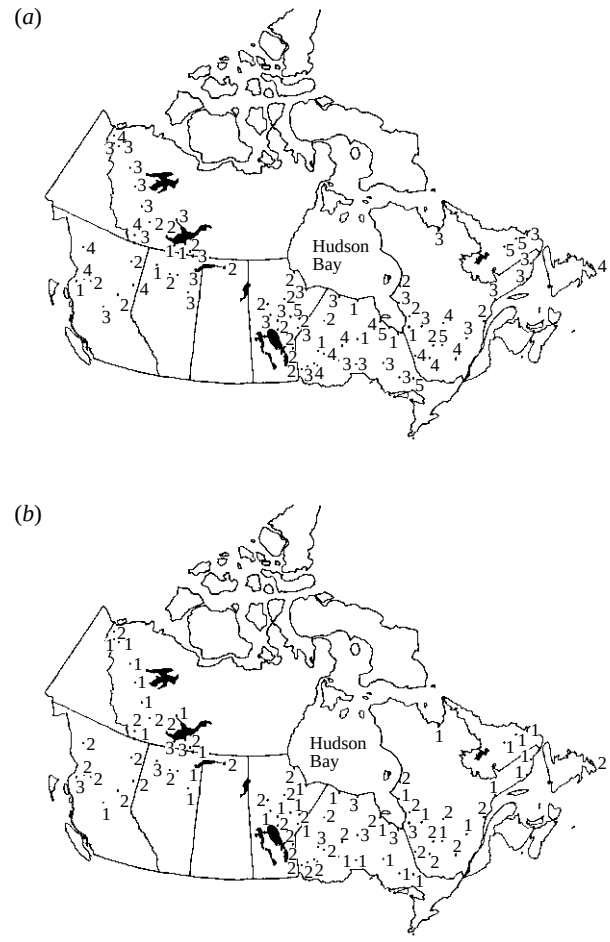

Figure 3. Classification of mink–muskrat posts based on tests for hypothesis $\mathcal{J}_0$. Locations of the posts labelled according to the classification with (*a*) five clusters, and (*b*) three clusters.

However, we may also divide the posts into three to five clusters according to the discrepancy in the variances. Figure 3 presents the two maps for such classifications with five and three clusters respectively. It seems that geographical location plays a less active role in determining the discrepancy of the variances.

(b) *Fitting the pooled data*

As an illustration, we report the results on modelling the pooled data for three selected groups, which consist of posts in clusters 1, 2 and 3 in figure 2*b*. Specifically, group 1 consists of the western posts in cluster 1, group 2 consists of the middle posts in cluster 2, and group 3 consists of all the eastern posts that belong to cluster 3. The locations of the posts in the three groups are shown in figure 2*c*. Model (15) is fitted to each group where the data is pooled from the posts within these groups. The parameters are estimated by least squares and the threshold r is searched within the 60% inner sample range. The estimated variance of noise $\hat{\sigma}^2$ is defined as the sum of squares of residuals divided by the sample size N minus seven (corresponding to seven free parameters in the model). The results are summarized in table 1. The standard errors of the estimated coefficients are reported in parentheses, which are calculated under the assumption

Table 1. *Fitting mink model (15) to the data in the selected groups*

group	N	regime	intercept	coefficient of Y_t	coefficient of X_t	r	$\hat{\sigma}^2$
1	432	lower	−0.14 (0.11)	0.48 (0.05)	0.36 (0.08)	−0.35	0.41
		upper	−0.12 (0.06)	0.63 (0.04)	0.18 (0.08)		
2	720	lower	−0.28 (0.12)	0.11 (0.05)	0.35 (0.09)	−0.58	0.46
		upper	−0.07 (0.04)	0.43 (0.04)	0.53 (0.05)		
3	216	lower	−0.46 (0.18)	0.64 (0.10)	−0.03 (0.15)	−0.26	0.65
		upper	0.15 (0.11)	0.45 (0.10)	0.06 (0.14)		

Table 2. *Fitting mink model (15) to the data in the western and central areas*

area	N	regime	intercept	coefficient of Y_t	coefficient of X_t	r	$\hat{\sigma}^2$
western	667	lower	−0.31 (0.09)	0.39 (0.05)	0.25 (0.08)	−0.35	0.44
		upper	0.09 (0.05)	0.61 (0.04)	0.20 (0.06)		
central	989	lower	−0.19 (0.11)	0.27 (0.05)	0.29 (0.08)	−0.57	0.46
		upper	−0.09 (0.03)	0.45 (0.03)	0.53 (0.04)		

that the threshold r is given (see Tong (1990) for an asymptotic justification).

The estimated coefficients of X_t are almost always positive. This reflects the fact that a large muskrat population will facilitate growth of the mink population, although the intensity of the increase is stronger in the middle core boreal forest (group 2) than in the western and eastern groups (groups 1 and 3). In fact, the food chain interaction of muskrat on mink is at its weakest in the eastern area where there may be a larger array of prey species for the mink to feed on. It should first be observed that the values of the estimated parameters are consistent with the predator–prey interaction specified in model (14). Our parameter estimates do specifically support an earlier suggestion (e.g. Errington 1961, 1963) that the mink and the muskrat relate to each other in a specialized food chain manner.

The coefficients of Y_t in model (15) correspond to the degree of self-regulation. The smaller are b_{11} and b_{21}, the stronger is the self-regulation. Hence, we may conclude that the weakest self-regulation in the mink is found in group 3. The strongest self-regulation within the mink population is found in group 2, where also the dependency on the muskrat in general is the largest. This suggests that group 2 corresponds to the region where there exists the closest predator–prey interaction between the mink and the muskrat. It seems plausible that this group corresponds to the core habitat of the muskrat in the boreal forest of Canada.

To reinforce the above analysis, we repeat the fitting for the pooled data from all posts in each of the western and central areas (without deleting the 'outliers'). The sample size is now 667 for the western area and 989 for the central area. The results are reported in table 2. Comparing it with table 1, the fitted models are not adversely different from the models based on the selected groups. The basic pattern described before is unchanged.

The threshold r defines the phases: the lower regime corresponds to the low and early increase phase, whereas the upper regime corresponds to the peak and decrease phase. Table 2 shows a clear phase dependency (see Framstad *et al.* 1997; Stenseth *et al.* 1998*b*) where the coefficients of the two regimes are significantly different for each of Y_t and X_t. The only exception is group 3 where the coefficients of X_t are very close to zero. The results for group 3 are different from those of the other two groups. In the light of what is known about this region, this should not be surprising: not only is there a larger array of prey species for the mink to feed on (making it less dependent on the muskrat), but also it is observed that foxes have a much more pronounced influence on the entire system of this region (e.g. Elton 1942).

4. CONCLUSIONS

By combining the information in panels of short time-series, we have been able to deduce the structure of the ecological process, which in our example is the predator–prey interaction between muskrat and mink over most parts of Canada. We have also deduced the existence of three ecological zones, each of which is characterized by different parameters in the ecological model. In the follow-up papers we will explore the ecological implications of the observed common ecological structure within the three regions. Over and beyond the ecological insight deduced from the model-fitting (briefly summarized in § 3(b)), we believe that our approach is of general interest and of wide applicability to the analysis of ecological and epidemiological time-series (both focusing on population biological processes) and to other fields wherever we are faced with panels of short time-series that share a common or similar underlying structure.

There has been a huge amount of literature on analysing panel data in econometrics (see, for example, Baltagi (1995) and references therein). In this paper we have focused specifically on the issue of population dynamics (as exemplified by the muskrat–mink interaction in Canada). It is our hope that the approach

developed in this paper will be of general applicability beyond the field of population biology. Indeed, we believe that the interplay between statistics and the areas posing difficult scientific problems may be greatly and mutually beneficial. Through the example used in this paper, we have shown the great merits of such work within the field of population ecology.

Both Q.Y. and H.T. were partially supported by the EPSRC grant L16385 and the BBSRC/EPSRC grant 96/MMI09785, H.T. was further supported by the Committee for Research and Conference Grants of the University of Hong Kong, B.F. was supported by the Wellcome Trust, and N.C.S. was partially supported by a Norwegian Science Council Grant and a special grant from the University of Oslo. We thank Professor C. J. Krebs for providing us with the map in figure 1*a*, and Dr M. Boyce and J. Erb for making available the Canadian mink–muskrat data. We are grateful to Sir David Cox, H. Viljugrein and reviewers for very helpful comments.

APPENDIX A. REGULARITY CONDITIONS

We always assume that $n_k = t_k n$ for some fixed $t_k \in (0, 1)$ when $n \to \infty$, $k = 1, \ldots, p$.

(a) ***Condition A1***

$\{\{\varepsilon_{ik}\}, k = 1, \ldots, p\}$ are p independent sequences. For each fixed k, $\{X_{ik}, \Upsilon_{ik}\}$ is jointly strictly stationary, and $\{\varepsilon_{ik}\}$ is a sequence of independent and identically distributed random variables with $E(\varepsilon_{ik}) = 0$ and $\mathrm{Var}(\varepsilon_{ik}) = \sigma_k^2$.

Furthermore, ε_{ik} is independent of $\{X_{jk}, j \leqslant i\}$ for each i and k.

(b) ***Condition A2***

All the second-order partial derivatives of $m(z, \theta)$ with respect to θ exist, and

$$E_{\theta_k}|\dot{m}_{j_1}(X_{s_1,k}, \theta_k)\dot{m}_{j_2}(X_{s_2,k}, \theta_k)\dot{m}_{j_3}(X_{s_3,k}, \theta_k)\dot{m}_{j_4}(X_{s_4,k}, \theta_k)| < \infty,$$

for $1 \leqslant j_i \leqslant d$, $1 \leqslant s_i \leqslant n_k$ and $1 \leqslant k \leqslant p$, (A1)

where $\dot{m}_i(x, \theta)$ denotes the partial derivative of m with respect to the ith component of θ.

(c) ***Condition A3***

$L_k(\theta) = E_{\theta_k}\{\Upsilon_{1k} - m(X_{1k}, \theta)\}^2$. Then $\partial L_k(\theta)/\partial\theta = 0$ if and only if $\theta = \theta_k$, and

$$\left.\frac{\partial^2 L_k(\theta)}{\partial\theta\partial\theta^\tau}\right|_{\theta=\theta_k} > 0, \qquad k = 1, \ldots p.$$

Further, for $L(\theta) \equiv \sum_{k=1}^{p} t_k \log\{L_k(\theta)\}$, there exists a unique θ_0 under hypothesis H_0 for which

$$\left.\frac{\partial L(\theta)}{\partial\theta}\right|_{\theta=\theta_0} = 0, \qquad \left.\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta^\tau}\right|_{\theta=\theta_0} > 0.$$

(d) ***Condition A4***

For $k = 1, \ldots, p$, $E_{\theta_k}(|\varepsilon_{1k}|^{4+\eta}) < \infty$, where $\eta > 0$ is a constant.

(e) ***Condition A5***

For each $1 \leqslant k \leqslant p$, $\{X_{ik}\}$ is ergodic, i.e. for any measurable f,

$$\frac{1}{n_k}\sum_{i=1}^{n_k} f(X_{ik}) \xrightarrow{a.s.} E_{\theta_k}\{f(X_{1k})\} \text{ as } n_k \to \infty,$$

provided $E_{\theta_k}|f(X_{1k})| < \infty$.

(f) ***Remark A1***

It is easy to see that $\partial L_k(\theta)/\partial\theta = 0$ if $\theta = \theta_k$. The assumption that θ_k is its unique root (as well as the uniqueness of θ_0) in condition (A3) is imposed to simplify the proof and is not essential.

Table C1. *The specification of the simulated models*

model	a	X_t	$\mathcal{Z}_t$	ε_t
(1)	1	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 0.25)$	$\mathcal{N}(0, 1)$
(2)	1	$U[0, 1]$	$U[-1, 1]$	$U[-\sqrt{3}, \sqrt{3}]$
(3)	1	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 4)$	$\mathcal{N}(0, 0.25)$
(4)	0	$U(0, 1)$	$\mathcal{N}(0, 1)$	$\mathcal{N}(0, 1)$

APPENDIX B. ASYMPTOTIC PROPERTIES

We now present some theoretical properties of our tests. Theorem B1 states the asymptotic distributions of the test statistics under their null hypotheses; theorem B2 presents the results for their bootstrap counterparts. The proofs of the theorems are available upon request. We always assume in this section that the regularity conditions listed in Appendix A hold.

(a) ***Theorem B1***

As the sample size n tends to ∞:

(i) T_H converges in distribution to a χ^2-distribution with $(dp - d)$ degrees of freedom under hypothesis H_0;

(ii) $T_{\mathcal{J}}$ converges in probability to the quadratic form $\boldsymbol{U}^\tau\phi\,\boldsymbol{U}$ under hypothesis $\mathcal{J}_0$, where $\boldsymbol{U}$ is a $d \times 1$ standard normal vector, and $\phi \equiv (\varphi_{ij})$ is a $p \times p$ symmetrical matrix with

$$\varphi_{ii} = \frac{1 - t_i}{2}\{E_{\theta_i}\{(\varepsilon_{1i}/\sigma_i)^4\} - 1\},$$

$$\varphi_{ij} = -\frac{1}{2}(t_i t_j)^{1/2}[E_{\theta_i}\{(\varepsilon_{1i}/\sigma_i)^4\} - 1]^{1/2} \times [E_{\theta_j}\{(\varepsilon_{1j}/\sigma_j)^4\} - 1]^{1/2}, \qquad i \neq j.$$

Further, if $E_{\theta_k}\{\varepsilon_{1k}/\sigma_k)^4\} \equiv 3$ for all $1 \leqslant k \leqslant p$, the asymptotic distribution of $T_{\mathcal{J}}$ is χ^2 with $(p - 1)$ degrees of freedom.

(b) ***Theorem B2***

Conditionally on $\{X_{ik}, \Upsilon_{ik}\}$, it holds almost surely as n tends to ∞ that

(i) T_H^* converges in distribution to a χ^2-distribution with $(dp - d)$ degrees of freedom, and

(ii) $T_{\mathcal{J}}^*$ converges in probability to the quadratic form $\boldsymbol{U}^\tau\phi\,\boldsymbol{U}$,

where $\boldsymbol{U}$ and ϕ are the same as in theorem B1.

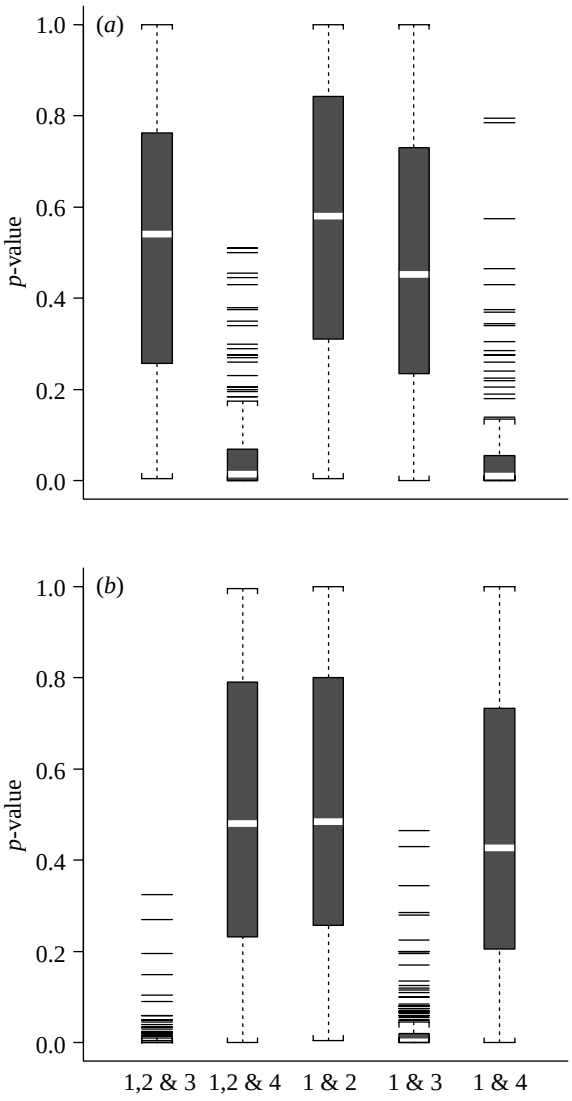

Figure C1. Simulation results for four linear models: (from left to right) boxplots of p-values for testing models 1, 2 and 3, models 1, 2 and 4, models 1 and 2, models 1 and 3, and models 1 and 4. (*a*) Bootstrap test for H_0. (*b*) Bootstrap test for $\mathcal{J}_0$.

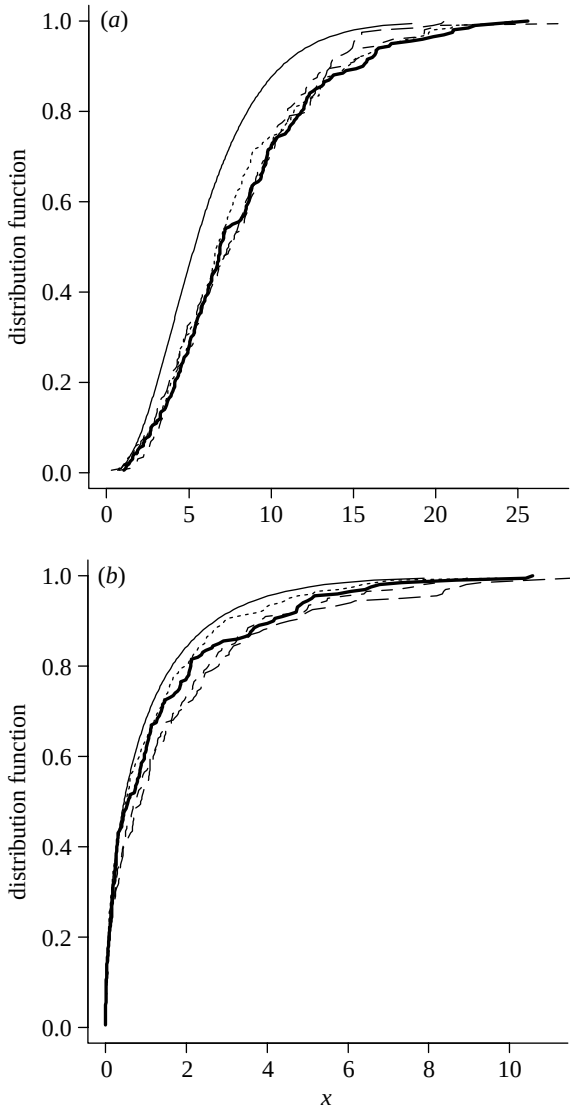

Figure C2. Comparison of bootstrap approximations and asymptotic approximations for distribution functions of test statistics: (*a*) T_H in testing H_0 for the first three models, and (*b*) $T_{\mathcal{J}}$ in testing $\mathcal{J}_0$ for models 1 and 4. The thick solid curve is the empirical distribution of the test statistic in a simulation with 200 replications; the thin solid curve is its asymptotic distribution; the three others are typical examples of bootstrap approximations.

APPENDIX C. NUMERICAL PROPERTIES

We illustrate the methods via a set of linear models. The basic model is set as

$$Y_t = a + 2.5X_t + 1.8Z_t + \varepsilon_t,$$

where X_t, Z_t and ε_t are independent. We allow the intercept a as well as the distributions of X_t, Z_t and ε_t to vary as in table C1.

Note that model 4 has a different mean function from the others, while the variance of the noise in model (3) differs from the other three models. We test H_0 and $\mathcal{J}_0$ for five different combinations of the models: the first three models, models 1, 2 and 4, models 1 and 2, models 1 and 3, and models 1 and 4. We set the sample size $n_k \equiv 24$; the bootstrap replications are 200 times. We repeat the simulation 200 times.

Figure C1*a*,*b* displays the boxplots of the p-values. We cannot reject H_0 when the first three models are considered together, although there is overwhelming evidence to reject $\mathcal{J}_0$ then. The hypothesis H_0 would be rejected but not $\mathcal{J}_0$ when models 1, 2 and 4 are considered. We could not reject both H_0 and $\mathcal{J}_0$ when we narrow our attention to models 1 and 2. The hypothesis $\mathcal{J}_0$ would be rejected but not H_0 when we compare models 1 and 3, and the hypothesis H_0 would be rejected but not $\mathcal{J}_0$ when we compare models 1 and 4. Note that the differences in the distributions of X_t, Z_t and ε_t are irrelevant to the hypotheses concerned. The reliable performance of the tests is further supported by table C2, which reports the simulated powers of the tests at the three different nominal levels (i.e. $\alpha = 0.01$, 0.05 and 0.10) in the above simulation.

We also compare the bootstrap approximations with the asymptotic approximations provided by theorem B1. Figure C2 presents the (approximated) distribution

Table C2. *Simulated powers of the bootstrap tests for H_0 and $\mathcal{J}_0$ for the four linear models*

models tested	hypothesis tested	average p-value	power ($\alpha = 0.01$)	power ($\alpha = 0.05$)	power ($\alpha = 0.10$)
1, 2, 3	H_0	0.510	0.01	0.04	0.09
	$\mathcal{J}_0$	0.011	0.84	0.95	0.98
1, 2, 4	H_0	0.063	0.47	0.70	0.79
	$\mathcal{J}_0$	0.493	0.01	0.05	0.10
1, 2	H_0	0.569	0.01	0.03	0.06
	$\mathcal{J}_0$	0.515	0.02	0.04	0.14
1, 3	H_0	0.482	0.02	0.06	0.11
	$\mathcal{J}_0$	0.029	0.70	0.82	0.92
1, 4	H_0	0.058	0.54	0.72	0.84
	$\mathcal{J}_0$	0.470	0.05	0.11	0.16

functions under the null hypotheses in two different settings: (i) the distribution of T_H in testing H_0 for the first three models, and (ii) the distribution of $T_{\mathcal{J}}$ in testing $\mathcal{J}_0$ for models 1 and 4. We plot the empirical distribution of a test statistic in a simulation with 200 replications, together with its asymptotic approximation, which is $\chi^2(6)$ in case (i) and $\chi^2(1)$ in case (ii), and three typical examples of its bootstrap approximations. A typical bootstrap approximation is selected in such a way that the corresponding p-value is equal to its 25th percentile (dotted curve), or the median (dot–dashed curve), or the 75th percentile (dashed curve). For the given sample size, bootstrap provides a better approximation than the asymptotic method, even for linear models.

REFERENCES

Baltagi, B. H. 1995 *Econometric analysis of panel data*. New York: Wiley.

Bjørnstad, O. N., Champely, S., Stenseth, N. C. & Saitoh, T. 1996 Cyclicity and stability of grey-sided voles, *Clethrionomys rufocanus*, of Hokkaido: evidence from spectral and principal component analysis. *Phil. Trans. R. Soc. Lond.* B **351**, 867–873.

Bjørnstad, O. N., Stenseth, N. C., Saitoh, T. & Lingjaerde, O. C. 1998 Mapping the regional transition to cyclicity in *Clethrionomys rufocanus*: spectral densities and functional data analysis. *Res. Popul. Ecol.* **40**, 77–84.

Chan, K. S., Tong, H. & Stenseth, N. C. 1997 Analysing abundance data from periodically fluctuating rodent populations by threshold models—a nearest neighbour bootstrap approach. Technical Report no. 258, Department of Statistics & Actuarial Science, University of Iowa.

Erb, J., Stenseth, N. C. & Boyce, M. S. 2000 Geographic variation in population cycles of Canadian muskrats (*Ondatra zibethicus*). *Can. J. Zool.* **78**, 1009–1016.

Efron, B. & Tibshirani, R. J. 1993 *An introduction to the bootstrap*. New York: Chapman & Hall.

Elton, C. S. 1924 Periodic fluctuations in the numbers of animals: their causes and effects. *Br. J. Exp. Biol.* **2**, 119–163.

Elton, C. S. 1942 *Voles, mice and lemmings*. Oxford, UK: Clarendon Press.

Errington, P. L. 1961 *Muskrats and marsh management*. Lincoln, Nebraska: University of Nebraska Press.

Errington, P. L. 1963 *Muskrat populations*. Ames, IA: Iowa State University Press.

Framstad, E., Stenseth, N. C., Bjørnstad, O. N. & Falck, W. 1997 Limit cycles in Norwegian lemmings: tensions between phase-dependence and density-dependence. *Proc. R. Soc. Lond.* B **264**, 31–38.

Gibbons, R. D. & Pratt J. 1975 *P*-values: interpretation and methodology. *Am. Stat.* **29**, 20–25.

Hanski, I., Turchin, P., Korpimäki, E. & Henttonen, H. 1993 Population oscillations of boreal rodents: regulation by mustelid predators leads to chaos. *Nature* **364**, 232–235.

Hjellvik, V. & Tjøstheim, D. 1999 Modeling panels of intercorrelated autoregressive time series. *Biometrika* **86**, 573–590.

Krebs, C. J., Boutin, S., Boonstra, R., Sinclair, A. R. E., Smith, J. N. M., Dale, M. R. T., Martin, K. & Turkington, R. 1995 Impact of food and predation on the snowshoe hare cycle. *Science* **269**, 1112–1115.

Lindström, J., Kokko, H., Ranta, E. & Linden, H. 1998 Predicting population fluctuations with artificial neural networks. *Wildl. Biol.* **4**, 47–53.

Liu, Y. & Singh, K. 1997 Notions of limiting p values based on data depth and bootstrap. *J. Am. Statist. Assoc.* **92**, 266–277.

May, R. M. 1981 Models for two interacting populations. In *Theoretical ecology* (ed. R. M. May), pp. 78–104. Oxford, UK: Blackwell.

May, R. M. 1986 When two and two do not make four: nonlinear phenomenon in ecology. *Proc. R. Soc. Lond.* B **228**, 241–266.

Mahalanobis, P. C. 1936 On the generalized distance in statistics. *Proc. Nat. Acad. India* **12**, 49–55.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 1992 *Numerical recipes in C*. Cambridge University Press.

Shao, J. & Tu, D. 1995 *The jackknife and bootstrap*. New York: Springer.

Sharma, S. 1996 *Applied multivariate techniques*. New York: Wiley.

Stenseth, N. C. 1999 Population cycles in voles and lemmings: density dependence and phase dependence in a stochastic world. *Oikos* **87**, 427–461.

Stenseth, N. C., Bjørnstad, O. N. & Falck, W. 1996*a* Is spacing behaviour coupled with predation causing the microtine density cycle? A synthesis of current process-oriented and pattern-oriented studies. *Proc. R. Soc. Lond.* B **263**, 1423–1435.

Stenseth, N. C., Bjørnstad, O. N. & Saitoh, T. 1996*b* A gradient from stable to cyclic populations of *Clethrionomys rufocanus* in Hokkaido, Japan. *Proc. R. Soc. Lond.* B **263**, 1117–1126.

Stenseth, N. C., Falck, W., Bjørnstad, O. N. & Krebs, C. J. 1997 Population regulation in snowshoe hare and Canadian lynx; asymmetric food web configurations between hare and lynx. *Proc. Natl Acad. Sci. USA* **94**, 5147–5152.

Stenseth, N. C., Falck, W., Chan, K. S., Bjørnstad, O. N., Tong, H., O'Donoghue, M., Boonstra, R., Boutin, S., Krebs, C. J. & Yoccoz, N. G. 1998*a* From patterns to processes: phase and density dependencies in the Canadian lynx cycle. *Proc. Natl Acad. Sci. USA* **95**, 15 430–15 435.

Stenseth, N. C., Chan, K. S., Framstad, E. & Tong, H. 1998*b* Phase- and density-dependent population dynamics in Norwegian lemmings: interaction between deterministic and stochastic processes. *Proc. R. Soc. Lond.* B **265**, 1957–1968.

Stenseth, N. C. (and 10 others) 1999 Common dynamic structure of Canada lynx populations within three climatic regions. *Science* **285**, 1071–1073.

Tong, H. 1990 *Non-linear time series: a dynamical system approach*. Oxford University Press.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.