# Stated probabilistic choice models

Stated probabilistic choice models

- Now, we observe a probability of choosing each alternative

## Stated probabilistic choice models

- Now, we observe a probability of choosing each alternative
- What does this information represent? It represents the person's uncertainty

Stated probabilistic choice models

- Now, we observe a probability of choosing each alternative
- What does this information represent? It represents the person's uncertainty
- Specifically, it represents resolvable uncertainty:

## Stated probabilistic choice models

- Now, we observe a probability of choosing each alternative
- What does this information represent? It represents the person's uncertainty
- Specifically, it represents resolvable uncertainty:
  - uncertainty about unspecified attributes or states of the world in which choices will ultimately be made

Example of resolvable uncertainty:

Example of resolvable uncertainty:

- Q. "Will you go for Mexican or Thai food on Friday night?"

Example of resolvable uncertainty:

- Q. "Will you go for Mexican or Thai food on Friday night?"
- A. "Well, I'm not sure what mood I'll be in that day, but probably Mexican"

Example of resolvable uncertainty:

- Q. “Will you go for Mexican or Thai food on Friday night?”
- A. “Well, I’m not sure what mood I’ll be in that day, but probably Mexican”
- Q. “What do you mean by ‘probably’?”

Example of resolvable uncertainty:

- Q. “Will you go for Mexican or Thai food on Friday night?”
- A. “Well, I’m not sure what mood I’ll be in that day, but probably Mexican”
- Q. “What do you mean by ‘probably’?”
- A. “A 78% chance”

Example of resolvable uncertainty:

- Q. "Will you go for Mexican or Thai food on Friday night?"
- A. "Well, I'm not sure what mood I'll be in that day, but probably Mexican"
- Q. "What do you mean by 'probably'?"
- A. "A 78% chance"

If the person thinks there is no such uncertainty, then they can report $p = 0$ or $p = 1$

## Estimating probabilistic choice models

Estimating probabilistic choice models

How do we proceed with estimation when $y$ is itself a probability (not $0/1$)?

Estimating probabilistic choice models

How do we proceed with estimation when $y$ is itself a probability (not $0/1$)?

We invert the logit formula

Estimating probabilistic choice models

How do we proceed with estimation when $y$ is itself a probability (not $0/1$)?

We invert the logit formula

Consider the binomial logit as an example:

$$P_{i1} = \frac{\exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}{1 + \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}$$

Start with the logit probability:

$$P_{i1} = \frac{\exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}{1 + \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}$$

Start with the logit probability:

$$P_{i1} = \frac{\exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}{1 + \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}$$

And the complement:

$$1 - P_{i1} = \frac{1}{1 + \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}$$

Start with the logit probability:

$$P_{i1} = \frac{\exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}{1 + \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}$$

And the complement:

$$1 - P_{i1} = \frac{1}{1 + \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}$$

Take the ratio (denominators cancel):

$$\frac{P_{i1}}{1 - P_{i1}} = \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)$$

Start with the logit probability:

$$P_{i1} = \frac{\exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}{1 + \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}$$

And the complement:

$$1 - P_{i1} = \frac{1}{1 + \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}$$

Take the ratio (denominators cancel):

$$\frac{P_{i1}}{1 - P_{i1}} = \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)$$

Take logs:

$$\log\left(\frac{P_{i1}}{1 - P_{i1}}\right) = \left(Z_{i1} - Z_{i2}\right)\gamma$$

Start with the logit probability:

$$P_{i1} = \frac{\exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}{1 + \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}$$

And the complement:

$$1 - P_{i1} = \frac{1}{1 + \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)}$$

Take the ratio (denominators cancel):

$$\frac{P_{i1}}{1 - P_{i1}} = \exp\left(\left(Z_{i1} - Z_{i2}\right)\gamma\right)$$

Take logs:

$$\log\left(\frac{P_{i1}}{1 - P_{i1}}\right) = \left(Z_{i1} - Z_{i2}\right)\gamma$$

Now we're in a world where we can (in principle) use OLS to estimate $\gamma$!

Stated probabilities provide more information than discrete choice or rank ordering

Stated probabilities provide more information than discrete choice or rank ordering

Consider the following responses for an individual:

| Option | Discrete Choice | Rank Ordering | Stated Prob 1 | Stated Prob 2 |
|---|---|---|---|---|
| A | 0 | 2 | 0.01 | 0.49 |
| B | 1 | 1 | 0.99 | 0.51 |
| C | 0 | 3 | 0.00 | 0.00 |

- The preference ordering is the same in each column

- The preference ordering is the same in each column
- But the implied preference intensity is much different in the last two columns

- The preference ordering is the same in each column
- But the implied preference intensity is much different in the last two columns
- A (0,1,0) discrete choice response corresponds to a (0,100,0) probability response

# Measurement error

## Measurement error

- One concern with stated probabilities is measurement error

## Measurement error

- One concern with stated probabilities is measurement error
- Rather than report 99.5%, someone may just write 100%

## Measurement error

- One concern with stated probabilities is measurement error
- Rather than report 99.5%, someone may just write 100%
- But if $p = 0$ or $p = 1$, $\log\left(\frac{P_{i1}}{1-P_{i1}}\right)$ is undefined!

## Measurement error

- One concern with stated probabilities is measurement error
- Rather than report 99.5%, someone may just write 100%
- But if $p = 0$ or $p = 1$, $\log\left(\frac{P_{i1}}{1-P_{i1}}\right)$ is undefined!
- In this case, we have to recode 0s or 1s to be small values (e.g. .001, .999)

Then, to avoid contamination, we need to use LAD instead of OLS. New equation:

$$\log\left(\frac{\tilde{P}_{i1}}{1-\tilde{P}_{i1}}\right) = (Z_{i1} - Z_{i2})\,\gamma + \eta_{i1}$$

Then, to avoid contamination, we need to use LAD instead of OLS. New equation:

$$\log\left(\frac{\tilde{P}_{i1}}{1-\tilde{P}_{i1}}\right) = (Z_{i1} - Z_{i2})\,\gamma + \eta_{i1}$$

where $\tilde{P}$ is the recoded probability and $\eta_{i1}$ is the difference in measurement errors

## Estimation with $J > 2$

## Estimation with $J > 2$

- The above is the equation for a 2-option choice set

## Estimation with $J > 2$

- The above is the equation for a 2-option choice set
- With more than 2 options, we have more observations per scenario

## Estimation with $J > 2$

- The above is the equation for a 2-option choice set
- With more than 2 options, we have more observations per scenario

If $J = 3$, we have:

$$\log\left(\frac{\tilde{P}_{i1}}{\tilde{P}_{i2}}\right) = (Z_{i1} - Z_{i2})\,\gamma + \eta_{i1}$$
$$\log\left(\frac{\tilde{P}_{i3}}{\tilde{P}_{i2}}\right) = (Z_{i3} - Z_{i2})\,\gamma + \eta_{i3}$$

If we see $N$ people each making $T$ choices, then our data has $NT(J-1)$ rows:

If we see $N$ people each making $T$ choices, then our data has $NT(J-1)$ rows:

| ID | Scenario | Alternative | Probability | $Z$ | $Z_j - Z_B$ | $\log(P_j/P_B)$ |
|---|---|---|---|---|---|---|
| 1 | 1 | A | 0.45 | 2.3 | 0.5 | 1.099 |
| 1 | 1 | B | 0.15 | 1.8 | 0.0 | — |
| 1 | 1 | C | 0.40 | 3.1 | 1.3 | 0.981 |
| 1 | 2 | A | 0.25 | 2.5 | -0.4 | -0.875 |
| 1 | 2 | B | 0.60 | 2.9 | 0.0 | — |
| 1 | 2 | C | 0.15 | 1.7 | -1.2 | -1.386 |
| 2 | 1 | A | 0.10 | 3.2 | -0.9 | -1.946 |
| 2 | 1 | B | 0.70 | 4.1 | 0.0 | — |
| 2 | 1 | C | 0.20 | 2.8 | -1.3 | -1.253 |

If we see $N$ people each making $T$ choices, then our data has $NT(J-1)$ rows:

| ID | Scenario | Alternative | Probability | $Z$ | $Z_j - Z_B$ | $\log(P_j/P_B)$ |
|---|---|---|---|---|---|---|
| 1 | 1 | A | 0.45 | 2.3 | 0.5 | 1.099 |
| 1 | 1 | B | 0.15 | 1.8 | 0.0 | — |
| 1 | 1 | C | 0.40 | 3.1 | 1.3 | 0.981 |
| 1 | 2 | A | 0.25 | 2.5 | -0.4 | -0.875 |
| 1 | 2 | B | 0.60 | 2.9 | 0.0 | — |
| 1 | 2 | C | 0.15 | 1.7 | -1.2 | -1.386 |
| 2 | 1 | A | 0.10 | 3.2 | -0.9 | -1.946 |
| 2 | 1 | B | 0.70 | 4.1 | 0.0 | — |
| 2 | 1 | C | 0.20 | 2.8 | -1.3 | -1.253 |

Drop the rows for alternative B before estimating LAD regression models