Machine learning for Tuna price prediction

Rodney Beard
Fisheries Economics Advisor
FFA

August 27, 2019

Some considerations

- Should the model be a causal model or a predictive model?
- Time series approaches (non-causal):
  - Univariate predictions
  - Extend this to a multivariate setting: ARIMA-X?
  - Previous work used ARIMA and ADL (autoregressive distributed lag)
  - ARIMA forecast from that work produces similar results

Prediction vs Causal inference

Policy changes impact the rent of the fishery in one of two ways:

1. Either policy increases rent directly and the magnitude of the increase depends on the price.
2. or Policy changes increase rent indirectly by increasing the price which in turn increases the rent.

Prediction vs Causal inference

- The latter implies a causal relationship between price and policy (policy causes prices to increase)
- Our ability to predict assumes that policy changes have no causal impact on prices (our model is not a causal model).
- We assume that the interest in predicting price is to evaluate the magnitude of a policy change's impact on rent which depends on the price level.
- The policy is therefore not relevant to the prediction task.

Prediction vs Causal inference

- A (pure) prediction problem arises when policy has no impact on price
- A causal inference problem arises when a policy change has a (causal) impact on price but no direct impact on rent.
- Prices would need to be predicted to evaluate the magnitude of the policy impact on rent, e.g reducing effort (TAE), may lead to an increase in rent or a decrease in rent, but the size of the impact will depend on what happens to the price (total revenue might increase or decrease).

Arguments based on: Kleinberg, Jon, et al. "Prediction policy problems." American Economic Review 105.5 (2015): 491-95.

What is the policy lever?

Candidate policy levers:

- TAE?
- Access fees?
- Something else?
- Answer impacts the estimation strategy

Policy: TAE or a Catch limit

Q1. about other ocean catches hints this might be the motivation behind the question.

Are catch and prices related?

Causal identification

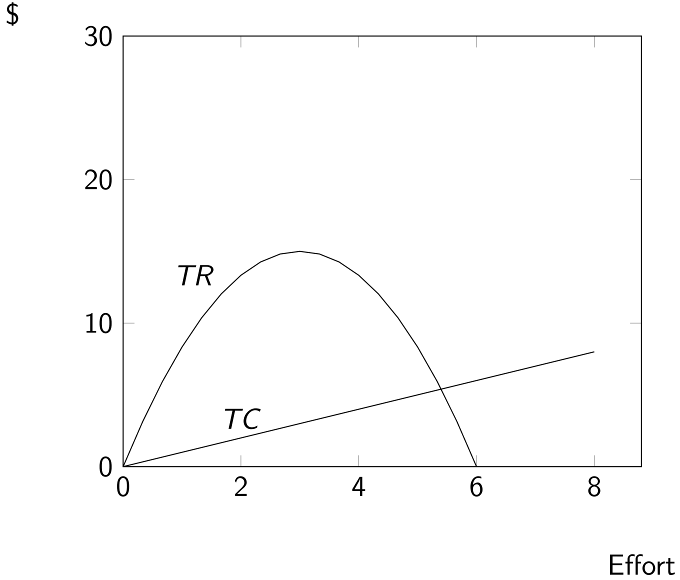
- To identify supply we would need to find a demand side instrumental variable
- Causal models are poor at prediction (they are designed for something else)
- However a properly identified predictive model would allow one to predict price using catch if a relationship exists. Caveats: Identification and such a relationship may not exist.
- We attempted to predict prices using catch with annual data without success, dropping catch from the model resulted in better performance in terms of prediction.

Are catch and prices related?

Supply and demand for Tuna

- Assuming that Skipjack/Purse-Seine fishery is a Schooling fishery, what does the supply curve for Skipjack look like?
- What does the demand curve for tuna look like? (demand studies indicate the demand is relatively elastic)

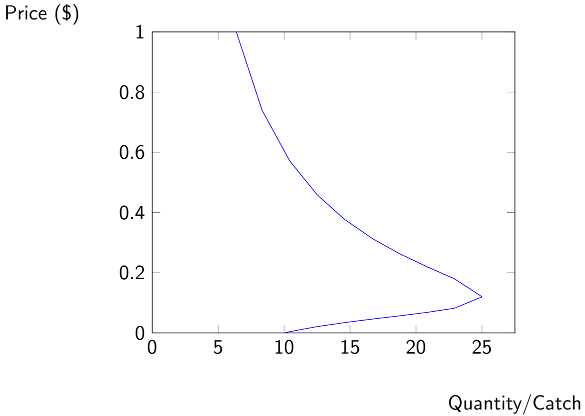
Hypothetical long-line fishery



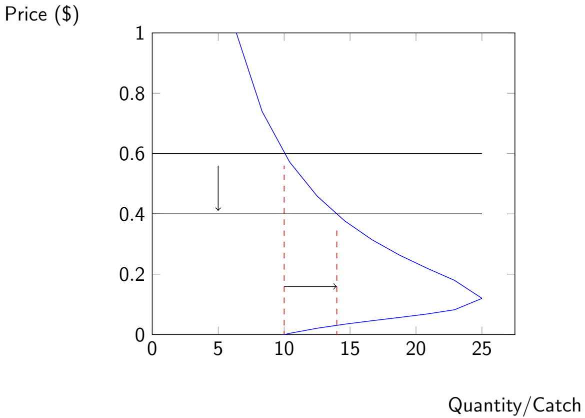
Supply curve

- Note: Typical supply curve for a (search) fishery is S-shaped, i.e. not hyperstable, this is true both for the open access fishery and the sole ownership fishery managed at MEY. It is not true if the fishery is managed to biological reference points.
- In the case of a schooling fishery, i.e. hyperstable case, it can be shown that the supply curve is perfectly inelastic above the shutdown point.

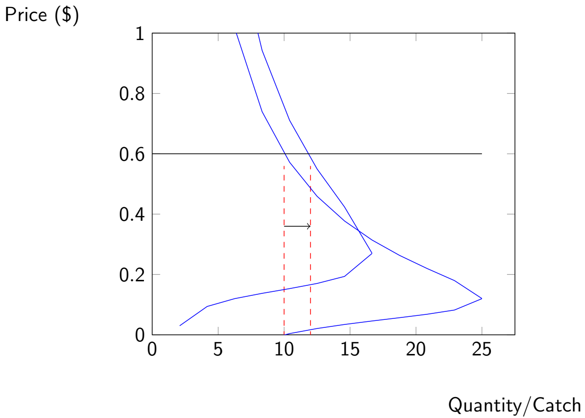
The "Textbook" fisheries supply curve (no hyperstability). Holds both for the open access fishery and the sole ownership fishery managed at MEY, e.g. zone based management. It does not hold if the fishery is managed to biological reference points.



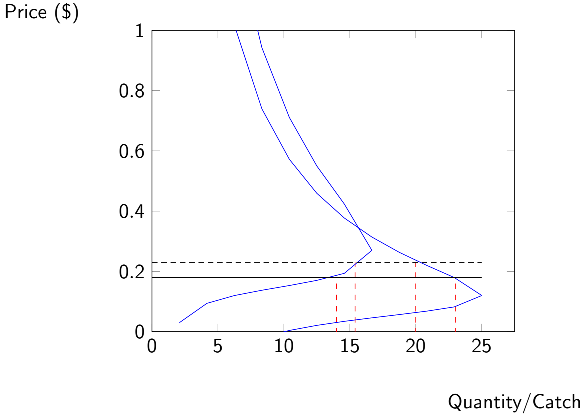
A fall in demand leads to a fall in price but an **increase** in the quantity demanded.



Increasing costs will rotate the supply curve and can lead to an expansion of the quantity demanded.



A demand shift can lead to the relationship between price and catch reversing depending on the supply situation.



Under these circumstances trying to use catch changes to predict price changes will fail

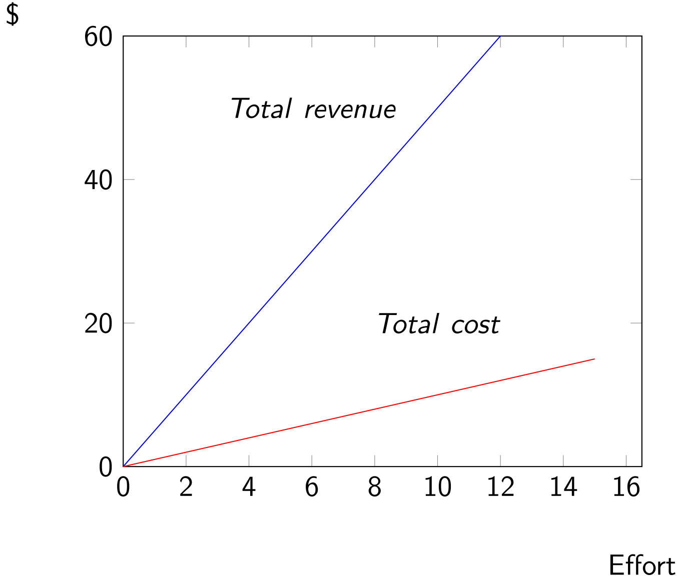
(Non-linear identification problem)

Consequences

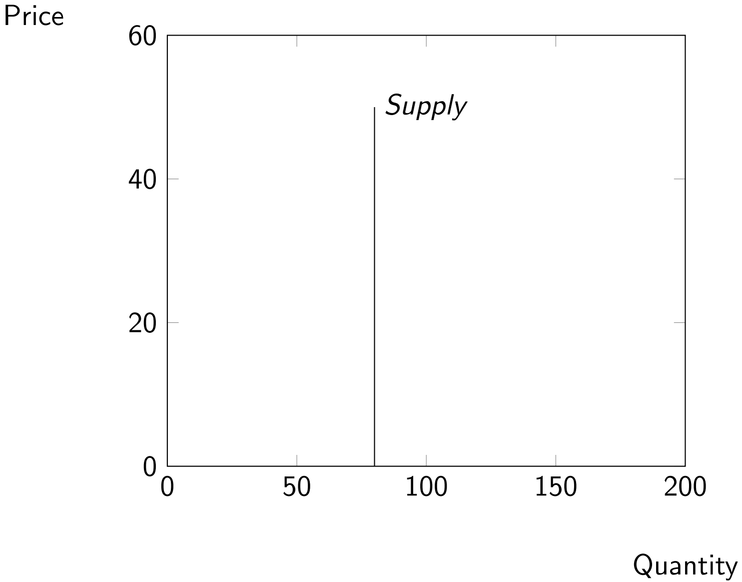
- A fall in supply leads to little to no increase in price depending on the elasticity of demand (elastic demand)
- An increase in supply leads to a fall or little to no change in price depending on the elasticity of demand (elastic demand)
- An increase in demand leads to an increase in price, but this might be associated with either an increase or decrease in the quantity demand (uncertain relationship between price and catch)
- A fall in demand leads to a reduction in price but this might be associated with an increase or decrease in the quantity demanded (uncertain relationship between price and catch)

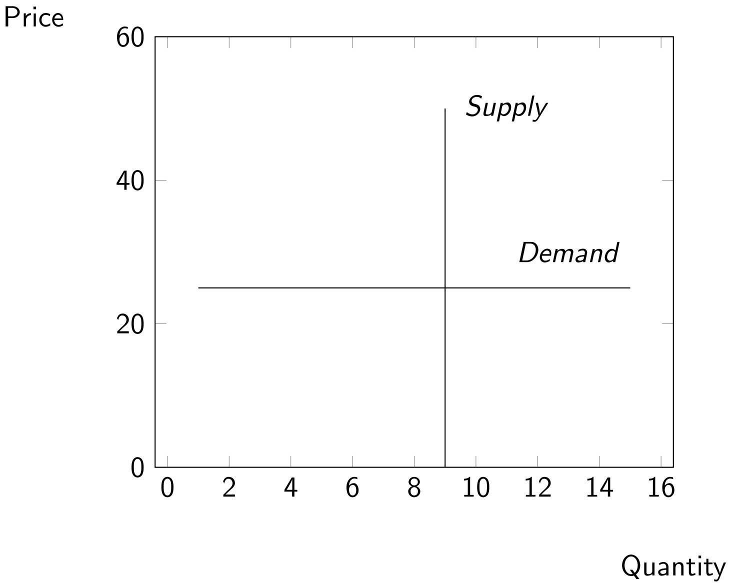
Result: Harder to identify a supply relationship than a demand relationship

Hypothetical purse-seine fishery

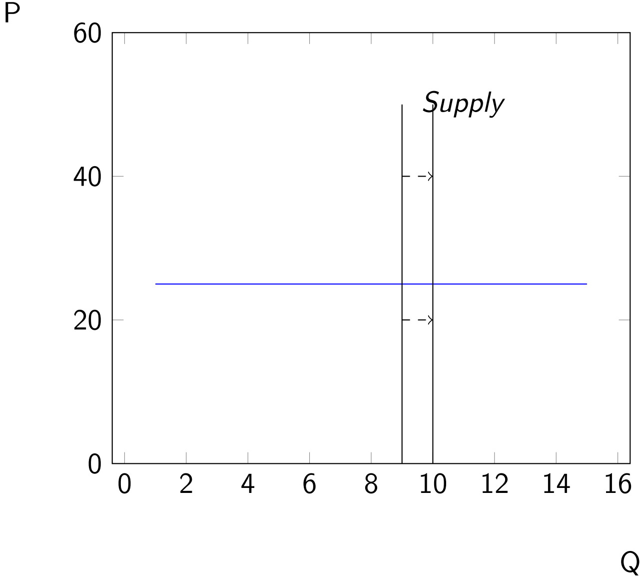


The fisheries supply curve (with hyperstability)




Price
60
40
20
0
Supply
Demand
0
2
4
6
8
10
12
14
16
Quantity

An increase in supply doesn't lead to a fall in the price



Putting this together

We would expect the following:

- Catch has no influence on price
- Price fluctuations are due to demand side factors and oceanographic factors other than catch (demand side influences were also identified in previous FFA study as the likely drivers of price fluctuations).
- This implies we have a pure prediction policy problem and can ignore causal effects (unless access fees are the policy lever).

We can therefore answer the first question in the negative, catch can't be used to predict skipjack prices.

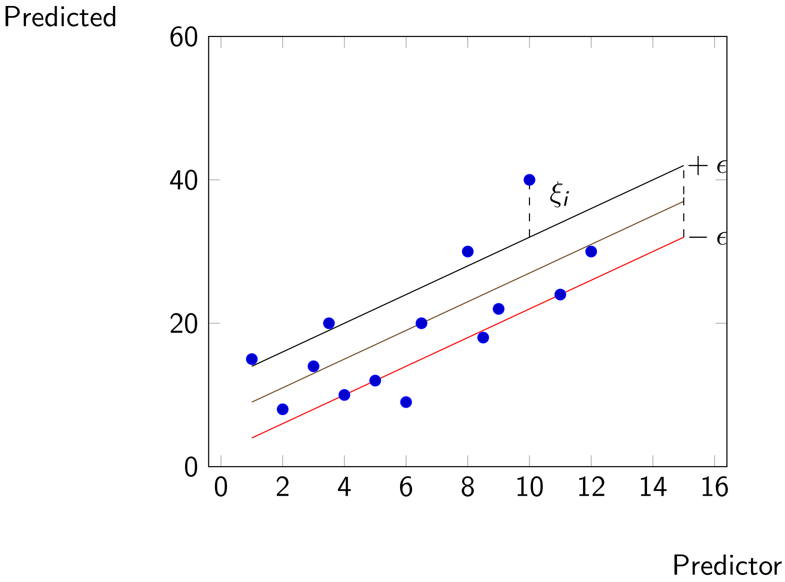
Price prediction

Predicting skipjack prices

- We looked at price prediction using ARIMA initially (gives good in sample fit and short range predictions are good but not so good for longer range predictions)
- We then looked at using ARIMA-X and including other variables however estimates weren't consistent with an autoregressive process.
- We also tried fitting a random forest model without success (poor out of sample performance)
- Finally, we tried support vector regression with some initial success which led us to pursue this approach (good out of sample performance)

Machine learning

- Support vector regression is a machine learning technique
- Non-parametric and distribution free
- Originally developed for image classification and pattern recognition
- Bias-Variance trade-off
- Traditionally econometrics aims to minimize bias whereas many machine learning methods trade off bias for low variance.


Predicted
60
40
20
0
0
2
4
6
8
10
12
14
16
ξ_i
$+\epsilon$
$-\epsilon$
Predictor

Support vector regression

Support vector regression is formulated as a quadratic programming problem (linear response function depicted here):

$$\min \tau(\mathbf{w}, \xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C\frac{1}{m}\sum_{i=1}^{m}(\xi_i + \xi_i^*)$$

$$\text{subject to } (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - y_i \leq \epsilon + \xi_i, i = 1, \ldots, m$$

$$y_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leq \epsilon + \xi_i^*, i = 1, \ldots, m$$

The constraints form a set of hyperplanes called support vectors, hence the name.

Bias-Variance tradeoff

$$\text{Mean-square error} = \text{Variance of Prediction} + \text{Bias}^2$$

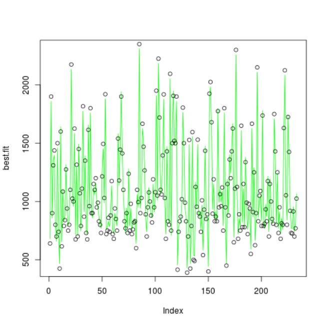
To get accurate prediction we want low variance (precision) and don't care much about bias.

Data

- Data is a mixed data-set consisting of pricing data collected monthly by FFA since 1984 and the Thai Union skipjack price series published by Thai Union since 2011.
- Data were split into training (known inputs and outputs), testing (known inputs only) and validation (unknown inputs and outputs, i.e. inputs were also predicted) data sets
- Features (other variables) were selected based on common sense, suggestions in the literature and their ability to predict out of sample on the test data set.
- The model was tuned based on out of sample predictions and 10000 fold cross validation
- Tuning involves choosing ϵ and an error ξ penalty weight C

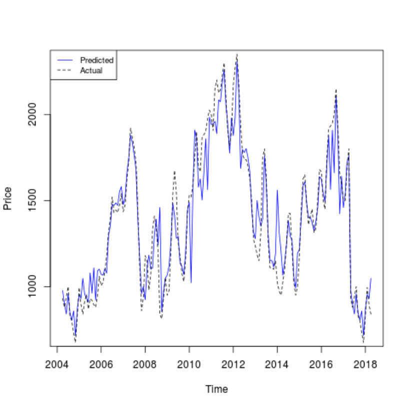
Training data

- Support vector machine was trained by choosing a penalty term of the errors and the radius of the "tube" around the data.
- Data are scaled and normalized before processing (unlike with econometric methods)
- Fit is a low priority because we are only interested in prediction



Test data

This shows predictions out of sample but using known input values (other variables)



Validation data

Other variables were predicted using ARIMA, Markov chain modelling, and the predictions fed into the support vector regression model (referred to as Model stacking)
Predictions for 2018 and 2019 were as follows:

Month 2018	Predicted	Actual
June	1764	1600
July	1439	1300
August	1408	1450
September	1558	1650
October	1493	1525
November	1561	1400
December	1551	1300

Note: prediction model did not account for US-China trade war.Chinese sales of Tuna into Bangkok were reported to have depressed prices.

How did we do?

- Average (Mean) Error = $78 error rate (average percent deviation from actual price) 5.8%
- Mean Absolute Error = $125 error rate (average absolute deviation from actual price) 8.9%
- Note: We are actually predicting 156 months ahead because of the way the data is split between training and holdout samples.
- To improve we need to model the other variables better: we need better feature engineering!

2019 Predictions

Month	Predicted	Actual	% error
Jan	1551	1280	21%
Feb	1526	1480	3%
Mar	1493	1600	6.68%
Apr	1514	1450	4.4%
May	1542	1200	28.5%
Jun	1533	1000	53.3%
Jul	1533	1200	27.75%
Aug	1371	?	
Sep	1371	?	
Oct	1533	?	
Nov	1533	?	
Dec	1533	?	

Some evidence that prices from Nov 2018 - Jan 2019 were impacted by US-China Trade War.
Ramadan May 6 - June 4
High inventories

Conclusion

- Long-range price prediction can be improved by drawing on additional information
- Machine learning predicts reasonably well out of sample compared with traditional time series methods (care needs to be taken, this is not obvious)
- ARIMA over smoothes: good for predicting average price level but not monthly prices
- SVR overfits (we try to avoid this using hold-out sampling)
- We will continue to refine monthly price prediction models
- We hope to develop a dashboard type reporting system (R Shiny dashboard) for members to access the latest price predictions