1 Introduction

We believe that to make progress on this problem, it is crucial to come to terms with several issues. First, the true abundance of a cohort is never measured without error, and this must be explicitly included in the model. That is, the dynamics of population abundance, must be considered in terms of parameters, i.e. true population abundance, that cannot be observed without error. Cox (1981) makes the useful distinction between models driven by observations and parameter driven models; here we restrict our attention to the latter. Second, population dynamics is inherently nonlinear, and models must be formulated accordingly. Third, there is not enough data on any one cohort to obtain accurate estimates of density-dependent mortality, and thus the results must be combined among cohorts. Kiefer and Wolfowitz (1956) noted that in estimation situations where the number of parameters increases to infinity, maximum likelihood parameter estimates are not consistent, but by treating parameters as coming from a distribution—that is, as random effects—consistency could be obtained. This is the approach taken here. Lastly, even with excellent data, it is rare that definitive conclusions can be reached from any one data set; i.e. it is necessary to combine data across many populations.

We use simulation based methods to obtain maximum likelihood estimates for this problem (Mariano, Schuermann, and Weeks 2000). A simulation based method is used to evaluate the integrals that occur in the nonlinear random effects part of the model. Such models are known as dynamic panel models in the econometrics literature.

2 Formulation

We consider a hierarchical random effects nonlinear and non-Gaussian state-space model.

3 The State Space Model

Our general model can be represented as two sets of equations, one to represent the measurement process and one to represent the transition form one age to the next. The number of fish (or the logarithm of fish) estimated in year class t at age a from survey s is given by $y_{t,a,s}$ While $y_{t,a,s}$ can be observed, i.e. estimated, the true abundance, which we denote as $\alpha_{t,a}$, is an unobservable random variable known as a state variable (**?**). The observable is related to the state the measurement equation:

$$(1) \qquad\qquad y_{t,a,s} = h_a(\alpha_{t,a}, \epsilon_{t,a,s})$$

for $t = 1, 2, \cdots, T$, where T denotes the number of cohorts modelled and for $a = 0, 2, \cdots, A$ is the number of ages modelled. The estimation error, $\epsilon_{t,a,s}$, is typically assumed to be mutually, independent, normal errors (if the log of the abundance is used); however, we will discuss examples where independence and normality is not assumed. It is useful to define $Y_{t,b}$ to be the information on cohort t set up to age b, i.e., $Y_{t,b} = \{y_{t,0}, y_{t,1}, \cdots, y_{t,b}\}$.

The second key equation of a state-space model is the transition equation, which in our case describes the density-dependent and density-independent mortality between two

ages. The transition equation has the form

$$(2) \qquad\qquad \alpha_{y,a} = f_a(\alpha_{t,a-1}, \eta_{t,a}),$$

where $\eta_{t,a}$ is the stochastic variation in density-independent mortality, which is assumed to be mutually independently distributed from $\epsilon_{t,a,s}$; which is typically assumed to be normally distributed.

The measurement, $f_a(\cdot, \cdot)$, and transition functions, $g_a(\cdot, \cdot)$, depend upon parameters that are usually estimated from the data. Similarly, the functions may depend upon exogenous variables. For example, it is easy to incorporate delayed density-dependent mortality in the above formulation, or environmental variables that affect mortality. In the above equations these have been omitted for simplicity.

Note that we have subscripted the measurement and transition function by age, a. This is to remind the reader that we generally assume that the measurement and transition function changes with age.

4   The Random Effects Model

Any one cohort will be only observed a very few times, thus, we must combine information from many cohorts to estimate density-dependent mortality. The above model leads naturally to a random effects model for mortality and cohort strength. For example, if $\alpha_{t,0}$ is the log of the true cohort abundance at age 0, then it is reasonable to link the cohort abundance by a common distribution. For example, it might be reasonable to

assume

$$\alpha_{t,0} = N(\mu_0, \sigma_0^2).$$

Similarly, if we assume that the stochastic variation in density-independent mortality comes from a common distribution, e.g. $\eta_{t,a} \sim N(0, \sigma_0^2)$, then the estimation of the variance in density independent mortality can be estimated by using all the data.

5 State space model

In order to estimates the unknown parameters in equations 1 and 2, the likelihood function is maximized. In what follows, we will use the generic notation $P(\ \cdot\ |\ \cdot\ )$ to indicate the probability density of the first argument given the second. For example, let $P(y_{t,a}|\alpha_{t,a})$ be the conditional probability density function derived from Eq. (1) and let $P(\alpha_{t,a}|\alpha_{t,a-1})$ be the conditional density function derived from Eq. (2). Because the true abundance is never observed without error, it is necessary to use a Kalman filtering approach. Specifically, we first calculate the probability of $\alpha_{t,a}$ given past observations of the cohort. This is known as a one step ahead perdiction and is given by

$$P(\alpha_{t,a}|Y_{t,a-1}) = \int P(\alpha_{t,a}|\alpha_{t,a-1})P(\alpha_{t,a-1}|Y_{t,a-1})d\alpha_{t,a-1}$$

where the above equation is calculated recursively. The next step is to combine the new sample data, $y_{t,a}$ with the past information $Y_{t,a-1}$. This is done use Bayes rule,

$$P(\alpha_{t,a}|Y_{t,a}) = \frac{P(y_{t,a})P(\alpha_{t,a}|Y_{t,a-1})}{\int P(y_{t,a}|\alpha_{t,a})P(\alpha_{t,a}|Y_{t,a-1})d\alpha_{t,a}}$$

The first equation is usually known as the prediction equation, whereas the second is known as the update equation. The initial condition, is typically given by allowing the initial population density be determined by out random effects distribution. For example, is we used the above assumption that the initially cohort abundance was lognormally distributed, then we would have

$$P(\alpha_{t,0}|Y_{t,0}) = N(\mu_0, \sigma_0^2).$$

Using such an initial condition, it is possible to calculate each of the above quantities recursively.

We begin by specifying the initial conditions, i.e. the distribution for the initial abundance at age 0. There are two cases. There are cases where it might be reasonable to assume that the inital abundance, is known without error. For example, for marine fish populations the initial egg production may be estimated from a virtual population analysis, which the researcher may believe is very accurate. In this case,

6 Likelihood

We write the likelihood in terms of the probability of observing each new data point, $y_{t,a}$, given the previous data, i.e. $Y_{t,a-1}$, for that age and cohort. This is the innovation form of the likelihood and is given by

$$\prod_{t=1}^{T} P(Y_{t,A}) = \prod_{t=1}^{T}\prod_{a=0}^{A} P(Y_{t,a}) = \prod_{t=1}^{T}\prod_{a=0}^{A} \int P(y_{t,a}|\alpha_{t,a})P(\alpha_{t,a}|Y_{t,a-1})d\alpha_{t,a},$$

Note that the integral is the denominator in the update equation.

7 Simulation Based Maximum Likelihood Estimation

The state space models can only be solved analytically in the linear, normal case, otherwise the integrals are intractable. Here we use a simulation based approach to carry these estimates. We have chose to use a resamping approach because it provides good accuracy, can be applied in virtually any situation, and is relatively easy to program (Tanizaki 2002).

We begin by specifying the initial conditions, i.e. the distribution for the initial abundance at age 0. There are two cases. There are cases where it might be reasonable to assume that the inital abundance, is known without error. For example, for marine fish populations the initial egg production may be estimated from a virtual population analysis, which the researcher may believe is very accurate. In this case,

describe how to do this for the normal.

For the next age, we first apply a one step ahead predition step. (Eq 33) to generage random samples based upon the previous step; these are denoted as $\alpha_{j,t,1|0}$. For other ages these random numbers would be denoted as $\alpha_{j,t,a|a-1}$, we will explain this interms of the geneeral case. These are easily obtained from the prediction equation, i.e.

$$\alpha_{i,t,a|a-1} = f_a(\alpha_{j,t,a-1|a-1}, \eta_{i,t,a}$$

where $\alpha_{j,t,a-1|a-1}$ is chosen with probability $1/N$ and $\eta_{i,t,a}$ is the ith random draw from the ηa distribution.

QUESTION FROM THE ABOVE SENTENCE: WHY DO WE HAVE TO CHOSE ALPHA WITH PROBABILITY 1/N, why not keep exactly the same random numbers, nothing is lost????? The last sentence is from page 19 of tanizaki above equ. 34.

The update equation is SEE EQU 34 AND 35 FROM TANIZAKI

References

Cox, D. R. 1981. Statistical analysis of time series: some recent developments. Scand. J. Statist. **8**(2): 93–115. With discussion by Gudmundur Gudmundsson, Georg Lindgren, Lennart Bondesson, Erik Harsaae, Petter Laake, Katarina Juselius and Steffen L. Lauritzen, With a reply by the author.

Kiefer, J., and Wolfowitz, J. 1956. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann. Math. Statist. **27**: 887–906.

R. Mariano, T. Schuermann, and M. J. Weeks (Eds.) 2000. Simulation-based inference in econometrics. Cambridge University Press, Cambridge. Methods and applications.

Tanizaki, H. 2002. Nonlinear and non-Gaussian state space modeling with Monte Carlo techniques: A survey a comparative study. *In* Handbook of Statistics, Vol.21: Stochastic Processes: Modeling and Simulation. North-Holland, pp. in press.