

2/11/2016

Stu --

Good question. I think it is easiest to answer this by describing the model(s) we will estimate. The crux of the matter is that the unit of observation in our analyses is an individual respondent. Seen in this way, our sample size is sufficient for all the models we propose.

Take the cancellation question for example. After viewing a wind farm offshore in a photo simulation, respondents are asked if it would cause them to cancel a trip. Each respondent answers three questions in this format at three different offshore distances. The distances shown to a respondent are randomly drawn from the set 2.5, 5, 7.5, 10, 12.5, 15, and 20 miles offshore.

This gives us dichotomous response data for cancellation – 1 for cancel; 0 for not cancel. We anticipate having roughly 1925 beachgoers in our sample, so that gives us 5775 (3 * 1925) responses to the cancellation question. Since the distances are randomly assigned, at each distance we have 825 responses (5775/3).

Because we are working with dichotomous response data, we intend to estimate a binary choice model (logit or probit regression). A simple linear version of the model is

$$pr(cancel)_{ij} = f(\beta d_{ij} + \alpha b_{ij})$$

$pr(cancel)_{ij}$ = probability respondent i cancels trip to beach j ,

d_{ij} = vector of dummy variables for each of the 7 distances ($d_{ij}^{2.5} = 1$ if wind farm is at 2.5 miles offshore for respondent i at beach j , 0 if not; $d_{ij}^5 = 1$ if wind farm is at 5 miles offshore for respondent i at beach j and 0 if not; etc.),

b_{ij} = vector of beach characteristics for beach j visited by respondent i ,

$f(\cdot)$ is a common logit or probit form, which I will not write out here,

β = a vector of parameters at the 7 distances: $\beta_{2.5}, \beta_5, \beta_{7.5}, \beta_{10}, \beta_{12.5}, \beta_{15}, \beta_{20}$, and

α = a vector of parameters for the beach characteristics

Estimates of the parameters β tell us how people respond to wind farms at different distances offshore. We expect $\beta_{2.5}$, for example, to be the largest of the coefficients indicating that the probability of cancelling is highest at the closest distance. As the wind farms get further offshore we expect the probability to decline and hence the

parameters β to fall as distance increases. We will also try a simple continuous measure for distance – entering distance as an explanatory variable instead of a series of dummies.

The estimates for α tell us how different beaches may get different responses to wind farms. For example, some argue that wind farms on developed beaches will have less impact than they would on more natural (park) beaches. Also, there may be different responses in different states. These propositions can be explored in the estimation results for parameters in α . We will also explore individual characteristics in the model, such as income, frequency of visiting the beach, etc. To simplify here, I have exclude these arguments.

With the model above we can predict the impacts on any beach in our model (beaches from MA to SC), by plugging in the relevant beach characteristics and offshore distance. This will be BOEM’s model for analysis of wind farms in different locations.

What is important for your comment is that the number of observation we have is more than sufficient to estimate this model and pick up a signal on the distance effects on cancellation. With a sample size well above 5000 and never more than a twenty parameters to estimate, we easily have >5000 degrees of freedom. We have estimated models like this with much less data and found significance over distance variables.¹

Several other models will be estimated in similar fashion. For example, the response to whether or not wind farms make your experience better or worse will be estimated using an ordered logit model since the response format has five outcomes which are ordered – better, somewhat better, no effect, somewhat worse, worse. Also, there is a question about whether or not a “special trip” would be taken (more or less a curiosity trip) to see a wind farm. All of the models will be estimated in much the same way as the cancellation model is described above and, for the same reason, there is sufficient data.

I hope that helps explain the study design and gives you reasoning why we believe our sample size is sufficient to answer the questions we have posed.

-- George Parsons, University of Delaware

¹ We recognize that the three response values from each person for cancellation should not be treated as independent of one another. This will be dealt with in the econometrics using fixed-effects. Still, the number of observations is more than enough for the analysis.