Addressing Onsite Sample Selection Biases in Discrete Choice Models: An Application of Propensity Score Based Weights

PAUL HINDSLEY
Assistant Professor
Department of Environmental Studies,
Eckerd College, St. Petersburg, Florida

Sample Selection Biases in Onsite Sampling

- Method Needed to Address Both:
  - Endogenous Stratification
  - Size-Biased Sampling (Avidity Bias)

Weights to Address Sample Selection Bias

$$\hat{L}\hat{L}_{ESS}(\beta) = \sum_{n=1}^{N} w(i_n) \ln \frac{P(i_n \mid z, s, \beta)}{\sum_{j \neq i} P(j \mid z, s, \beta)}$$

$$w(i_n) = \frac{Q(i_n)}{H(i_n)}$$

Weighted Exogenous Sample Maximum Likelihood (WESMLE)
Manski & Lerman (1977), Cosslett (1981)

$$\hat{L}\hat{L}_{ESS}(\beta) = \sum_{n=1}^{N} w(i_n, z, s) \ln \frac{P(i_n \mid z, s, \beta)}{\sum_{j \neq i} P(j \mid z, s, \beta)}$$

$$w(i_n) \equiv \frac{Q(i_n, z, s)}{H(i_n, z, s)} = \frac{\Pr(t=1 \mid i_n, z, s)}{1 - \Pr(t=1 \mid i_n, z, s)}$$

Propensity Score Weight Evaluation

Figure 1: A comparison of balance between the unweighted variables and the variables weighted using the propensity score.

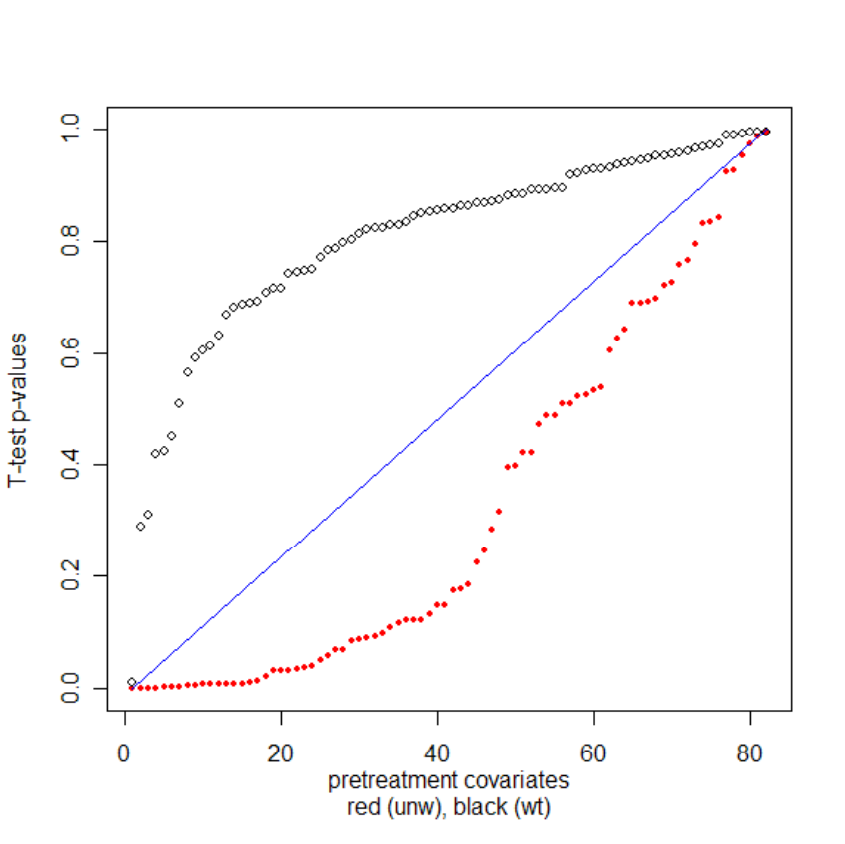
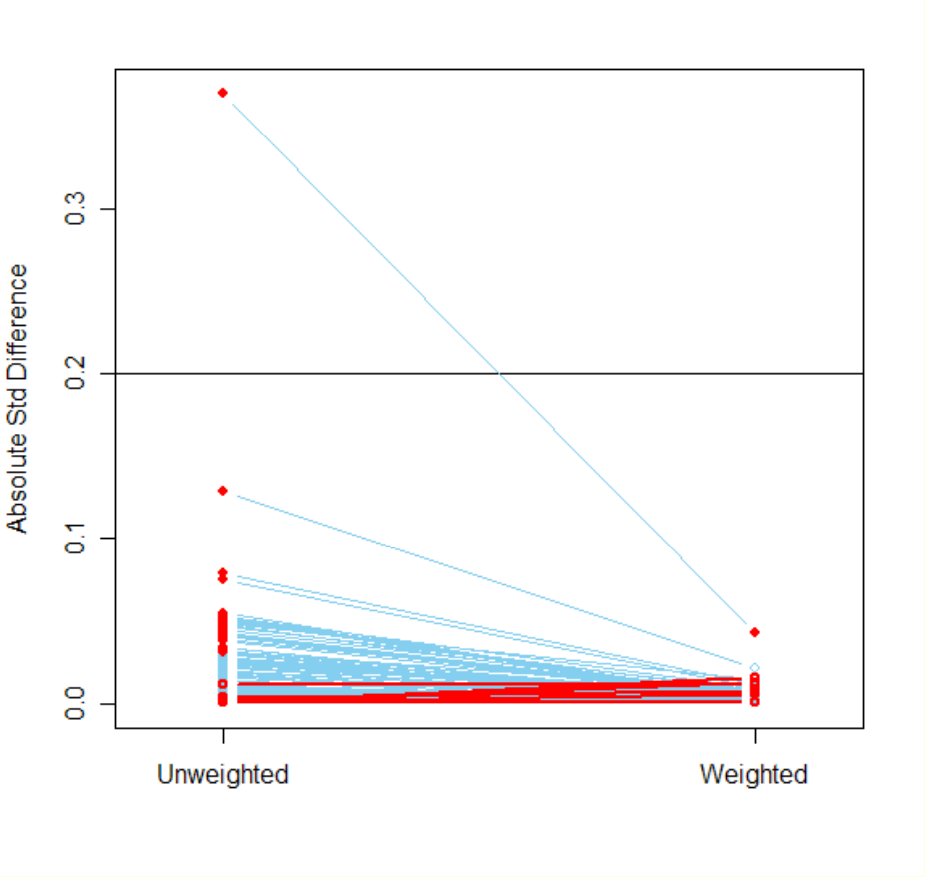
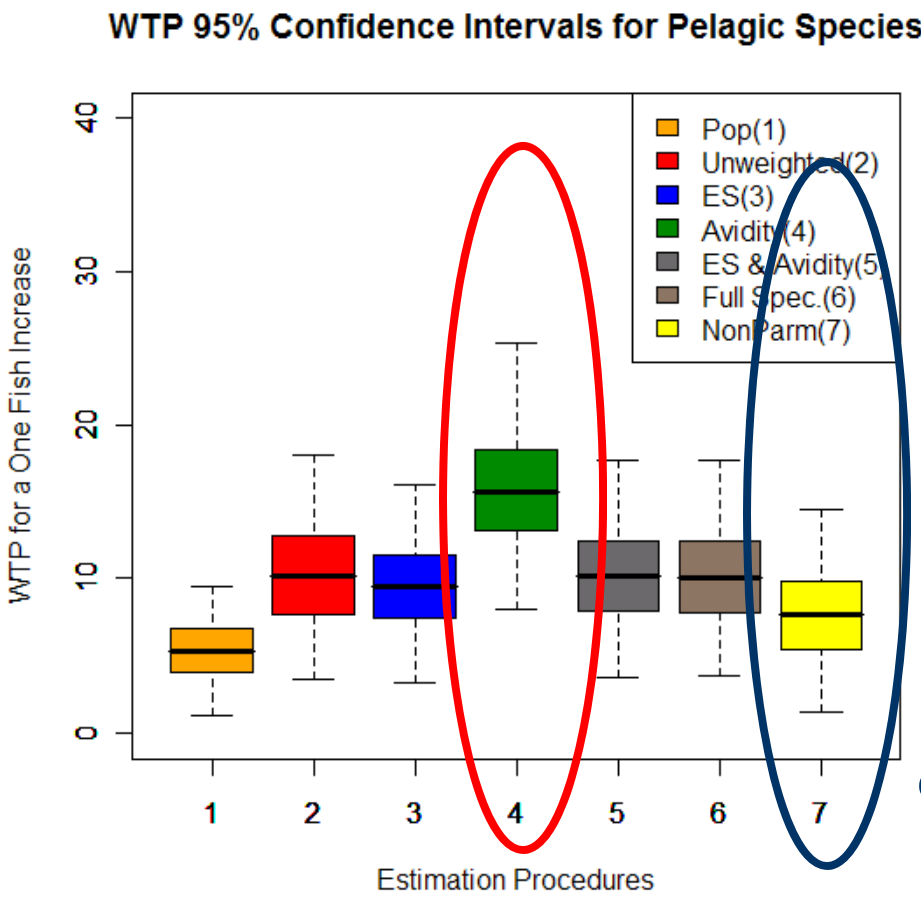


Figure 2: A comparison of the absolute standard differences between the unweighted variables and the variables weighted using the propensity score.



Pelagic Species



	WTP	Difference (Percent	Average Difference All Species
Population	$5.21		
Unweighted	$10.30	0.98	0.45
Site Only	$9.37	0.80	0.39
Avidity Only	$15.51	1.98	0.89
Site & Avidity	$10.17	0.95	0.40
All Variables	$10.02	0.92	0.40
All Variables (Flexible Form)	$7.46	0.43	0.26

Conclusions

- Balanced WESMLE can reduce multiple forms of bias in MNL models when auxiliary information is available
- Model and Variable Selection Play an important role in bias reduction
- Balanced WESMLE not appropriate for Nested Models or Models utilizing Mixing Distributions
- Additional research needed to address these biases while also addressing shortcomings of MNL model