LECTURE / DISCUSSION

Forecasting with Logit

Simulation with an Estimated Logit Model

Issue

- Given an estimated model, how can we use it to simulate impacts of changes in attribute levels on **aggregate** market shares?

Example Applications

- What is the expected change in the market share of gas heating due to a price reduction of 10%?

- What is the forecasted market share of gas heating under projections of changing relative energy prices, operating efficiencies, and installation costs?

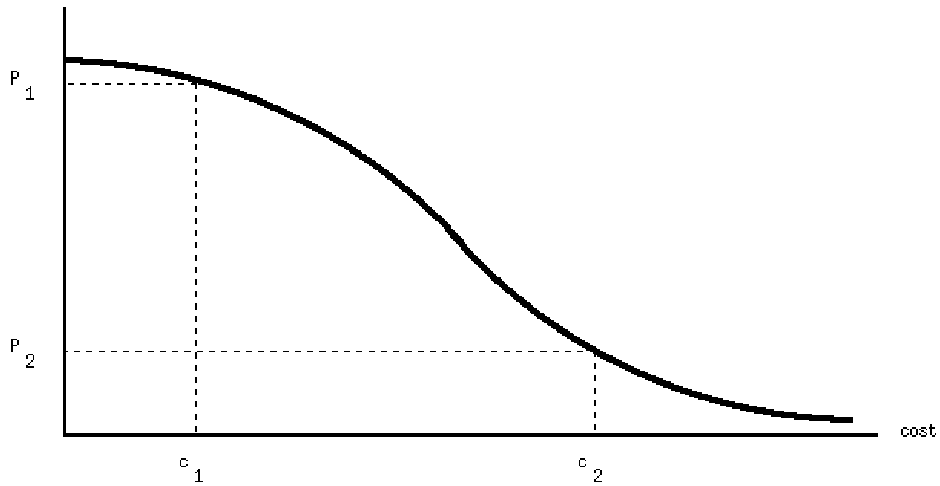
- What is the forecasted penetration of a new technology?

Contrast to Hypothesis Testing

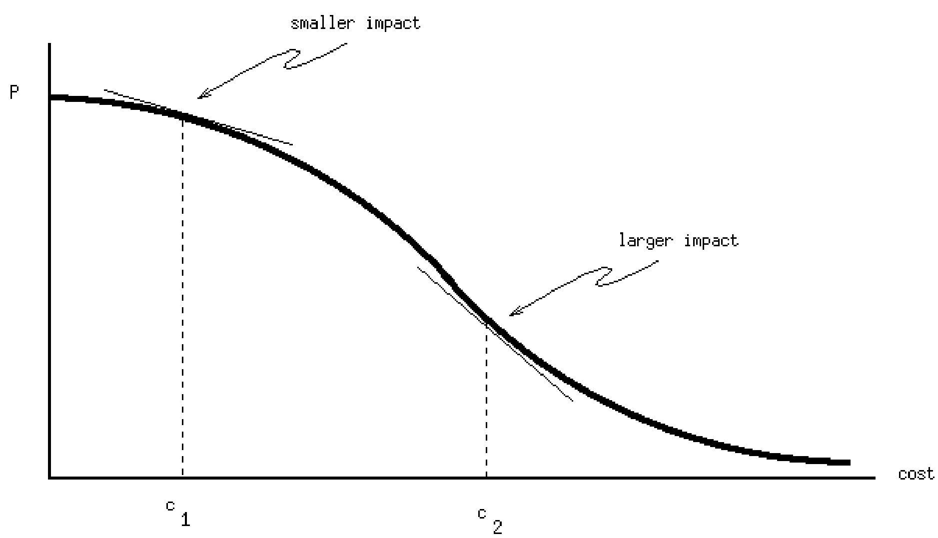
- Is there a significant relation between operating costs and installation rates of gas space heating?

The Problem

1. Probability is different for each person.



2. Impact of a change is different for each person.



Forecasting with Linear Regression

With linear-in-variables regression, forecasting can be performed using averages of explanatory variables.

$$y_n = \beta x_n + \varepsilon_n$$

Average dependent variable is average explanatory variables times their coefficients.

$$\bar{y} = \beta \bar{x}$$

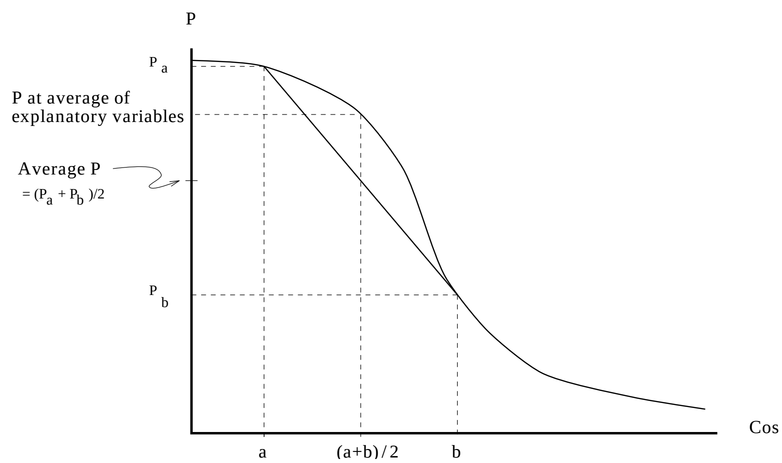
Impact of a change in explanatory variables is change in average explanatory variables times their coefficients.

$$\bar{y}_1 - \bar{y}_2 = \beta(\bar{x}_1 - \bar{x}_2)$$

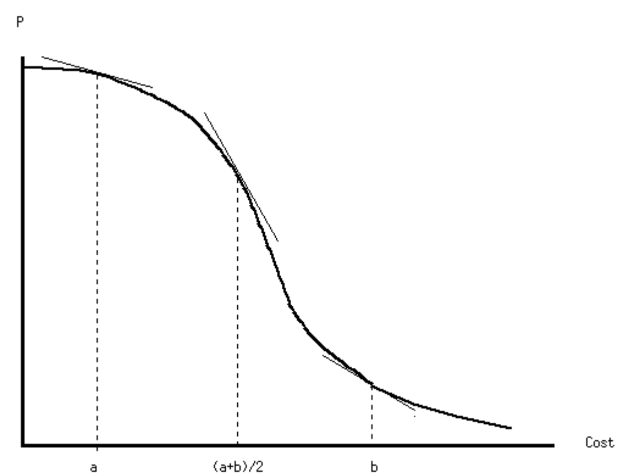
Because logit model is nonlinear, forecasting ***cannot*** be performed using only the averages of explanatory variables.

Aggregation Bias with Nonlinear Logit Models

1. Average probability is not equal to the probability at the average of explanatory variables.

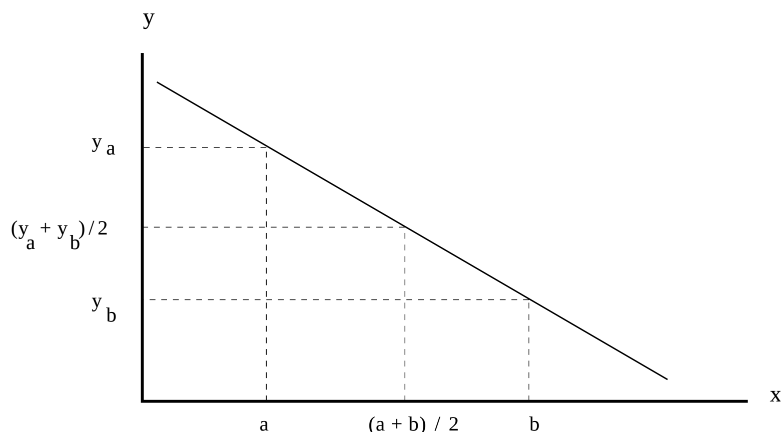


2. Average impact of a change is not equal to the impact calculated at the average of the explanatory variables.

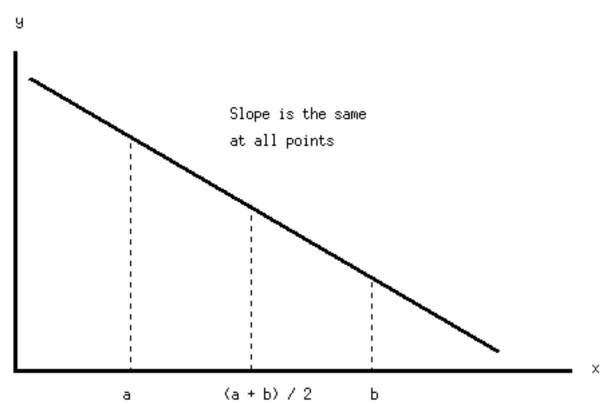


Contrast with linear-in-variables regression:

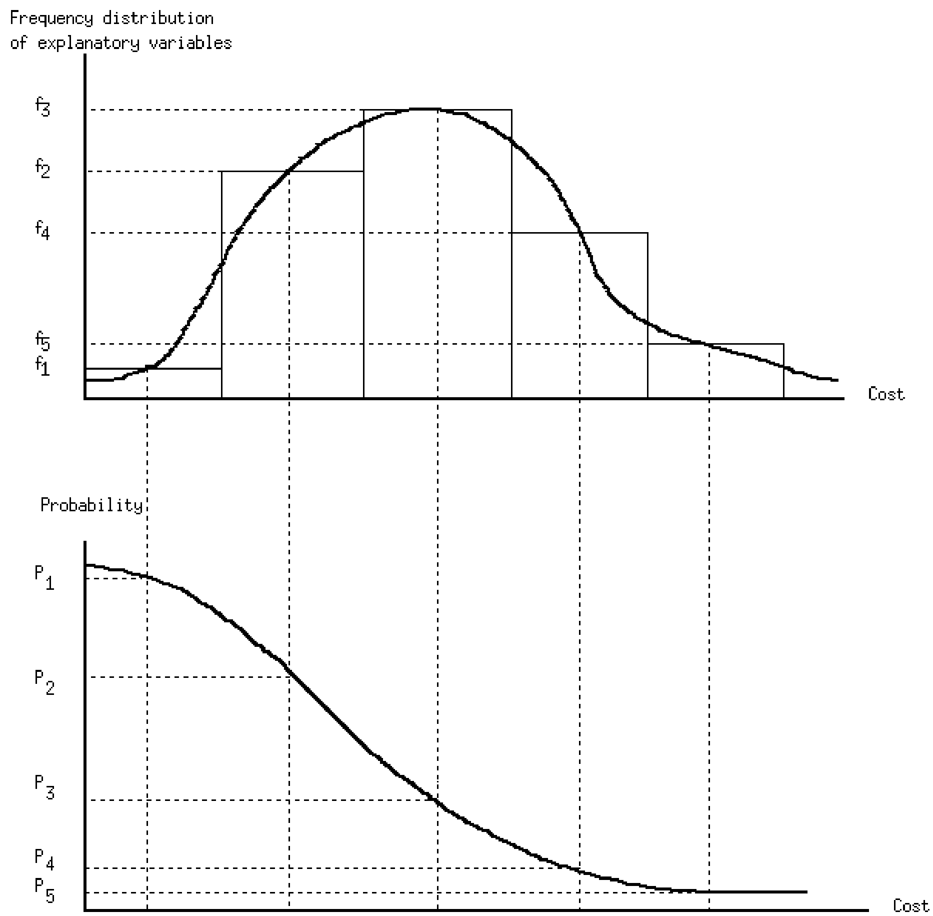
1. Average dependent variable equals the dependent variable evaluated at the average of the explanatory variables.



2. Average impact equals the impact evaluated at average explanatory variables.



Need to represent distribution of explanatory variables in the population:



$$\overline{P}_i = \sum_j P_{ij} \times f_j$$

Methods of Representing Joint Distribution of "Choice Settings" and Customer Characteristics

1. Average individual

   Generally risks aggregation bias, but it works in "degenerate" cases.

2. Analytical expression

   Generally infeasible for any application except simple models.

3. Market segmentation

   Partitions population into exhaustive set of all levels of all characteristics and computes frequencies in each segment. Number of segments may become unmanageable.

4. Sample enumeration

   Represents population through sample. Subject to sampling error. Straightforward for "static" simulations.

Simulation by Market Segmentation

1. Segment population into M groups representing distinct values of explanatory variables.

2. Determine proportion of decisionmakers in each segment: $f(x_m)$.

3. Calculate choice probabilities for each segment: $P_i(x = x_m)$.

4. Average probability in population is:

$$\overline{P}_i = \sum_{m=1}^{M} P_i(x = x_m)\, f(x_m)$$

Example of Market Segmentation

Mode choice with time as only explanatory variable:

$$U_c = \beta T_c + \varepsilon_c$$

$$U_b = \beta T_b + \varepsilon_b$$

Car Time	Bus Time			
	< 15 min.	15-30 min.	30-60 min.	> 60 min.
< 15 min.	f_1	f_2	f_3	f_4
15-30 min.	f_5	f_6	f_7	f_8
30-60 min.	f_9	f_{10}	f_{11}	f_{12}
> 60 min.	f_{13}	f_{14}	f_{15}	f_{16}

Frequency for 16 segments

Calculate the fraction of decisionmakers in each segment, f_m.

Calculate probability for each segment. Take weighted average of probabilities.

Simulation by Sample Enumeration

1. Represent population by means of a sample.

2. Estimate the proportion choosing *i* as

$$\overline{P}_i = \frac{1}{N}\sum_n P_{in}$$

   where *n* indexes the observations in the sample, n = 1,...,N.

3. Change explanatory variables to represent scenario.

4. Calculate new P^*_{in} for each *n*. The new estimated proportion for choosing *i* is

$$\overline{P}^*_i = \frac{1}{N}\sum_n P^*_{in}$$

5. The impact of scenario is

$$\overline{P}^*_i - \overline{P}_i$$

6. For stratified sample, probabilities would be weighted by inverse sampling weights.

Difficulties with market segmentation approach:

1. Aggregation bias within each segment, unless segmentation is "exhaustive ".

2. Number of segments can be very large to adequately represent range of explanatory variables.

3. Data on the proportion of decisionmakers in each segment are often not available or hard to obtain.

Note:

- Sample enumeration is an approximation to the complete market segmentation approach. The approximation becomes better as sample size increases.

Forecasting over Time

Sample enumeration:

Adjust sample such that it "looks like" a sample drawn in the future year.

Change:
- decisionmaker's characteristics
- weights for each sampled decisionmaker

Market segmentation:

Adjust proportions in each segment to represent changes in population over time.

WORKSHOP ON SIMULATION

1. Run model 2 from the previous workshop. This time, calculate the average probabilities (that is, the predicted shares) explicitly rather than using the PROB subop. This will help in steps 2 and 3 below. The SST command file **sim.cmd** will do this. Be sure you understand each line in this file.

2. Now suppose the energy utility is considering whether to offer rebates on heat pumps. The energy utility wants to predict the effect of the rebates on the heating system choice of its customers. The rebates will be set at 10% of the installation cost; the new installation cost for heat pumps is therefore

   $$nic5 = .90 * ic5$$

   Using the estimated coefficients from step 1, calculate new probabilities and predicted shares using **nic5** instead of **ic5**.

3. Suppose a new technology is developed that provides more efficient central electric heating. The new technology costs $200 more than the central electric system that we have specified as our alternative 3; however, it saves 25% of the electricity, such that its operating cost is 75% of the operating cost of our alternative 3.

   We want to estimate the potential market penetration of this technology in a market with the ***original*** five alternatives in Question 1.

   Note that there are now six alternatives. Calculate probabilities and expected shares for the six alternatives, using the estimated coefficients from step 1.

   What is the expected penetration (i.e., share) for the new technology? From which of the original five systems does the new technology draw the most customers?

DISCUSSION OF WORKSHOP RESULTS

Impact of Reduced Heat Pump Costs

	Predicted Shares		Change in Shares
	Original Costs	Heat Pump Cost Down by 10%	
Gas central	.517	.482	-.035
Gas room	.240	.224	-.016
Electric central	.104	.097	-.007
Electric room	.051	.048	-.003
Heat pump	.087	.149	.062

Elasticity of Heat Pump Market Share

Aggregate arc elasticity:

$$E = \frac{\Delta MS}{\Delta X} \times \frac{X}{\overline{MS}}$$

$$= \frac{0.062}{-104.6} \times \frac{1046}{.118} = -5.25$$

Point elasticity:

$$\varepsilon = \frac{\partial P}{\partial X} \cdot \frac{X}{P}$$

$$= \beta(P - P^2) \cdot \frac{X}{P} = \beta X(1 - P)$$

Potential Penetration of New Technology

	Predicted Shares		Change in Shares
	Without New Technology	With New Technology	
Gas central	.517	.493	-.024
Gas room	.240	.229	-.011
Electric central	.104	.098	-.006
Electric room	.051	.049	-.002
Heat pump	.087	.083	-.004
New electric central	---	.048	+.048
	1.00	1.00	0.0