Utility maximization implies

$$u_{ij} + \epsilon_{ij} > u_{ik} + \epsilon_{ik} \text{ for all } k \neq j \tag{1}$$

Utility maximization implies

$$u_{ij} + \epsilon_{ij} > u_{ik} + \epsilon_{ik} \text{ for all } k \neq j \tag{1}$$

With the ϵ's unobserved, the probability of i making choice j is given by:

$$\begin{aligned} P_{ij} &= \Pr(u_{ij} + \epsilon_{ij} > u_{ij'} + \epsilon_{ij'} \; \forall \; j' \neq j) \\ &= \Pr(\epsilon_{ij'} - \epsilon_{ij} < u_{ij} - u_{ij'} \; \forall \; j' \neq j) \qquad (2) \\ &= \int_{\epsilon} I(\epsilon_{ij'} - \epsilon_{ij} < u_{ij} - u_{ij'} \; \forall \; j' \neq j) f(\epsilon) d\epsilon \end{aligned}$$

Regardless of the distribution of the ϵ's, probabilities don't change when we:

- Add a constant to the utility of all options

  or

- Multiply by a positive number

Regardless of the distribution of the ϵ's, probabilities don't change when we:

- Add a constant to the utility of all options

  or

- Multiply by a positive number

Thus, we need to make some normalizations to identify preferences:

- Only *differences* in utility matter (location normalization)

- We need to scale the variance of the ϵ's (scale normalization)

Suppose we have two options with the following observable utilities:

$$u_{i1} = \alpha Tall_i + \beta_1 X_i + \gamma Z_1$$
$$u_{i2} = \alpha Tall_i + \beta_2 X_i + \gamma Z_2$$

Suppose we have two options with the following observable utilities:

$$u_{i1} = \alpha Tall_i + \beta_1 X_i + \gamma Z_1$$
$$u_{i2} = \alpha Tall_i + \beta_2 X_i + \gamma Z_2$$

Since only differences in utility matter:

$$u_{i1} - u_{i2} = (\beta_1 - \beta_2) X_i + \gamma (Z_1 - Z_2)$$

Suppose we have two options with the following observable utilities:

$$u_{i1} = \alpha Tall_i + \beta_1 X_i + \gamma Z_1$$
$$u_{i2} = \alpha Tall_i + \beta_2 X_i + \gamma Z_2$$

Since only differences in utility matter:

$$u_{i1} - u_{i2} = (\beta_1 - \beta_2) X_i + \gamma (Z_1 - Z_2)$$

- We can't tell whether tall people are happier (overall) than short people, but we can tell whether they more strongly prefer one option
- We can only obtain *differenced* coefficient estimates on X's
- We can only obtain an estimate of a coefficient that is constant across choice alternatives if its corresponding variable varies by alternative

Because only differences in utility matter, we end up having one fewer dimension of ϵ

Rewriting the last line of (2) as a $J-1$ dimensional integral over the *differenced* ϵ's:

$$P_{ij} = \int_{\tilde{\epsilon}} I(\tilde{\epsilon}_{ij'} < \tilde{u}_{ij'} \ \forall\ j' \neq j) g(\tilde{\epsilon}) d\tilde{\epsilon} \tag{4}$$

where $\tilde{\epsilon}_j \equiv \epsilon_j - \epsilon_J$, etc. ($J$ is reference alternative)

- The need to normalize scale means we can never estimate the variance of $G(\tilde{\epsilon})$

- This contrasts with linear regression models, where we can easily estimate MSE

- The need to normalize scale means we can never estimate the variance of $G(\tilde{\epsilon})$

- This contrasts with linear regression models, where we can easily estimate MSE

- The scale normalization means our β's and γ's are implicitly divided by an unknown variance term:

- The need to normalize scale means we can never estimate the variance of $G(\tilde{\epsilon})$
- This contrasts with linear regression models, where we can easily estimate MSE
- The scale normalization means our β's and γ's are implicitly divided by an unknown variance term:

$$
\begin{aligned}
u_{i1} - u_{i2} &= (\beta_1 - \beta_2)X_i + \gamma(Z_1 - Z_2) \\
&= \tilde{\beta}X_i + \gamma\tilde{Z} \\
&= \frac{\beta^*}{\sigma}X_i + \frac{\gamma^*}{\sigma}\tilde{Z}
\end{aligned}
$$

$\tilde{\beta}$ is what we estimate, but we will never know β^* because utility is scale-invariant

	(1)	(2)	(3)
Price ($)	-0.145	-0.112	-0.098
	(0.023)	(0.028)	(0.031)
Distance (miles)		-0.067	-0.054
		(0.015)	(0.018)
Weekend			0.423
			(0.089)
N	2,456	2,456	2,456
Pseudo R^2	0.142	0.198	0.234

Important: We cannot directly compare coefficient magnitudes across specifications due to the scale normalization. Each model implicitly rescales coefficients by $1/\sigma$, where σ varies with the set of included variables. ($\sigma \downarrow$ as more covariates are included)

Instead, use marginal effects for comparison across specifications