Journal of Bioeconomics manuscript No.
(will be inserted by the editor)

Removing Biases in Forecasts of Fishery Status

Christopher Costello · **Olivier Deschenes** ·
Ashley Larsen · **Steven Gaines**

Received: date / Accepted: date

Abstract A recent highly cited paper from this journal develops a model predicting maximum sustainable yield (*MSY*) of a fishery using the historical maximum catch ($MaxCatch$). The model is parameterized with a small sample of fisheries from the United States, and is subsequently applied to all of the world's fisheries to estimate the benefits of fishery recovery. That empirical relationship has been adopted for many subsequent high-profile analyses. Unfortunately, the analysis suffers from two important oversights: (1) because the model is non-linear, it suffers from "retransformation bias" and therefore the results significantly understate *MSY* and (2) the analysis is parameterized from of a very limited data set and so generalizability of the fitted empirical relationship between *MSY* and *MaxCatch* to global fisheries is questionable. In this note, we rectify both oversights and provide an updated estimate of the relationship between *MSY* and $MaxCatch$.

Keywords Retransformation bias · Fisheries · Maximum Sustainable Yield
JEL Classification: Q22 D57 F01

We thank Dan Ovando for assisting with data collection

C. Costello
Bren School, University of California, Santa Barbara, CA 93106
Tel.: +805-893-5802
Fax: +805-893-7612
E-mail: costello@bren.ucsb.edu *Visiting Researcher Lameta, Montpellier, France*

O. Deschenes
Department of Economics, University of California, Santa Barbara, CA 93106

A. Larsen
Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA 93106

S. Gaines
Bren School of Environmental Science and Management, University of California Santa Barbara, 93106

1 Introduction

A cornerstone of bioeconomic analysis in fisheries management is the ability to generate predictions into the future or across fisheries. A common approach involves estimating a model with a relatively small dataset, and then applying the parameter estimates of that model to a much larger dataset, thus generating out-of-sample predictions. An important recent paper in this journal used 25 data points from US fisheries to establish a relationship between maximum recorded catch and maximum sustainable yield (Srinivasan et al. 2010) . They estimated the following regression model:

$$\log_{10}(MSY_i) = \alpha_0 + \alpha_1 \log_{10}(MaxCatch_i) + \varepsilon_i. \tag{1}$$

For most of the world's fisheries, the variable MSY is not available, but the variable $MaxCatch$ commonly is. Thus, the parameter estimates $\hat{\alpha}_0$ and $\hat{\alpha}_1$ were then used to generate predictions of MSY for these fisheries, as follows:

$$\widehat{MSY}_i = 10^{(\hat{\alpha}_0 + \hat{\alpha}_1 \log(MaxCatch_i))} \tag{2}$$

Owing to its ease of use, broad applicability, and intuitive appeal, this predictive equation has been subsequently applied in many high-profile analyses including: (1) estimating the global increase in food provision from global fisheries (Halpern et al. 2012), (2) estimating regional fisheries losses from mis-management (Srinivasan et al. 2012) and (3) estimating the value of rebuilding global fisheries (Sumaila et al. 2012).

Unfortunately, the analysis provided in Srinivasan et al. (2010) suffers from two fundamental oversights. First, because their model is non-linear it must be corrected for a well-known bias called the "retransformation bias." Without doing so, the estimate given by Equation 2 will systematically underestimate MSY. Second, because the model is parameterized off of a very small dataset with a narrow range of input variables, it is unclear how the model will apply when extrapolated beyond the range of input data. In this note we describe and correct both the retransformation bias problem and the extrapolation problem, and ultimately provide an updated version of the prediction of MSY from $MaxCatch$.[1]

2 *Retransformation Bias*

Srinivasan et al. (2010) sought predictions of MSY for many of the world's fisheries from which to derive the consequences of fisheries recovery. Because Equation 1 has a non-linear transformation on the left-hand-side, however, the prediction given by Equation 2 is biased, and will thus under-estimate MSY. This result is well-known in the statistics literature, where it is called the *retransformation bias* (Duan 1983). While it has been recognized in some of the fisheries literature (See Loland et al. 2007, Leathwick et al. 2008, and Costello et al. 2012), many prominent papers in

[1] Although one can also quibble with the model itself: e.g., the form of non-linearity or whether there are missing variables (i.e., life history variables), we do not address these broader issues in this paper.

the bioeconomics literature are seemingly unaware, and also suffer from this bias. Indeed, any model of the form:

$$f(Y_i) = X_i\beta + \varepsilon_i \tag{3}$$

for non-linear function $f(Y)$ will suffer from this problem, so the simple prediction $f^{-1}(X_i\beta)$ will be a biased estimate of the expectation of Y_i. Fortunately, there is a compact, but accessible statistics literature describing the retransformation bias problem in general and providing useful corrections for it. To keep this discussion concrete, we focus solely on models where the left hand side contains either a ln (because it is most commonly used) or $\log_{10}$ (because that is what Srinivasan et al. (2010) used); generalizing to any non-linear function is straightforward. Consider the model:

$$ln(Y_i) = X_i\beta + \varepsilon_i \tag{4}$$

from which we estimate β. The expected value of Y_i is given by:

$$E(Y_i) = \exp(X_i\beta)E(\exp(\varepsilon_i)). \tag{5}$$

Assuming $\varepsilon_i \sim Normal(0, \sigma^2)$, it is easy to see that exponeniating this error term results in a lognormal distribution with mean $\exp(0.5\sigma^2)$. Thus, $E(\exp(\varepsilon_i)) = E(\exp(0.5\sigma^2)) > 1$ (generally), which implies $E(Y_i) > \exp(X_i\beta)$. In other words, ignoring the term $E(\exp(\varepsilon_i))$ causes the predicted value of Y_i to be underpredicted. The extent of the bias will depend on the magnitude of $E(\exp(\varepsilon_i))$, which depends on the error variance, σ^2. A practical, and straightforward correction is called the "smearing estimate" (Duan 1983) and simply involves using the residuals: $\hat{\varepsilon}_i \equiv ln(Y_i) - X_i\widehat{\beta}$, and using the correction factor:

$$\widehat{S}_1 = 1/n\sum_{i=1}^{n} \exp(\hat{\varepsilon}_i). \tag{6}$$

Then the final, unbiased estimate of the expected value of Y_i is simply:

$$\widehat{E}(Y_i) = \exp(X_i\beta)\widehat{S}_1 \tag{7}$$

A very similar process can be employed to correct bias stemming from a log_{10} transformed dependent variable. Consider the model,

$$log_{10}(Y_i) = X_i\beta + \varepsilon_i \tag{8}$$

from which we again estimate β. The expected value of Y_i is then given by:

$$E(Y_i) = 10^{(X_i\beta)}E(10^{\varepsilon_i}). \tag{9}$$

Again, unless the error is exactly zero, $E(10^{\varepsilon_i}) > 1$ and thus again $E(Y_i) > 10^{(X_i\beta)}$. The log_{10} retransformation does not have the convenient mathematical characteristics of the exponential function, but nonetheless, the corresponding smearing estimate can be used where the residuals, $\hat{\varepsilon}_i \equiv log_{10}(Y_i) - X_i\widehat{\beta}$, can be used in the correction factor,

$$\widehat{S}_2 = 1/n\sum_{i=1}^{n} 10^{\hat{\varepsilon}_i} \tag{10}$$

In this case, the unbiased estimate of the expected value of Y_i is given by:

$$\widehat{E}(Y_i) = 10^{(X_i\widehat{\beta})}\widehat{S}_2 \tag{11}$$

2.1 Extrapolation Problem

A second concern is that the data used by Srinivasan et al. (2010) to parameterize regression Equation 1 represent a very restricted range of *MSY* and *MaxCatch* relative to global fisheries. Both variables are concentrated at the center of their respective populations. If the extrapolation (Equation 2) is to be applied for fisheries with similar characteristics, then no extrapolation problem exists. But since the authors seek to extrapolate to most world fisheries, many of which have significantly smaller or larger values of these variables, then the estimates produced by Srinivasan et al. (2010) do not generalize.

This extrapolation problem is well-documented and well-understood in the literature. A simple means of rectifying this apparent oversight is simply to include a more broadly representative set of input data from which to parameterize the model. When the paper in question was written, however, no broad global dataset containing *MSY* and *MaxCatch* was available, so the authors made use of what data were available. Now, the RAM II dataset is publicly available, and over 100 of the fisheries in the database contain these two variables.[2] Whether parameterizing Equation 1 with this broader data set will change the estimates remains an empirical question.

3 Correcting the Oversights

We have illuminated potential biases in the main result from Srinivasan et al. (2010) arising from retransformation and extrapolation. Here, we correct those oversights and produce a new estimate of the link between *MaxCatch* and *MSY* for global fisheries. The original data and the regression line estimated by Srinivasan et al. (2010), are shown in Figure 1.

Our first task is to correct the estimates for the retransformation bias. The smearing estimate ($\widehat{S}_2$ in Equation 10) for this model is $\widehat{S}_2 = 1.23$, suggesting that the retransformation bias is considerable. The point estimate suggests that the authors have understated any given fishery's level of *MSY* by about 23%.

An even more dramatic picture emerges when we explore the possible consequences of using such a small, limited-range data set. We assembled *MSY* and *MaxCatch* for the 109 stocks for which these variables are available in the RAM II global database of stock assessments (Ricard et al. 2011). Using these data, we repeated the analysis of Srinivasan et al. (2010). The results of this analysis are shown in Figure 2, along with the comparable data used in the original paper, (18 species from 25 stocks available from North East Fisheries Science Center).

Two key observations are apparent. First, the new input data themselves represent a much wider range, suggesting that they are more representative of global fisheries. Second, the regression lines actually cross: for small and medium sized fisheries, the old line appears to overstate *MSY*, but for very large fisheries the old result appears to understate *MSY*.

[2] *MaxCatch* is from reported catch statistics and *MSY* is estimated from various forms of stock assessments.

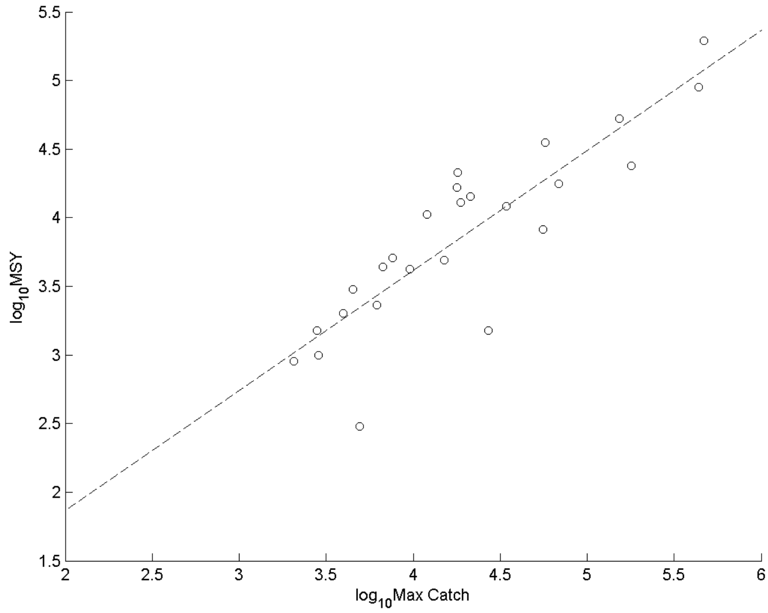

Fig. 1 Original data and regression line from Srinivasan et al. (2010).

Of course, the regression line calculated using the RAM II data is itself plagued by the retransformation bias problem identified above. It must be corrected, using Equation 10 for an unbiased estimate of MSY. Because these data are more spread than were the previous data, the smearing estimate is appreciably larger. Using these data, the smearing estimate is $\widehat{S_2} = 1.78$ suggesting that a naive application of Equation 2 to this new result would need to be 78% larger to erase the bias. Taking all of this together, the final, bias-corrected estimate of the expected value of MSY for a fishery is:

$$\widehat{E}(MSY) = 1.78 * 10^{(-.8644+1.0976\log(MaxCatch))} \tag{12}$$

Overall, this final estimate suggests that Srinivasan et al. (2010) significantly over-stated MSY for small fisheries and significantly understated MSY for large fisheries of the world. These errors could significantly alter predictions of the benefits of fishery reforms.

References

1. Srinivasan, U.T., Cheung, W.W.L., Watson, R., and Sumaila, U.R., Food security implications of global marine catch losses due to overfishing, Journal of Bioeconomics, 12, 183-200(2010)
2. Halpern, B.S., Longo, C., Hardy, D., McLeod, K.L., Samhouri, J.F., Katona, S.K., Kleisner, K., Lester, S.E., O'Leary, J., Ranelletti, M., and others, An index to assess the health and benefits of the global ocean, Nature, 488, 615-620(2012)

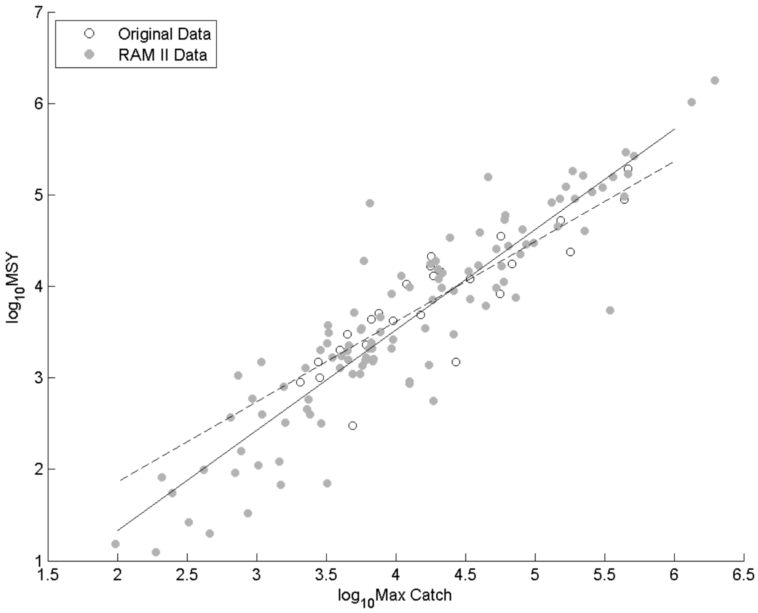

Fig. 2 Data and regressions from Srinivasan et al. (2010) (white circles, dashed line) and new data from RAM II (gray circles, solid line).

3. Duan, N., Smearing estimate: a nonparametric retransformation method, Journal of the American Statistical Association, 78, 605-610(1983)
4. Srinivasan, U.T., Watson, R., Rashid Sumaila, U., Global fisheries losses at the exclusive economic zone level, 1950 to present, Marine Policy, 36, 544-549(2012)
5. Sumaila, U.R., Cheung, W., Dyck, A., Gueye, K., Huang, L., Lam, V., Pauly, D., Srinivasan, T., Swartz, W., Watson, R. and others, Benefits of rebuilding global marine fisheries outweigh costs,PLoS One, 7, e40542(2012)
6. Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O., Lester, S.E., Status and solutions for the world's unassessed fisheries, Science, 338, 517-520(2012)
7. Leathwick, J, Moilanen, A., Francis, M., Elith, J., Taylor, P., Julian, K., Hastie, T., and Duffy, C., Novel methods for the design and evaluation of marine protected areas in offshore waters, Conservation Letters, 1, 91-102(2008)
8. Loland, A., Aldrin, M., Ona, E., Hjellvik, V., and Holst, J.C., Estimating and decomposing total uncertainty for survey-based abundance estimates of Norwegian spring-spawning herring, ICES Journal of Marine Science, 64,1302-1312(2007)
9. Ricard, D., Minto, C., Jensen, O.P. and Baum, J.K., Evaluating the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database, Fish and Fisheries, 13, 380-398(2011)