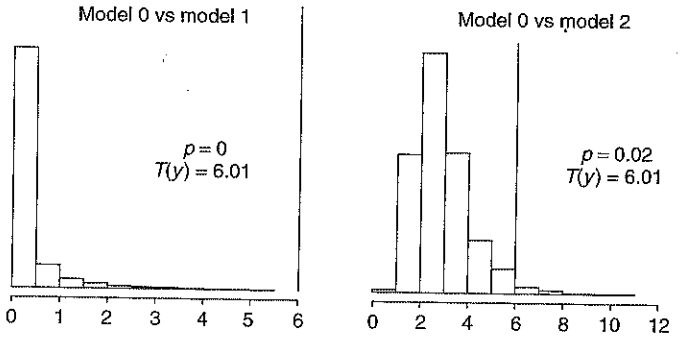

Figure 26.3 The posterior-predictive check. In each of the two histograms, the observed likelihood ratio test statistic (the vertical line) is compared with the posterior-predictive distribution of the test statistic under Model 0.

on the likelihood ratio test statistic,

$$T_i(y_{\text{rep}}) = \log\left\{\frac{\sup_{\theta\in\Theta_i} L(\theta|y_{\text{rep}})}{\sup_{\theta\in\Theta_0} L(\theta|y_{\text{rep}})}\right\}, \quad i = 1, 2,$$

where Θ_0, Θ_1, and Θ_2 represent the parameter spaces under Models 0, 1, and 2 respectively, and y_{rep} is a replicate data set. We can generate a sample from the posterior-predictive distribution of $T_i(y_{\text{rep}})$ under Model 0; we use the EM-type algorithms described above to compute $T_i(y_{\text{rep}})$. Histograms of $T_1(y_{\text{rep}})$ and $T_2(y_{\text{rep}})$ appear in Figure 26.3. Comparing these distributions with the observed values of the test statistics yields the posterior-predictive p-values in Figure 26.3. There is strong evidence for the presence of the emission line in the spectrum. Thus, Models 1 and 2 are preferable to Model 0.

27

Improved predictions of lynx trappings using a biological model

Cavan Reilly and Angelique Zeringue[1]

27.1 Introduction

Often statistics is viewed, and taught, as a series of procedures. In this view, methods are developed on the basis of some hypothesized data structure. The perspective that there are fixed data structures that can be treated as a whole misses the fascinating specificity of real-world problems. The field of time series prediction provides an excellent example of a well-defined data structure with a well-defined problem. In short, we assume we have a real-valued stochastic process that depends on time and our goal is to predict values of this process at some point in the future. If we assume the process is stationary, then there are representation theorems that provide us with a parameterized representation of any such series. Hence, to predict the series, we fit one of these parameterized forms and extrapolate. There are other classes of stochastic processes that have been developed to deal with nonstationary series, and while none of these has the same status as autoregressive moving averages, the same strategy is advocated: find a suitable parametric form from a class and estimate the parameters.

This general approach to statistics is often not the best approach to data analysis. As an example, we will consider prediction of the often-analyzed series of Canadian

[1]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minn.

Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives.
Edited by A. Gelman and X-L. Meng © 2004 John Wiley & Sons, Ltd ISBN: 0-470-09043-X

lynx trapped in the Mackenzie River area from 1821 to 1934 (Elton and Nicholson, 1942). We will develop a model using just the first 80 years, and then use this model to predict the series for the next 34 years. These data have been analyzed dozens of times see (Tong, 1990 for a review), often by methods that have no basis in population biology. For example, several early analyses fit a sine curve to the population over time and cleaned up the remaining lack of fit with an autoregression (Bulmer, 1974; Campbell and Walker, 1977). But why would a sine curve describe the dynamics of the lynx population? Clearly the lynx population fluctuates, but sine curves, or even finite linear combinations of such curves, are certainly not the only periodic functions. Perhaps such a model even provides good predictions, but could we do better using knowledge of the biology involved?

Our statistical model of the lynx series should be based on the biological context. This means that the model should attempt to describe fluctuations in the series in terms of the source of the fluctuations. As mentioned above, most approaches to statistical models of the lynx series have modeled the series as having fluctuations that are attributable to some form of autocorrelation in the series without attempting to understand why there would be such autocorrelation. The approach presented here assumes that these fluctuations are due to fluctuations in the primary food source of the lynx, namely, the snowshoe hare. The problem with this approach is that there is no data on the hare population for this period; hence we will need to impute the hare population, at least implicitly.

To understand the basis of the model developed below, we first note an important fact about the Canadian lynx. The Canadian lynx is an unusual predator in terms of its diet. This predator relies almost exclusively on a diet of snowshoe hare. When the hare become scarce in a region, the lynx will either move to other regions or slowly starve to death rather than switch their food source (McCord and Cardoza, 1982; Keith, 1990; Poole, 1994; Slough and Mowat, 1996; Brand and Keith, 1979). Other similar predators, such as the bobcat, will change their diet according to what food sources are available. Hence, our statistical model should attribute the source of fluctuations in the lynx population to fluctuations in the size of the hare population.

27.2 The current best model

There have been many attempts to model the lynx series: indeed, this series is considered a benchmark by many who work in nonlinear time series analysis. A rather comprehensive treatment of methods existing up to 1990 can be found in Tong (1990). As mentioned in the introduction, the first attempts at modeling this series combined autoregressions with sine curves. In 1980, Tong and Lim published a paper in which they used a self exciting threshold autoregression (SETAR) to model the lynx series. They had noticed that the series increased at a different rate than it declined, hence sine curves were inappropriate. SETAR models can display this behavior. Basically this model fits a different autoregression to the upswings and the downturns in the population. For model selection issues, they employed

Akaike's information criterion. Many other models have been fit to this data with varying degrees of success. Almost all of these models have been based on some proposed form of autocorrelation in the series. In reviews of various treatments, Lim (1987) and Lai (1996) both rated Tong's SETAR model to be the best in overall fit.

27.3 Biological models for predator prey systems

The most fundamental model of the interaction of a predator species with a prey species is provided by the Lotka–Volterra equations. These equations assume that the number of hare would increase exponentially in the absence of predation and the number of lynx would decay exponentially in the absence of hare. In addition, when there are lynx present in the system, the hare population will decrease exponentially at a rate depending on the population of lynx, and similarly the population of lynx will increase exponentially at a rate depending on the hare population. If $u_1(t) =$ the number of lynx at time t, and $u_2(t) =$ the number of snowshoe hare at time t, then this simple framework implies the following set of differential equations that describe the dynamics of the interaction between these two species

$$\frac{du_1}{dt} = -\alpha_1 u_1 + \beta_1 u_1 u_2$$
$$\frac{du_2}{dt} = \alpha_2 u_2 - \beta_2 u_1 u_2,$$

where α_j, β_j for $j = 1, 2$ are positive parameters.

From a biological perspective, this model has the obvious shortcoming that it does not consider the effect of other predators on the population of snowshoe hare. That is, to have a model that represents the interaction of species in this habitat, we should have more terms in the second equation of the form $-\beta_j u_j u_2$ for $j = 3, \ldots, J$, where $J - 1$ is the number of predators that consume snowshoe hare. Indeed, one can imagine a system of equations where there is an equation for each predator and an equation for each prey that describes which animals consume each other in a habitat. What makes the equation for the lynx unique is that it only depends on the hare population. To take advantage of this property of the lynx equation, we suppose there are two types of snowshoe hare: those that ultimately are consumed by lynx and those that are not. We can split the equation for the total hare population into two equations: one of the two equations will govern the dynamics of the population of hare that are consumed by lynx and one equation for all the other hare. The first of these equations will not depend on the population of any other predator and will be exactly of the form of the second equation above. These two equations will be related, but we assume that the effect of competition between hares is negligible compared to the effect of birth and death on the population. Such an assumption is a basic tenet of the Lotka-Volterra equations. Hence the effect of other predators is just

that now in the basic Lotka-Volterra equations presented above, $u_2(t)$ = the number of snowshoe hare alive at time t that are ultimately consumed by lynx. Of course, we cannot measure the number of hare today that will eventually be consumed by lynx, but it is nonetheless a well-defined concept. Actually, just determining the number of hare in a given habitat is a hard problem.

Another biological shortcoming of this model is the assumption that in the absence of predators, the snowshoe hare population will increase without bound. Clearly this is not realistic, as ultimately the food source of the hare will become depleted. To remedy this shortcoming, other terms are often added to the right side of the equations that include powers of the population of the species on the left side of the equation so that this behavior is ruled out. Rather than taking this route, we think of the system of equations as a useful model only when conditions are such that neither species dies out. That these conditions are applicable to the lynx/hare system over the last several hundred years, and that therefore this model is appropriate for the lynx/hare system, is obvious from the continued survival of both species.

A mathematical aspect of this model that has led some to conclude that it is not useful as a model in practice is that these equations are not structurally stable: small changes in the parameter values can lead to radical changes in the behavior of solutions. This has led some to abandon these equations or modify them to obtain a system that is better behaved. While this instability does make model fitting difficult, we can still use this set of equations to estimate parameters and make predictions, as we demonstrate in what follows. We do not think this structural instability makes the model unrealistic, as the world is full of phenomena that are quite sensitive to parameters.

27.4 Some statistical models based on the Lotka-Volterra system

Our first statistical model is based on the Lotka–Volterra system presented above. We observe the number of lynx trapped each year, $y(t)$ for 80 years. Although the number of lynx and hare can only take integer values, we model these quantities by real valued processes, as in the biological models presented above. We suppose that the expected proportion of lynx trapped each year is some constant proportion of the total number of lynx residing in the region, so that

$$y(t) = \alpha_0' u_1(t)\delta(t),$$

where $\delta(t), t = 1, \ldots, 80$ is a sequence of unit mean iid random variables that are independent of $u_1(t)$. For the purposes of conducting inference, we further assume these are lognormally distributed errors. The resulting model has 8 parameters: α_1, α_2, β_1, β_2, $u_1(1)$, $u_2(1)$, α_0', and σ, the standard deviation of the lognormally distributed errors.

This model is poorly identified; hence, we turned to the scientific literature in an attempt to construct informative priors. There are several methods that have been suggested for estimating parameters in the system. For example, one can construct an artificial habitat for hare so that no predation takes place. Observations on the hare population in such a setting could provide estimates of the birth rate of hare. But even in such situations, it is not clear that the birth rate is what it would be if there were lynx present. In any event, we can then assume that the birth rate of hare that ultimately get consumed by lynx is the same as the overall hare birthrate and obtain an informative prior for the birth rate parameter α_2. Other methods have been used to estimate the birthrate of hare, such as counting the mean number of young surviving. Similar techniques have been used to estimate the death rate of lynx (Poole, 1994; Slough and Mowat, 1996; Brand and Keith, 1979).

Unfortunately, we found that unless we used prior distributions with smaller standard deviations than the prior information really indicates, the posterior is too diffuse, as we describe below in the section on posterior simulation. For this model, the model parameters and the predictions themselves diverged as the Metropolis algorithm proceeded. Despite this, the predictions of the model at the best local mode we could find were very good, but we are reluctant to recommend the use of such predictions in general.

A simple reparameterization leads to a model with six parameters, and the resulting model behaves much better. This reparameterization can be thought of as just changing the units of the system. By letting $\theta_1(t) = \log\big(\beta_2 u_1(t)\big)$ and $\theta_2(t) = \log\big(\beta_1 u_2(t)\big)$ we obtain the system,

$$\log\big(y(t)\big) = \alpha_0 + \theta_1(t) + \epsilon(t)$$
$$\frac{d\theta_1}{dt} = e^{\theta_2} - \alpha_1$$
$$\frac{d\theta_2}{dt} = \alpha_2 - e^{\theta_1},$$

where $\epsilon(t)$ for $t = 1, \ldots, 80$ is a sequence of independent normal measurement errors. We then have six parameters in the model ($\theta_1(1)$, $\theta_2(1)$, α_0, α_1, α_2, and σ). Unfortunately, although it is not immediately transparent, these six parameters are not identifiable.

To understand the nature of the identifiability problem here, we need to consider the trajectories of the system. The system has a non-hyperbolic fixed point at $(\theta_1 = \log\alpha_2, \theta_2 = \log\alpha_1)$. If we take the ratio of the equations that define the system, we obtain the differential equation

$$\frac{d\theta_1}{d\theta_2} = \frac{e_2^{\theta} - \alpha_1}{\alpha_2 - e_1^{\theta}},$$

which can be solved to yield an equation that describes the trajectories of the system in phase space

$$\alpha_2\theta_1(t) - e^{\theta_1(t)} + \alpha_1\theta_2(t) - e^{\theta_2(t)} = \alpha_2\theta_1(1) - e^{\theta_1(1)} + \alpha_1\theta_2(1) - e^{\theta_2(1)}.$$

If we define $f(x) = \alpha_2 x - e^x$, then f is concave and has a unique maximum at $\log \alpha_2$, hence provided $\alpha_2\theta_1(1) - e^{\theta_1(1)} + \alpha_1\theta_2(1) - e^{\theta_2(1)} + e^{\theta_2(t)} - \alpha_1\theta_2(t) < \alpha_2(\log \alpha_2 - 1)$, there are two distinct solutions to the previous equation, one less than $\log \alpha_2$ and one greater than $\log \alpha_2$. We can repeat this argument using a condition on $\theta_1(t)$ too, hence the set of trajectories implied by the model is a collection of closed curves. Moreover, we can see from the equation that for trajectories near the fixed point, these curves will be approximately ellipses. For the lynx data, given this parameterization, the data supports the trajectory being very close to the fixed point for the θ_2 dimension, hence an elliptical trajectory with respect to that dimension. But if the trajectory is an ellipse and we only have data related to the θ_1 axis, then any translation of the trajectory along the θ_2 axis will yield the same fit to the data. When we attempted to find the posterior mode or generate samples from the posterior, we noticed that α_2 and $\theta_2(1)$ always moved together—this is what we expect given the elliptical trajectories. Given this identifiability problem, we simply fix $\theta_2(1)$, the rescaled initial number of hare that are ultimately consumed by lynx, at some arbitrary value and use noninformative priors for the other parameters in the model. In general, fixing $\theta_2(1)$ may reduce the set of possible trajectories, but this does not appear to be the case for this data set. Also, by fixing $\theta_2(1)$, we clearly cannot interpret α_1, but α_2 is still interpretable. The resulting model has five parameters that we estimate from the data.

Prior information on the system

There have been a large number of field studies aimed at understanding the population dynamics of lynx and hare. None of these have generated long time series of the sort on which we will base our predictions. Instead, these studies typically observe the numbers of animals over a short time period. Of the facts that these studies have identified, a consistent observation has been that the lynx population reaches its peak 1 to 2 years after the hare population reaches its peak. That is, once the hare population starts to decline, the lynx population follows suit. The Lotka–Volterra system has the property that periodic solutions have a fixed period, hence we use a prior distribution on the system that states that the difference in time between the two peaks is 1.5 years with a standard deviation of 0.25. When we discuss computing the posterior at a location in parameter space we will make clear how one can use this prior information.

27.5 Computational aspects of posterior inference

Given the structure of our model, computation is quite difficult. Note that we have no data on the number of hare at any point in time. The point of using the Lotka–Volterra system is to have a functional form for the number of lynx over time that is consistent with models from population biology. Although we think the formulation of the system in terms of the number of lynx and hare is quite intuitive, one can take the hare out of the system and obtain a second order

differential equation for the lynx dynamics. Since we ultimately solve the system numerically, we end up converting back to two first-order equations in any event.

Computing the posterior at a location in parameter space

Since there is no explicit solution to the system of equations presented above, computation of the likelihood is not straightforward. We compute the log-likelihood at a point in parameter space $(\theta_1(1), \theta_2(1), \alpha_0, \alpha_1, \alpha_2, \sigma)$ by first computing the contribution to the log-likelihood of the first observation $y(1)$. Since $\log y(1) \sim N(\alpha_0 + \theta_1(1), \dot{\sigma}^2)$ this term is straightforward. To compute the contribution of $y(2)$ to the log-likelihood, we first numerically integrate the system forward in time one step to obtain $\theta_1(2)$ and $\theta_2(2)$, then we use $\log y(2) \sim N(\alpha_0 + \theta_1(2), \sigma^2)$ to determine the contribution of the second time point to the log-likelihood. Note that $\theta_1(2)$ will be a function of α_1 and α_2. If we iterate this process, we can compute the log-likelihood for all of the data in this fashion. Finally, given that we have computed the log-likelihood we simply add the terms from the log-prior to obtain the log-posterior.

To perform the numerical integration, we use the fourth-order Runge–Kutta method (for implementation see Press et al., 1992). In order to use a prior distribution on the distance between the peaks of the series, we need to modify the basic procedure outlined above. As described above, we will only have the values of the solution to the system of differential equations at integer values. While this is adequate for computing the log-likelihood, we actually need the values of the solution for times between the integer valued times in order to determine at what time the peak of each series occurs. To this end, we integrate the system forward in time and save the solution each tenth of a year. Then we examine the value of the solutions over this finer time scale in order to determine when the peaks occur in each series. From the time of the peaks of the two series, it is easy to get the distance between the peaks implied by the set of parameter values $(\theta_1(1), \theta_2(1), \alpha_0, \alpha_1, \alpha_2)$. We then use this distance between the peaks in the term for the log-prior. Since the distance between the peaks is the same for all peaks, we can save some computational time by only integrating over this fine scale for the first pair of peaks.

Finding posterior modes

Although our posterior is only five-dimensional, finding posterior modes is quite difficult since the posterior is computed by numerically solving a system of differential equations. We found that using the simulated annealing algorithm for optimization of functions with continuous arguments presented in Press et al. (1992) allowed us to find posterior modes with some success.

Since the use of that algorithm is not at all standardized, we briefly indicate how we were able to successfully use the method. The simulated annealing algorithm of Press et al. is a stochastic mode-finding algorithm based on the downhill

simplex method combined with a Metropolis-type algorithm. This algorithm has three parameters whose values greatly influence the utility of the approach: the initial computational temperature, the number of iterations at each temperature, and the percentage the computational temperature should decrease when lowered. We found that using an initial computational temperature of 1 that gets lowered every 500 iterations by 90% was useful for finding local modes here. Choice of the initial computational temperature has, in our experience, been the most important parameter when using this algorithm. One should monitor the best solution as the temperature is decreased. If the initial temperature is selected too high, then these best solutions tend not to be as good as the initial value. If this value is selected too low, then the algorithm usually converges quickly to a local mode.

Simulating from the posterior distribution

Since we can compute the log-posterior as described above, we can use the Metropolis algorithm to draw simulations from the posterior distribution. While we are actually only concerned with predictions based on the posterior mode, we used the Metropolis algorithm as a check on the propriety of the posterior distribution. We used the general strategy outlined in Gelman, Carlin, Stern, and Rubin (2003): a multivariate normal jumping distribution with an estimated covariance matrix that is scaled so that 30 to 40% of the jumps are accepted. Since we were not able to successfully compute the numerical derivatives of the log-posterior with adequate accuracy, we ran the chain for several thousand iterations to obtain an estimate of the covariance matrix, then used this estimate in the next run of the chain. It was by using the Metropolis algorithm with multiple chains that never converged that we were able to conclude that the model with six parameters and noninformative prior distributions did not give a proper posterior distribution. Similarly, when we used priors constructed from the literature, as previously mentioned, the chains still did not mix adequately to declare convergence of the chains (using Gelman and Rubin's $\sqrt{\hat{R}}$). As sometimes happens, although the posterior is mathematically proper when we use informative priors, if these priors are not adequately informative, the posterior can numerically behave as if it is not proper.

27.6 Posterior predictive checks and model expansion

While the model performs quite well in terms of prediction, if we perform diagnostic checks just using the first 80 years of data and our fitted model, we discover an important discrepancy between the model and the data. In Figure 27.1, we see a graph of the residuals at the posterior mode and a graph of the mean of

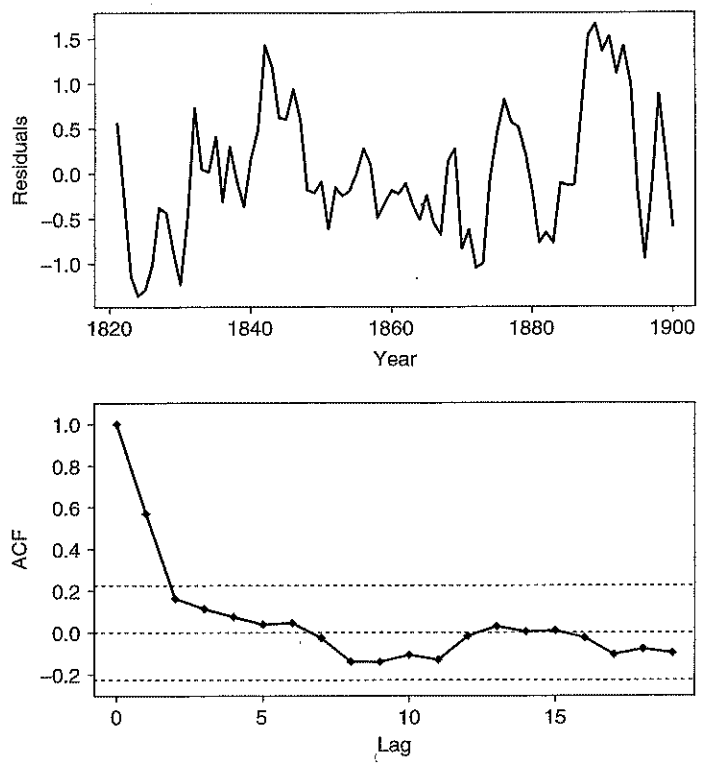

Figure 27.1 The residuals at the posterior mode and the mean of the posterior distribution of the residuals when there is no autoregressive component. There is autocorrelation at one lag.

the posterior distribution of the autocorrelation function of the residuals. We do not need to compute the posterior predictive distribution of the residuals in this example even though we are doing a posterior predictive check because we simply have iid Gaussian noise added to a functional form; hence, we know how large the autocorrelation function should be if there is really no autocorrelation. There is evidently substantial autocorrelation at lag one. This is not surprising given that there is an extensive literature indicating the presence of autocorrelation in this series, and here we see how posterior predictive checks can automatically detect such deviations from iid errors. There are basically two potential sources for this autocorrelation: the model dynamics are inadequate or the equation relating the dynamics to the measurements is incorrect. Since the model dynamics are based on the biological background, we expand our model to consider more realistic

models for the way the number of lynx trappings relate to the number of lynx. In particular, the assumption that the proportion of lynx trapped is constant over time seems questionable. We would expect that the effort of trappers to capture lynx is a function of the demand for lynx pelts. As lynx pelts are luxury items, the demand would be greatly affected by fluctuations in the business cycle. To model this effect, we suppose that the measurement errors are a realization from an autoregression. To determine the order of the autoregression, we fit the smallest number of terms to this autoregression so that there is no autocorrelation in the posterior predictive residuals. This exercise led us to conclude that a first-order autoregression (with parameter ϕ) is adequate to describe the deviation from iid errors. In Figure 27.2, we see the residuals at the mode and the mean of the posterior distribution of the residuals. In Figure 27.3, we see the fitted curve and the predictions for the lynx and the scaled hare population (scaled to fit on the graph). In particular, notice the asymmetry of the rise and decline in the populations over time.

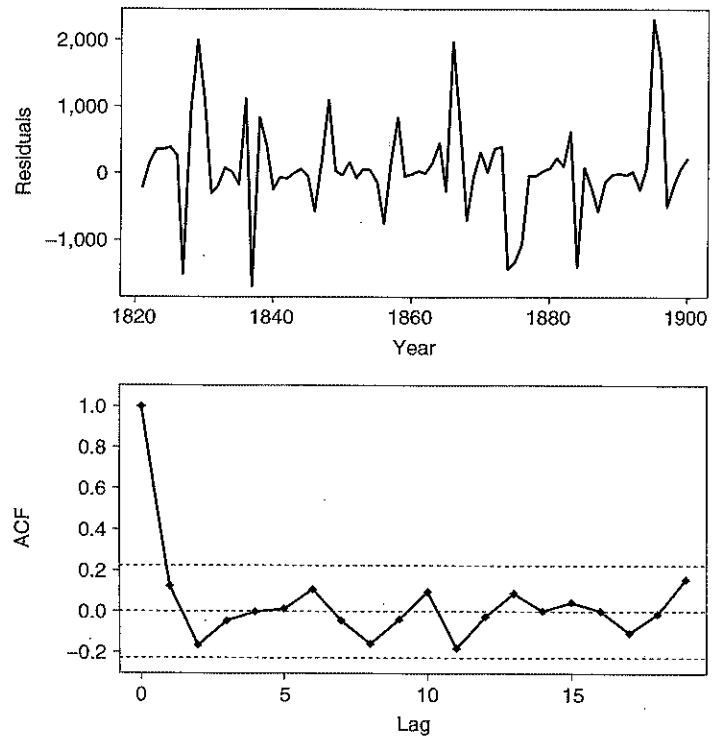

Figure 27.2 The residuals at the posterior mode and the mean of the posterior distribution of the residuals when there is a first-order autoregressive component. There is no evidence for autocorrelation.

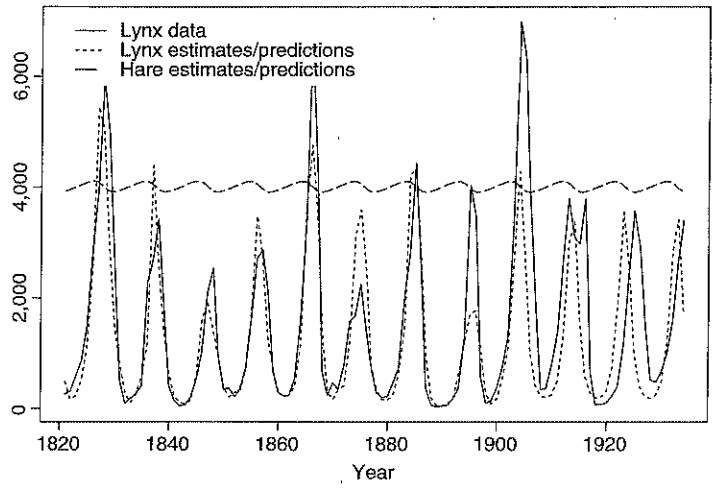

Figure 27.3 The lynx series, and the fitted values for the lynx and the hare. (The hare are scaled to fit in the graph.) The model is only fit to the first 80 years: the fitted values beyond year 1900 represent predictions.

27.7 Prediction with the posterior mode

Of course, without some regularity on the log-posterior we can never be sure that we have really found the global optimum. After running the simulated annealing algorithm for many iterations with many restarts from new locations in parameter space, we eventually became convinced that the best of the modes we had identified is the global optimum. Then we used this global optimum to make predictions. To obtain the predictions, we use the parameter values we found at the optimum and integrate the system forward in time starting from year 80 (the parameter values are $\alpha_0 = 14.4309$, $\alpha_1 = 804.209$, $\alpha_2 = 0.0006318$, $\theta_1(1) = -8.2474$, $\theta_2(1) = 6.6888$, $\sigma = 0.7151$, $\phi = 0.7431$). Although there are perhaps better ways to quantify the quality of a set of correlated predictions, we use the root-mean-square error of the predictions to quantify the quality of the predictions. For the above model, this quantity is 1,481.6. As noted above, perhaps the most widely supported model for this series is Tong's SETAR model. Tong fit his model to the entire series of 134 observations and using some model fit criteria, he eventually arrived at a 14-parameter model. To compare Tong's model to the model proposed here, we used Tong's parameter estimates (obtained from the entire series) and with his model made predictions starting from year 80 for the rest of the series. Strictly speaking, we should compare the predictions from our model to the predictions from a SETAR model fit to only the first 80 years of the series. In any event, the root-mean-squared error from Tong's model is 1,599.3; hence our model is better

in terms of prediction even though Tong got to use more data and his model has more than twice as parameters. While the model developed here generates accurate predictions, there are some large discrepancies between the fitted curve and data (e.g., around 1865). A better fit could be obtained by allowing noise in the system, that is, use a system of stochastic differential equations. Such a model would be more realistic as we would expect stochastic disturbances (e.g., the weather) to impact animal populations.

27.8 Discussion

We have shown here how using models based on the science at hand, when combined with state-of-the-art statistical methods, can greatly improve our long-term predictive ability. Similar phenomena are known to exist in prediction of economic time series, but in that case it is usually accepted that nonstructural models, such as time series models, can outperform structural models (those based on economic theory) in the short term. We have also illustrated that nonlinear dynamical models can be of use in applications, and are not useless pieces of theory from textbooks. The numerical challenges of such model fitting are not to be underestimated, but they are not insurmountable.

28

Record linkage using finite mixture models

Michael D. Larsen[1]

28.1 Introduction to record linkage

A goal of record linkage is to identify pairs of records (a, b), a from file A and b from file B, that correspond to the same person or entity. If there are no unique codes that identify the matching pairs of records, then links can be designated by comparing variables contained in the two files. In US census operations, social security number (SSN) is not collected, but first and last name, street address and house number, and other information are recorded. Often a great deal of work, including name and address parsing and standardization, is required to prepare files for comparison. If unique SSN's were recorded accurately for all individuals in both files, then the linkage task would be greatly simplified.

At the US Bureau of the Census, record linkage is an important step in undercount estimation and coverage evaluation. In order to evaluate the 1990 census, the Bureau of the Census conducted a post-enumeration survey (PES). The PES database was matched to census records. The number of individuals counted in both the census and the PES and the numbers counted in one but not in the other canvas, under an assumption of independence between enumerations, yields an overall estimate of the population. The actual estimation procedure is much more complicated in its details, but the idea is essentially the same. The 1990 PES is discussed in articles in volume 88 of *Journal of the American Statistical Association*

[1]Department of Statistics and CSSM, Iowa State University, Ames, Iowa.

Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives.
Edited by A. Gelman and X-L. Meng ISBN: 0-470-09043-X