Bias Estimation of Biological Reference Points.

Nick Grunloh, E.J. Dick, Herbie Lee

Introduction

- Hello. My name is Nick Grunloh.
- Thanks to everyone in attendance and thanks to my committee for coming together on π day 2022 to listen to my talk today.
- I'll be talking to you today about a Metamodeling approach for assessing estimation bias in important population dynamics models (the Schaeffer model), but also some extensions to this work as I finish up my PhD.
- This is work in collaboration with NOAA NMFS, and largely funded by NOAA Sea Grant, and I'd like to thank my collaborators with NOAA.

Data and Basic Modeling Structure

- The modeling context here is that of population dynamics model as they might be used for stock assessment.
- I'll be focusing on the

(Mangel et.al., 2013) Canadian Journal of Fisheries and Aquatic Science

- I'll drop us into that work in the setting of a BH-SRR production model
  - The primary insight of Mangel et.al. that I want to focus on is that while the space of $\frac{B^*}{B_0}$ and $\frac{F^*}{M}$ RPs is an entire 2D space, we quickly limit ourselves as we model.
  - With a 2 parameter production function, such as BH, if M is fixed this RP space is limited to a 1D curve.
  - Further if steepness (h) is specificed the space is further reduced to a

0D point along that curve, and we may have unintentionally selected our RPs before model even meets data.

- **Right Panel:** On the right I show an emperical way of noticing this
  - The black are posterior samples of the RPs for a 3 parameter Shepherd-like model fit to cowcod rockfish data
  - The red are posterior samples of the RPs for a BH model with fixed M
- **Next:**
  - Notice how the posterior has been squashed into the red curve $\left(\frac{1}{\frac{F^*}{M}+2}\right)$
  - Mangel et. al. suggests looking into 3-parameter curves to avoid back ourselves into a corner and unintentionally overspecifying our models in the prior.

Pella-Tomlinson Production Model

- Here I formulate a slightly simpler setting for investigating how models in the 2 parameter limited setting might be biased when fit to data from a 3 parameter production function.
- In particular I have a PT Production Model
- The model starts with a nuisance observation layer
- Biomass is driven by the given ODE.
- Production is given by R; That's the PT 3 parameter production function.
- **Right Panel:** On the right you can see the PT production function and how it uses its third γ parameter to lean the production function to the left or right.
  - When $\gamma = 2$; PT=Logistic Prodcution function and the curve is symmetric about B^*.
- **Next:** Recall the logistic production function is paramterized in terms of:
  - the slope at the origin (r) as seen in the slope of the blue line
  - and the right-hand x-intercept (K) as seen with the purple vertical line
- **Next:** Due to the symmetry of the shaefer model, the RP space is limited to the horizontal line 1/2 @ all F^*
  - For brevity, later I'll refer to this line in this space as the shaefer line.
  - PT can get off of this line by changing the γ parameter to lean left or right.

Simulation

- The goal is to investigate bias induced my fitting PT data with the restricted shaefer model.
- I simulate data off of this shaefer line and subsequently fit those data with a limited 2 paramter model
- Here I show a grid of location in RP space where I will simulate data and once the shaefer model is fit, that estimate will neccessarily have to fit on that $\frac{1}{2}$ line.
- In particular I'll point out these 4 red X's in the 4 corners, Since I will refer back to these examples in about 2 slides.
- these are examples of large model misspecification.

Catch

- I've looked at this basic setup across a range of different fishing behaviors.
  - Here I show 3 different fishing patterns
  - I've parameterized fishing relative to F^* so that this first constant line at 1 on the left represents constant fishing at F^*.
  - The second curve represents a more typical ramp up of fishing and subsequent backing off toward F^*.
  - And the third curve is probably a completely uncommon fishing behavior just to see.
- **Next** For Brevity here we will only look at results of the constant fishing case.
- We will see that the detail you can get out of this simulation setting is plenty rich and the simplicty of this constant catch is helpful for understanding the mechanisms by which bias is induced when fitting a two parameter production function to even slightly more complicated data.

Curves

- This slide visualizes the posterior fit of those data in the 4 large model misspecification corners of RP space
- In all cases the red lines represent the posterior fit of the Scheafer model, and the black is the truth of each of quantity.

- On the left I show the production curves of those 4 corner max misspecification fits with each plot in the relative position of each corner.
- In the middle column of small plots I shoe the posterior fit to biomass
- In the far right column of small plots I show the posterior fit to depletion
- **Left:** So starting on the left its immediatly possible to notice a few trends from these fits of the prodction function.
  - When the data are generated with $\frac{B^*}{B_0} > 1/2$ (above the scheaffer line) F^* is over estimated
    - We see that in the slope at the origin being too steep
  - When the data are generated with $\frac{B^*}{B_0} > 1/2$ (above the scheaffer line) F^* tends to be under estimated
  - Looking at these pictures you can start to understand why
    - When fishing is held at F^* the population simply declines exponentially from K to B^*.
    - The model only observes the right half of the true SRR
    - Due to the leaning of the true PT curves, and the symmetry of the logistic parabola, the logistic curve is learning about its slope at the origin entirely from data where depletion$> \frac{1}{2}$, and above the schaefer line PT is steeper that on the right half than it is on the left, and so we over estimate F^* for data generated above the line.
    - The vice versa phenomena occurs below the schaffer line.
    - Data is only obsevered on the right half of the production function $\Rightarrow$ PT is shallower on the right than on the left $\Rightarrow$ and so the logistic parabola estimate tends to under estimate F^*.
  - Thats my F^* story, but we can also observe some trends in biomass RPs.
  - Notice that the fits tend to match up the location of the humps fairly well
  - So the model tend to be estimating B^* fairly well, but since we are fitting a restrictive parabola something has to give and you can see that we are totally missing on K.
- **Biomass:** In the center column we can see for the most part we doing ok on Biomass
- **Depletion:** But when you rescale things to consider depletion, we can see that the posterior estimate of ratio of biomass realtive to K, is ending up completely wrong for the highly misspecified cases.

- So that can be a small lesson that even if our models manage to predict faily well, model misspecification can completely lead us to incorrect inferences for some quantites.

Heat Map

- The particular model fits from tyhe last slide are only as helpful as their standard errors allow, but when you observe trends on repeated sampling you can start to gain confidence.
- With my grid of locations in RP space I am able to get a grid of fits, and across that grid I am able to establish bias tends in all of the major latent model quanities, and thats what I show on this slide.
- **Top Right:** In the top right I'm showing the entire space of PT data and when those data are fit with a Scheafer model how do RP map onto the schaeffer line.
- For all other plots Red indicates over estimation of the modeled quanity and blue indicates underestimates of the modeled quantity.
- **Bottom Right** In the bottom right you can see the generalized version of the our F^* story
  - Below the line we underestimate F^*
  - Above the line we overestimate F^*
- **Bottom Middle** Next to that we see a very similar pattern, but this time the quantity being graphed is MSY

- **Top Middle** I show the bias in $\frac{B^*}{B_0}$
  - this picture is a law of nature for this simulation setting (estimate must land on the line)
- Whats more interesting is whats is the behavior of bias in the numerator (B^*) and the demoniator (K) not divided but independently
- Whatever the individual patterns are they need to divide back up to give this top middle picture.
- **Left:** On the left I show those quantities, the bias in the quantity B^* is shown *top left*, and the bias in K is shown on the *bottom left*.
- Interestingly, B^* shows large swaths of relatively little bias, and most of the pattern in the top middle panel comes from bias in K.
- The world did not have to be this way, but this says that B^* is often a robustly estimated quantity, but due to restrictive model misspecification, its a zero sum game and accuray in B^* often come at the cost of estimates of K.

Summary

- The statistician George Box famously said "All model are wrong, but some models are useful"
- My question is how useful are our models?, and if our models are useful. How useful are they? and when are they useful?
- This is a simulation based method for starting to understand that those questions.
- We want to expand these methods to more commonly used production functions
  - BH and Ricker
- and we want to try to get a similar analysis of these assumptiomtions when embeded inside of an age structured or delay difference model.
- but in this simple PT/Shaeffer setting we can see
  - as model misspecification increases biases in some quantities can become very large
  - estimates in B^* are tending to be less sensative to model misspecification than K
  - and F^* bias is going to tend to be strongly catch dependent