# Assessing the value of information in data poor stocks and providing best practice guidelines for data poor assessment approaches

Laurence T. Kell
Centre for Environmental Policy, Imperial College London, Weeks Building, 16-18 Princes Gardens, London
SW7 1NE, UK
laurie@seaplusplus.co.uk

Rishi Sharma
Food and Agricultural Organization, Fishery and Aquaculture Policy and Resources Division, Rome, Lazio,
00153, Italy
rishi.sharma@fao.org

Henning Winker
Joint Research Centre (JRC), European Commission, TP 051, Via Enrico Fermi 2749, 21027 Ispra, VA,
Italy.
e-mail address

March 21, 2021

**Abstract**

**Keywords:** biomass dynamic; catch-only methods; data poor; evaluation; stock assessment

# 1 Introduction

Fisheries are important both economically and socially, however, they are also a source of conflict since stocks often straddle Exclusive Economic Zones (EEZs) or are conducted in Areas Beyond National Jurisdiction (ABNJ). They may also impact endangered, threatened and protected species (ETP) or vulnerable marine ecosystems (VMEs). There are therefore potentially many stakeholders with conflicting objectives and divergent views, which may mean that uncertainties are used to support stakeholder positions in order to strengthen or weaken management measures [**?**].

Fish stocks can not be observed directly and estimates of stock status rely on models with a variety of assumptions, that use fishery-dependent and independent data sets. The quality of data from many small-scale fisheries is insufficient to allow for the application of conventional assessment methods [e.g. **?**]. In such case data-poor methods, such as those based on catch-only methods are used to estimate productivity and reconstruct historical abundance trends by making assumptions about final biomass relative to unfished and initial biomass [e.g. **???**]. However, simulation testing has indicated that catch-only methods only perform well only when assumptions regarding final relative abundance are met [**?**]. Despite these reservation estimates of global stock status require assessments of both data poor and data rich stocks are required for strategic planning. There is a need, therefore, for the validation of catch-only methods to increase trust amongst the public, stake and asset-holders and policymakers.

Validation requires model estimates to be compared to known values (i.e. observations) or well estimated historical values [**?**]. The only observations used in catch-only-method are catches, and if catch observations are removed then the method can not be run. Therefore to evaluate catch-only methods a reference set of data-rich stocks for a variety of regions, species and fisheries obtained from the RAM legacy database[1] was used.

We then evaluate the knowledge requirements, in the form of priors for population growth rate ($r$) and initial and final depletion, and the form of the production function for catch-only methods to provide assessments of stock status relative to maximum sustainable yield ($MSY$) targets reference points. We also compare catch-only methods to assessments that use an index of relative abundance for calibration. To do this we use Receiver Operating Characteristic (ROC) curves to evaluate the ability of models and reference points to identify and ranking stocks with respect to being overfished.

[1]RAM Legacy Stock Assessment Database. 2018. Version 4.44-assessment-only. Released 2018-12-22. Accessed [Date accessed 2020-10-30]. Retrieved from DOI:10.5281/zenodo.2542919.

# 2 Material and Methods

The RAM legacy database was used to simulate data-poor datasets and then a catch-only method based on a biomass dynamic model was used to estimate stock status relative to $MSY$ reference points, and compared to the original data-rich estimates. The objective is to evaluate the ability of data-poor methods to identify system state relative to targets reference points based on $MSY$. To do this we use Receiver Operating Characteristic (ROC) curves [?]

## 2.1 Material

The RAM Legacy database collates stock assessment data and estimates derived from a variety of methods. Assessments may be based on integrated statistical models using length and age data which estimate reference points as part of the fitting process, age-based models based on virtual population analysis where reference points are estimated in post-processing, or biomass dynamic models where assumptions related to density dependence (i.e. growth, mortality and recruitment) are modelled by a production function. Despite these difference, all models assume a production function, either explicitly as in the case of biomass dynamic or implicitly in the case of age-based models. The production function provides the basis of maximum sustainable yield ($MSY$) based on reference points [?]. Age-based models provide estimates of spawning stock biomass (SSB) and instantaneous fishing mortality, while biomass dynamic model estimates correspond to exploitable biomass and harvest rate.

Therefore time-series of biomass and exploitation in the RAM Legacy database are provided in a variety of ways. Trends in biomass are based either on $SSB$ or total biomass $B$, while trends in exploitation rate are based on either instantaneous fishing mortality ($F$) or annual exploitation rate ($U$).

Trends in $F/F_{MSY}$, $B/B_{MSY}$, and $Catch/MSY$ are shown in figures 1, 2 and 3 respectively. Individual stocks are represented by the faint lines, the median trend by the thick line,interquartiles by the thick dashed lines and the $90^{th}$ percentiles by the thin dashed lines. There has been a gradual increase in catches peaking around 1990, after which catches declined. Fishing mortality also peaked in 1990 and stayed around the $F_{MSY}$ until today, while the stocks have continued to decline. A noticeable feature is that some stock has shown wide variability while others have shown smooth trends. The median shows that stock have declined since the start of the series in the 1950s until 2000, after which the stocks stabilised. There is however a high degree of variability on a stock-by-stock basis. This is because fishing mortality had increased until the 1990s after which it varied just below the $F/F[MSY]$ level. This due to the adoption of target reference points based on $MSY$ by many management bodies after the adoption of the Precautionary Approach. Currently catches are less than $MSY$ and yields follow the general trends in biomass and fishing mortality, however yield and biomass are lagged biomass responding to catch a few years later. Fishing mortality is variable, reflecting that management is generally based on catch and that biomass is also influenced by process error, e.g. variability in year-class strength.

## 2.2 Methods

Rather than using a data poor package with a set of often poorly documented assumptions we used the JABBA biomass dynamic model [**?**] as it presents a unifying, flexible framework, based on a production function that can be used to estimate stock status and reference points under a variety of assumptions. JABBA is also used to conduct stock assessments for data moderate and rich stocks and so allows the value of improving data and knowledge to be evaluated.

A Pella Tomlinson production function [**?**] was assumed as this allows the shape of the production function to be varied, to represent alternative assumptions about productivity, stock status and reference points to be evaluated. At the data limited end of the stock assessment spectrum JABBA can be set up to approximate the behaviour of CMSY**?**, sampling from prior distributions to obtain parameter values that given a catch history that does not crash the population and satisfy priors for initial and final depletion. At the data rich end JABBA can be fitted to an index of relative abundance or catch-per-unit-effort data with priors for the production function, i.e. population growth rate ($r$) and virgin biomass ($K$), and initial and final depletion. All modelling was performed in R using the FLR simulation framework [**?**].

Two scenarios were considered for shape and $r$, since the shape of the production function is difficult to estimate from data alone even in data-rich assessments as it is determined by the form of density dependence assumed and parameters such as natural mortality ($M$) and the steepness of the stock-recruitment relationship ($h$) are difficult to estimate. Therefore the shape of the production function was assumed to be either logistic (Schaeffer) or Gompertz (Fox). In the latter case, production is maintained at the lower stock size, i.e. $B_{MSY}$ is found at a smaller fraction of virgin biomass $K$ than the former. The population growth rate at low stock size ($r$) can be derived from life-history parameters, however, it depends on the assumed values of $M$ and $h$. In many studies when developing uncertainty grids $M$ and $h$ are varied independently, however, $h$ and $M$ are related as $h$ describes density dependent mortality of recruits. Therefore we set up two scenarios when estimating $r$, based on low or high $M$ and $h$.

To provide priors for the other parameters, i.e. $K$ and initial and final depletion, each of the 4 scenarios (for assumed shape and $r$) were fitted to a perfect index of abundance, based on the RAM estimates of biomass. This provides unbiased priors consistent with the assessment estimates, CVs were set to 30%. This allows the Value-of-Infomation to be evaluated, i.e. how improving priors would increase yield. In addition heuristics for $K$, initial and final depletion was evaluated.

Table **??** summarises the alternative stock assessment assumptions made when running JABBA as a catch only method (COM).

We used Receiver Operating Characteristic curves to visualise the ability of models and reference points to identify whether a stock is overfished. To be useful for classification an indicator must have a high true positive rate (TPR) together with a low false-positive rate (FPR). A ROC curve, therefore, plots the TPR against the FPR. The area under the ROC curve gives an idea about the usefulness of a test, as the greater the area under a curve the better the test. The areas under ROC curves can also be used used to compare the usefulness of tests. ROC curves were first employed in the study of discriminator systems for the detection of radio signals in the presence of noise in the 1940s, following the attack on Pearl Harbor. The initial research was motivated by the desire to determine how the US RADAR "receiver operators" had missed the Japanese aircraft.

To be useful for classification an indicator must have a high true positive rate (TPR) together with a low false-positive rate (FPR). A ROC curve, therefore, plots the TPR against the FPR. The area under the ROC curve gives an idea about the usefulness of a test, as the greater the area under a curve the better the test. The areas under ROC curves can also be used used to compare the usefulness of tests. ROC curves were first employed in the study of discriminator systems for the detection of radio signals in the presence of noise in the 1940s, following the attack on Pearl Harbor. The initial research was motivated by the desire to determine how the US RADAR "receiver operators" had missed the Japanese aircraft.

A ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The ROC curves are constructed by sorting the observed values from the OM ($B/B_{MSY}$ and $F/F_{MSY}$) by their predicted scores based on the LBI with the highest scores first. The cumulative True Positive Rate (TPR) and True Negative Rate (TNR) are then calculated for the ordered observed outcomes. The ROC curve is then generated by plotting the cumulative distribution function (area under the probability distribution from to the discrimination threshold) of the detection probability in the y-axis versus the cumulative distribution function of the false-alarm probability on the x-axis. A ROC analysis, therefore, provides a tool to select the best candidate indicators.

The best performing indicators should pass through the top lefthand corner, as only positive cases should be identified, i.e. TPR=1 & FPR=0. Since the ROC is a probability curve the area under the curve (AUC) is an important metric for measuring performance. For example, a coin toss would produce a curve that fell along the $y = x$ line and the AUC would be equal to 0.5. The AUC, therefore, is a measure of how well an index is capable of distinguishing between states. The higher is the AUC, the better the model is at predicting. The ROC can also be used to identify the performance of a reference point, since the reference point is used as a discriminate threshold it should be the close point to TPR=1 & FPR=0.

# 3 Results

Estimates of $B : B_{MSY}$ are compared to the RAM legacy values in figure 6. If the COMs were able to assess the stock trends without error the points would lie along the $y = x$ line, while if they were able to classify stocks with respect to $B : B_{MSY}$ then points would either fall in the top right or bottom left quadrants defined by the red lines corresponding to $B : B_{MSY}$. The positive condition (P) is defined as the number of cases where $B \geq B_{MSY}$ in the reference set and the negative condition (N) where $B < B_{MSY}$. The number of real negative cases in the data true positive (TP) correctly classified are termed true negative (TN), while those incorrectly classified are known as false positive (FP) equivalent to a false alarm or Type I error; false negative (FN) are equivalent to a Type II error. Sensitivity ($\frac{TP}{TP+FN}$) measures the proportion of positives that are correctly identified, while specificity ($\frac{TN}{TN+FP}$) measures the proportion of negatives that are correctly identified.

Figure 7 shows ROC curves for COMs, comparing perfect knowledge for initial and final depletion (i.e. actual) to the heuristics. The blue dark line corresponds to using the heuristics alone without data. Scenarios were run for the shape of the production function (Schaeffer or Fox) and the $r$ prior (low or high). These are compared within a panel for using the known (actual) value for initial and final depletion or the heuristic. The curves were similar within a panel while the points identifying the reference level, based on the estimates $B : B_{MSY}$ and $F : F_{MSY}$vary across scenarios. This shows that the choice of production function and initial prior for $r$ do not have a major effect on ranking but are important for classification. Only when final depletion was known (first column) did the COM perform well. When the heuristic for final depletion was used the data had no effect.

To evaluate how well a biomass dynamic model with an index as well as catch would perform figure **??** shows the ROC curve for JABBA fitted to a perfect index with a CV of 30%.

The trends in $B : B_{MSY}$ by stock are shown in figure **??**. The black lines are the RAM values and the blue and red lines are the estimates using the perfect index, for two production functions (add to legend) and dashed/solid is for the $r$ prior; the horizontal lines indicate 120%, 100% and 0 of $B : B_{MSY}$. A range of trends are seen, although all stock has declined some continue to decline, some have stabilised (i, ii, ...) and others have started to recover (I, II, ...). Another characteristic is inter-annual variability since some stocks (A, B, ...) show smooth trends while others show fluctuations that appear to be independent of fishing (X, Y, ...)

# 4 Discussion

- Previous studies have used packages that implement the same basic algorithm with extra features such as ways for setting priors, the shape of the production function and heuristics for depletion. They have shown that the choices made are essential to get a good estimate of depletion in the final year [?]. However, even for data rich stocks productivity depends on difficult to estimate parameters such as the steepness of the stock recruitment relationship and natural mortality.

- A problem is how to make an objective choice between alternative COM runs when conducting sensitivity tests since diagnostics such as goodness of fit based on residuals, retrospective analysis and cross-validation are not applicable.

- We used a variety of stocks, regions and fisheries, however in many cases stock from a limited region are used to derive heuristics. However, as we showed heuristics had more impact than the data. A better alternative may be to use an indicator stock or fishery

- We used an assessment method that has been used to provide advice on stock status relative to target and limit reference points and to set quotas for data rich stocks. This allowed us to evaluate the impact of the different data types, assumptions, knowledge and priors on advice.

- We used ROC to compare the ability of methods to both rank and categorise stocks.

- COMs cannot be used for management, which requires monitoring the effectiveness of regulations, since to fit a COM this must already be known. Fisheries independent data is important to provide a robust assessment of stock status, as the data used to set catch are the same as the management regulation then it is unlikely that COMs can provide robust estimates.

- Or to combine indicators into a empirical rule that can be tested using MSE [?].

# 5 Conclusions

- Previous studies have used packages that implement the same basic algorithm with extra features such as ways for setting priors, the shape of the production function and heuristics for depletion. They have shown that the choices made are essential to get a good estimate of depletion in the final year [?]. However, even for data rich stocks productivity depends on difficult to estimate parameters such as the steepness of the stock recruitment relationship and natural mortality.

- We used JABBA a biomass based stock assessment method that is used to provide advice for a number of data rich stocks. This allowed us to evaluate the value-of-infomation across the data poor to rich spectrum.
- $F/F_{MSY}$ poorly estimated due to mining and lags.
- ROC curves, distance of threshold from [0,1] shows how good at classification of state, AUC shows how good at ranking.
- Data have no affect
- Snapshot, but then need to know depletion, i.e. priors are developed by region
- Can not validate, or use in MSE as a feedback controller
- Solution is to collect indices of abundance
- Develop MPs [?]

# 6 References

# References

# 7 Tables

# Figures

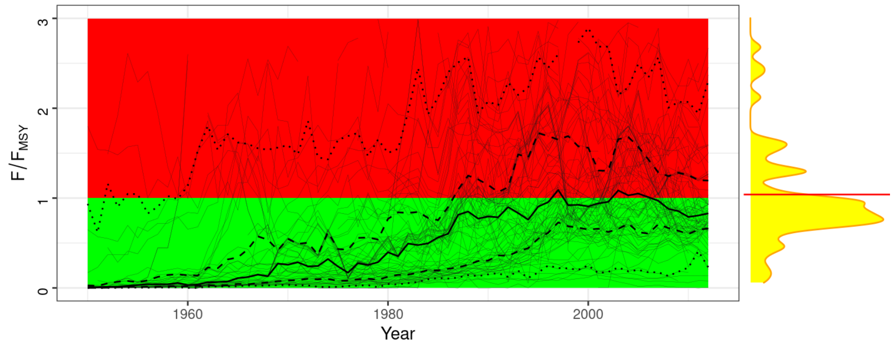


Figure 1: Time series of $F : F_{MSY}$ for the RAM Legacy database assessments.

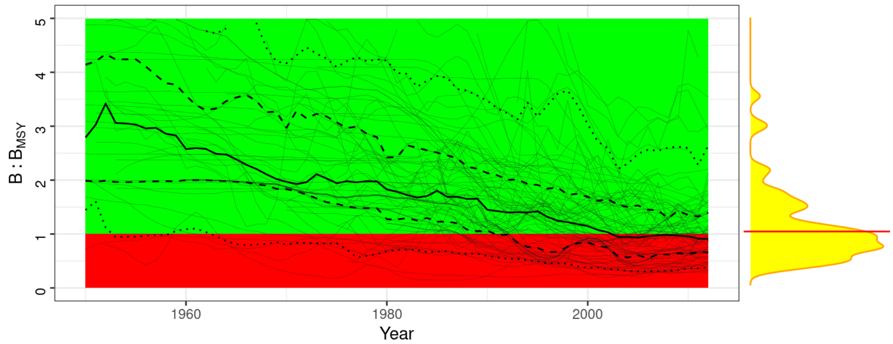


Figure 2: Time series of $B : B_{MSY}$ for the RAM Legacy database assessments.

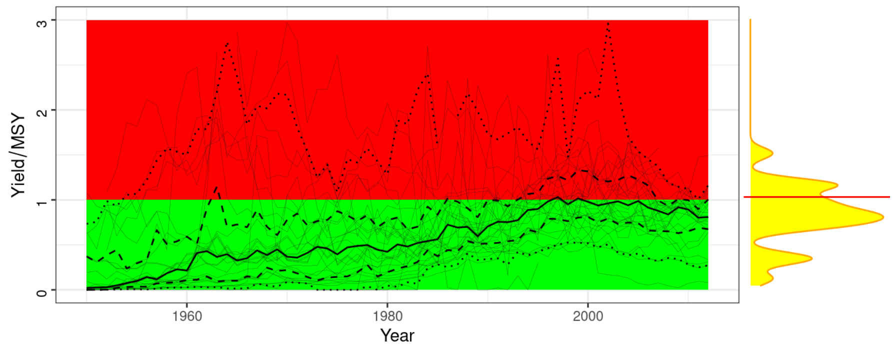


Figure 3: Time series of $Catch : MSY$ for the RAM Legacy database assessments.

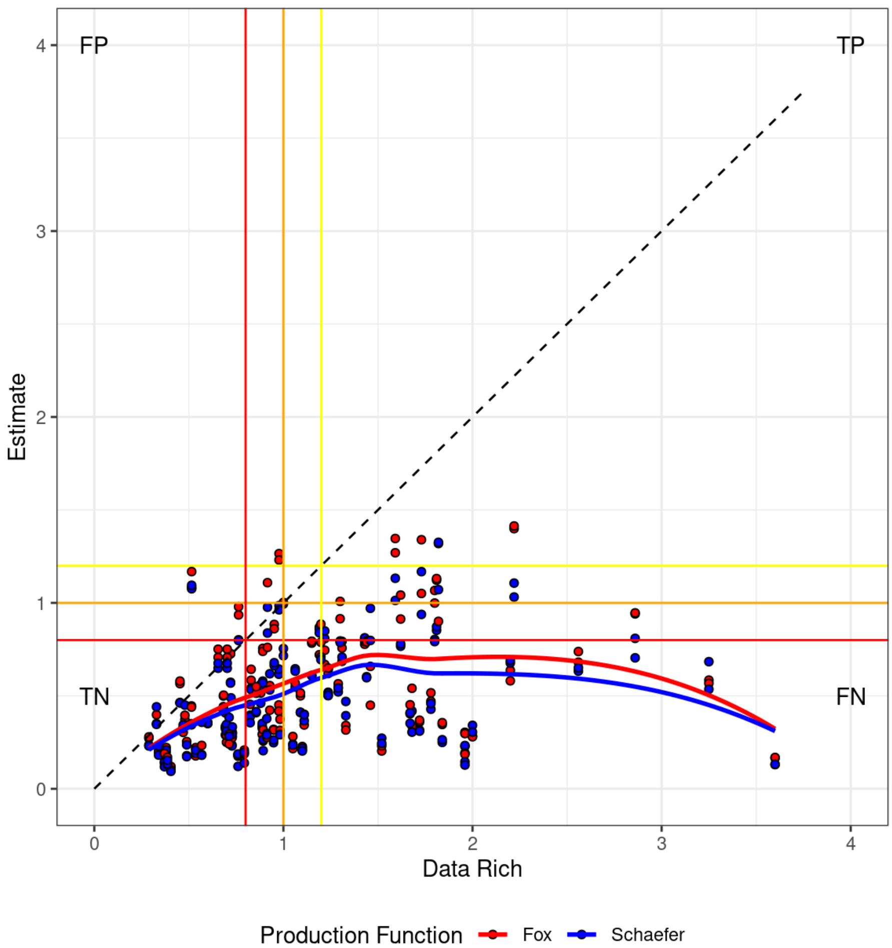


Figure 4: Comparisons of $B : B_{MSY}$, If the COM was unbiased y=x, if COM was biased but tunable then the smoother should be linear.

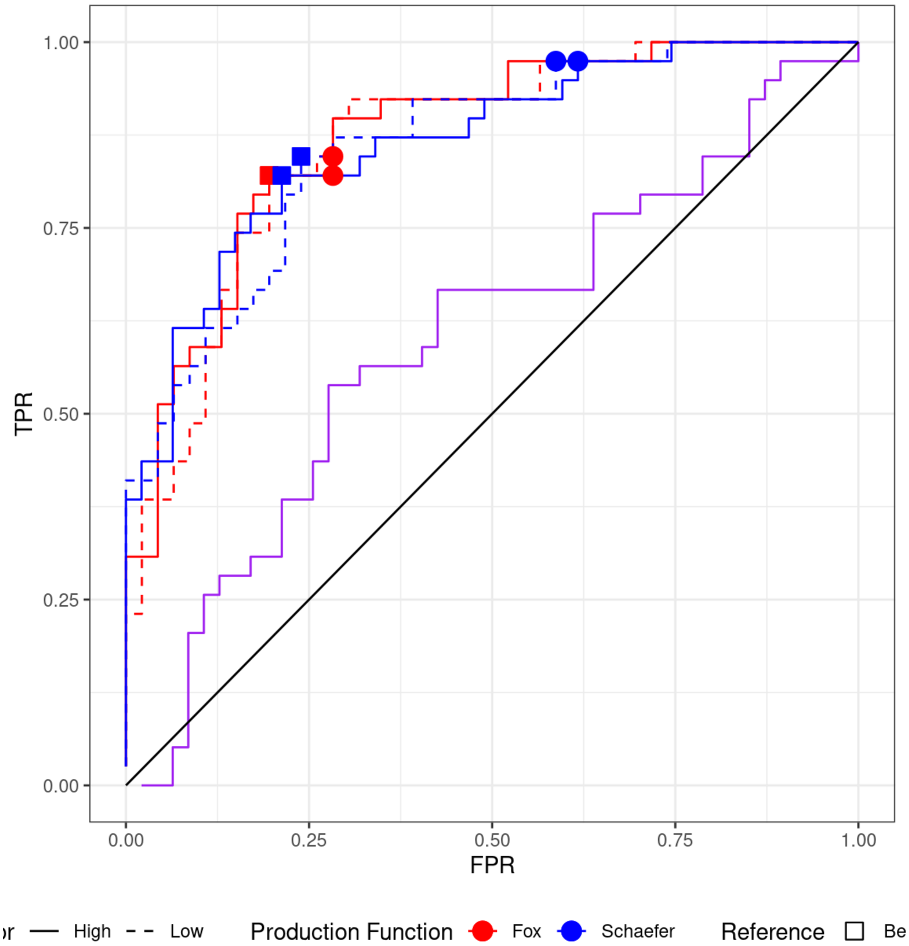


Figure 5: ROC curves for COMs.

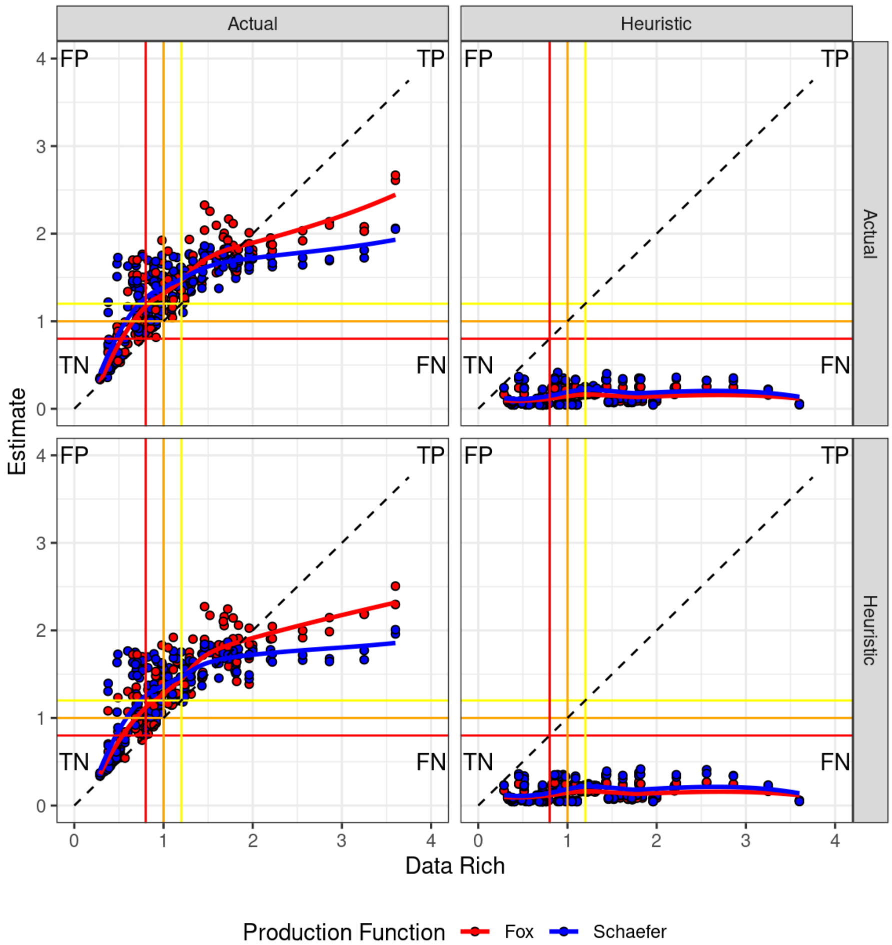


Figure 6: Comparisons of $B : B_{MSY}$, If the COM was unbiased y=x, if COM was biased but tunable then the smoother should be linear.

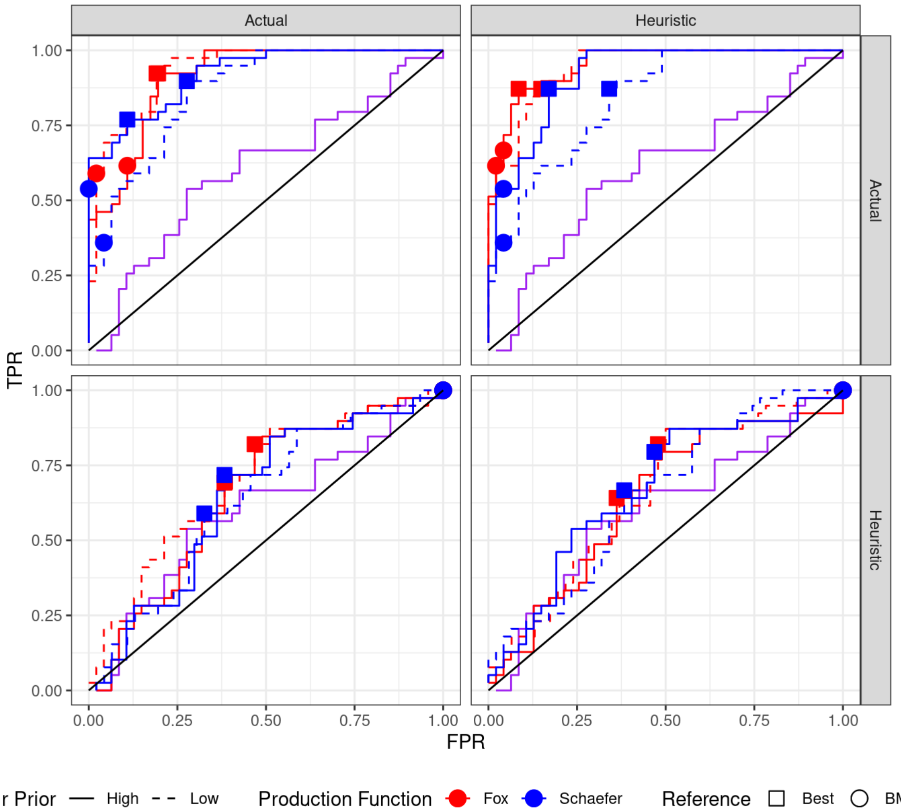


Figure 7: ROC curves for COMs.

## Supplementary Material