Topic 3:
Choice models with multiple alternatives

Takeaki Sunada[1]

[1]Simon Business School
University of Rochester

Multinomial logit models

- Binary logit models ignore "what consumers do if they don't buy KB". Not buying KB is not buying KB. However, "buy KR instead of KB" vs "buy nothing" may imply completely different things.
- Today we use "multinomial logit models" to consider situations where consumers face more than two choice alternatives. We now incorporate product substitution, which is relevant in competitive pricing policies and product line (multi-product) pricing.
- We still assume that demand is homogeneous (one set of parameters apply to all consumer's demand). Hence no segmentation yet.

Full choice data

id	week	trip	price.0	price.KB	price.KR	price.MB	choice
1	96	1	0	1.43	1.43	1.43	0
2	14	1	0	1.43	1.43	1.65	0
2	25	2	0	1.43	1.43	1.65	0
2	26	3	0	1.43	1.43	1.65	0
2	31	4	0	1.43	0.88	1.65	KB
2	89	5	0	1.43	1.43	1.32	KB
2	94	6	0	0.9	0.89	1.43	KR
2	96	7	0	1.43	1.43	1.43	0
2	114	8	0	1.43	1.43	1.34	MB
2	126	9	0	1.43	1.43	1.26	MB
2	146	10	0	1.47	1.43	1.33	KR

Notations

- Up until now, we denoted the event "a consumer buys KB" by "$q^{KB} = 1$".
- As a convention of multinomial choice models, we will refer to the same event by "$y = KB$" from now on. y represents the alternative that the consumer selects.
- Hence $Pr(q^{KB} = 1 \mid P^{KB}) = Pr(y = KB \mid P^{KB})$.
- In multinomial choice environments, this notation is easier to deal with.

Binary logit models

- In a binary logit environment, we only needed to specify the probability that a consumer buys "KB".

$$Pr(y = KB \mid P^{KB}) = \frac{\exp(\beta_0 + \beta_1 P^{KB})}{1 + \exp(\beta_0 + \beta_1 P^{KB})}.$$

- Note that although consumers face two choice options (buy and not buy), we only specified one choice probability - probability of buying. We know the probability of not buying is $1 - Pr(y = KB \mid P^{KB})$, because probabilities have to sum to one.

Multinomial logit models

- Suppose now that we add one more option, "KR" with price P^{KR}. Consumers can choose from the two products, or not buy. How many choice probabilities do we need?
- We need two choice probabilities - say $Pr(y = KB \mid P^{KB}, P^{KR})$ and $Pr(y = KR \mid P^{KB}, P^{KR})$. Then the probability for not buying is $1 - Pr(y = KB \mid P^{KB}, P^{KR}) - Pr(y = KR \mid P^{KB}, P^{KR})$.
- Generally, when consumers face J choice options (including no purchase), the model needs to specify $J - 1$ choice probabilities.

Multinomial logit models

- Suppose we want to model $Pr(y = KB \mid P^{KB}, P^{KR})$ and $Pr(y = KR \mid P^{KB}, P^{KR})$. What is the best way?
- The choice probabilities need to be "consistent" with each other - we want them to satisfy some conditions. For example:
  - They need to be between zero and one (same as binary case), and their *sum* needs to be between zero and one (new).
  - More popular product should be assigned a higher probability (new).
  - If the price of a product goes up, we expect that choice probability of that product goes down (same as binary case), and that of its rival product goes up (new).

Multinomial logit models

- Here's how we construct multinomial logit models.
- Define the probability that a consumer chooses KB as follows.

$$Pr(y = KB \mid P) = \frac{\exp(\beta_0^{KB} + \beta_1 P^{KB})}{1 + \exp(\beta_0^{KB} + \beta_1 P^{KB}) + \exp(\beta_0^{KR} + \beta_1 P^{KR})},$$

  where $P = \{P^{KB}, P^{KR}\}$.

- The parameters of the model is now $\{\beta_0^{KB}, \beta_0^{KR}, \beta_1\}$.

Multinomial logit models

- Let's compare the choice probability with that of binary logit model. Here's this model.

$$Pr(y = KB \mid P) = \frac{\exp(\beta_0^{KB} + \beta_1 P^{KB})}{1 + \exp(\beta_0^{KB} + \beta_1 P^{KB}) + \exp(\beta_0^{KR} + \beta_1 P^{KR})}.$$

  Here's binary case.

$$Pr(y = KB \mid P^{KB}) = \frac{\exp(\beta_0^{KB} + \beta_1 P^{KB})}{1 + \exp(\beta_0^{KB} + \beta_1 P^{KB})}.$$

Multinomial logit models

- Let's compare the choice probability with that of binary logit model.

  Here's this model.

$$Pr(y = KB \mid P) = \frac{\exp(\beta_0^{KB} + \beta_1 P^{KB})}{1 + \exp(\beta_0^{KB} + \beta_1 P^{KB}) + \exp(\beta_0^{KR} + \beta_1 P^{KR})}.$$

  Here's binary case.

$$Pr(y = KB \mid P^{KB}) = \frac{\exp(\beta_0^{KB} + \beta_1 P^{KB})}{1 + \exp(\beta_0^{KB} + \beta_1 P^{KB})}.$$

Multinomial logit models

- Imagine that each product gets a score: score of KB is $\exp(\beta_0^{KB} + \beta_1 P^{KB})$, and that of KR is $\exp(\beta_0^{KR} + \beta_1 P^{KR})$.
- The probability that KB is selected is "the score of KB, divided by the sum of all the scores $+1$".

$$Pr(y = KB \mid P) = \frac{\exp(\beta_0^{KB} + \beta_1 P^{KB})}{1 + \exp(\beta_0^{KB} + \beta_1 P^{KB}) + \exp(\beta_0^{KR} + \beta_1 P^{KR})}.$$

Multinomial logit models

- We can define $Pr(y = KR \mid P)$ analogously:

$$
\begin{aligned}
Pr(y =& KR \mid P) \\
&= \frac{\exp(\beta_0^{KR} + \beta_1 P^{KR})}{1 + \exp(\beta_0^{KB} + \beta_1 P^{KB}) + \exp(\beta_0^{KR} + \beta_1 P^{KR})}
\end{aligned}
$$

- Then we can also get the probability that a consumer buys nothing.

$$
\begin{aligned}
&Pr(y = 0 \mid P) \\
&\quad = 1 - Pr(y = KB \mid P) - Pr(y = KR \mid P) \\
&\quad = \frac{1}{1 + \exp(\beta_0^{KB} + \beta_1 P^{KB}) + \exp(\beta_0^{KR} + \beta_1 P^{KR})}
\end{aligned}
$$

Multinomial logit models

- These expressions have all the properties we wanted:

$$Pr(y = KB \mid P) = \frac{\exp(\beta_0^{KB} + \beta_1 P^{KB})}{1 + \exp(\beta_0^{KB} + \beta_1 P^{KB}) + \exp(\beta_0^{KR} + \beta_1 P^{KR})}.$$

- There's own-effect.
  - Each product has its own β_0^j. Think of it as some underlying popularity (or consumers' innate preference) of the product - a product with a higher β_0^j is selected with a higher probability.
  - A product with higher P^j faces lower choice probability through $\beta_1 < 0$.

Multinomial logit models

$$Pr(y = KB \mid P) = \frac{\exp(\beta_0^{KB} + \beta_1 P^{KB})}{1 + \exp(\beta_0^{KB} + \beta_1 P^{KB}) + \exp(\beta_0^{KR} + \beta_1 P^{KR})}.$$

- and there's competitive effect through demand substitution.
  - The choice probability of product j goes down if there's another product j' with higher $\beta_0^{j'}$.
  - Similarly, a product is less likely to be selected if rival price goes down.
- The sum of the two choice probabilities $(Pr(y = KB \mid P) + Pr(y = KR \mid P))$ is less than one.

Multinomial logit models

- In general, in an environment with $j = \{1, ..., J\}$ products, a multinomial logit model takes the following form.

$$Pr(y = j \mid P) = \frac{\exp(\beta_0^j + \beta_1 P^j)}{1 + \sum_{j'=1}^{J} \exp(\beta_0^{j'} + \beta_1 P^{j'})},$$

where $P = \{P^1, ..., P^j, ..., P^J\}$.

- "Buying nothing" is then

$$Pr(y = 0 \mid P) = \frac{1}{1 + \sum_{j'=1}^{J} \exp(\beta_0^{j'} + \beta_1 P^{j'})},$$

- In this case, the model parameters are $\beta = \{\beta_0^1, ..., \beta_0^j, ..., \beta_0^J, \beta_1\}$.

Exercise 1

- Suppose that there exists 1000 consumers with the choice probability governed by $\beta_0^{KB} = 3$, $\beta_0^{KR} = 2$, and $\beta_1 = -2$. Your product KB is more popular than the rival, KR.
- The rival's price is $P^{KR} = 1$ and your unit cost is 1. What is the price you should set?

Code for this exercise

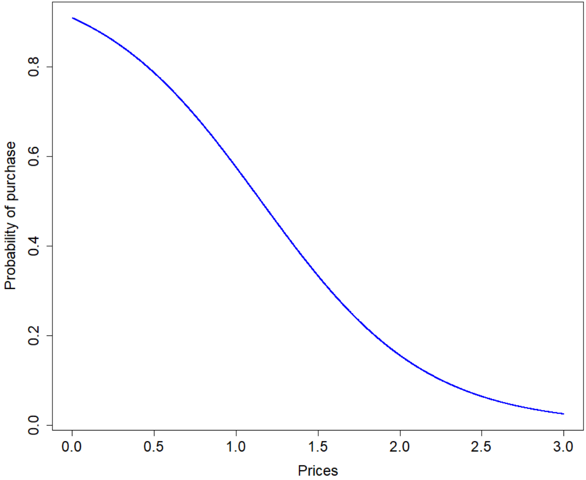
- Just like the binary case, first define the probability of purchase as a function of prices and parameters.

```
#Multinomial logit: illustration
#Write choice probability as a function
demand=function(priceKB,priceKR,beta0KB,beta0KR,beta1){
  prob=exp(beta0KB+beta1*priceKB)/(1+exp(beta0KB+beta1*priceKB)+exp(beta0KR+beta1*priceKR))
  return(prob)
}
```

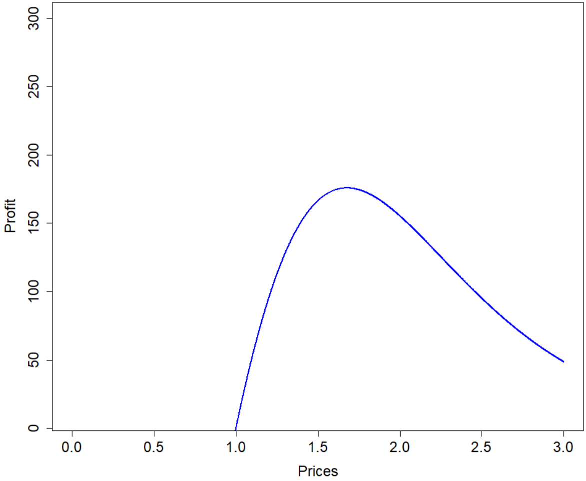
- Then call that function with assigned parameters and prices.

```
#Demand case 1
pricespace=seq(0,3,0.001)
plot(pricespace,demand(pricespace,1,3,2,-2),type='l',xlab='Prices',
     ylab='Probability of purchase',col="blue",lwd=2
     ,cex=2,cex.lab=1.5, cex.axis=1.5, cex.main=1.5, cex.sub=1.5)
```

Choice probability of KB



Profit of KB

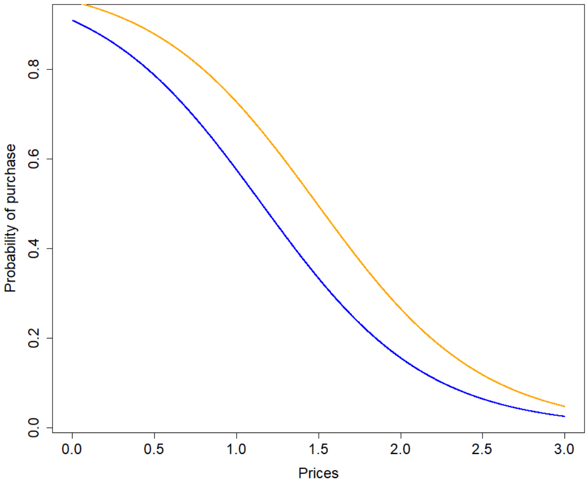


- The optimal price is 1.68.

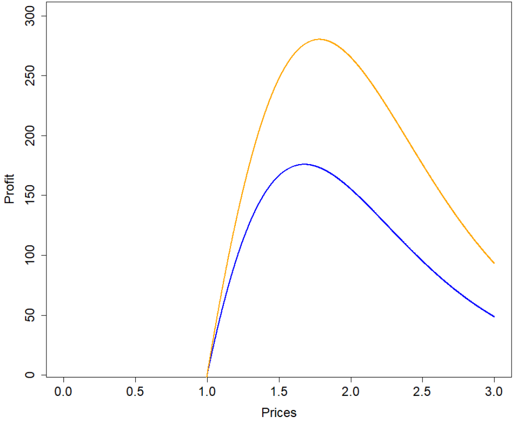
Exercise 2

- Suppose that the rival increased its price to $P^{KR} = 3$. Should you change your price too? If yes, by how much?

Demand shift



Profit shift

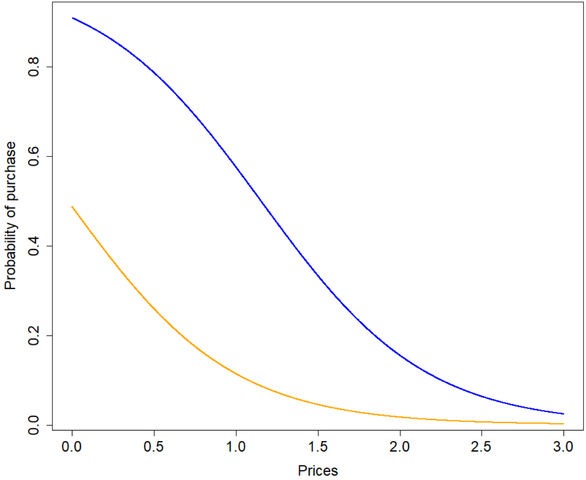


- The optimal price is now 1.78.

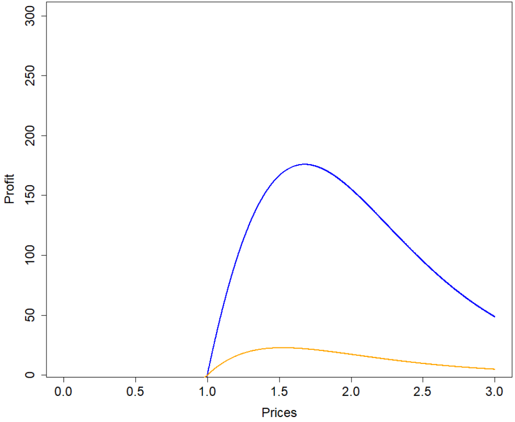
Exercise 3

- Suppose that the rival ran an intensive promotion. As a result the rival's β_0^{KR} has increased to 5: the rival is now more popular than you. Should you change the price? If yes, what is the new price you should charge?

Demand shift



Profit shift

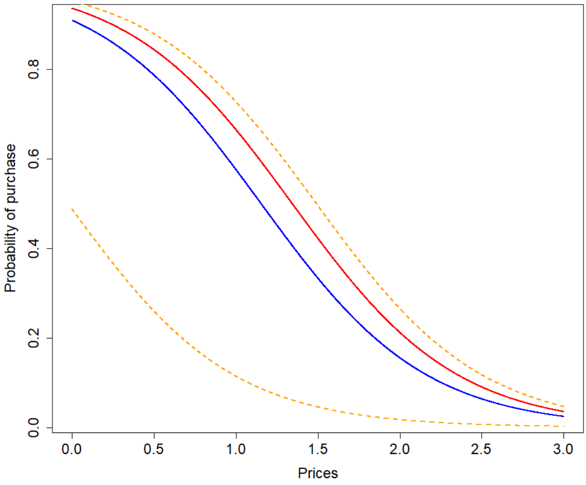


- The optimal price is now 1.52.

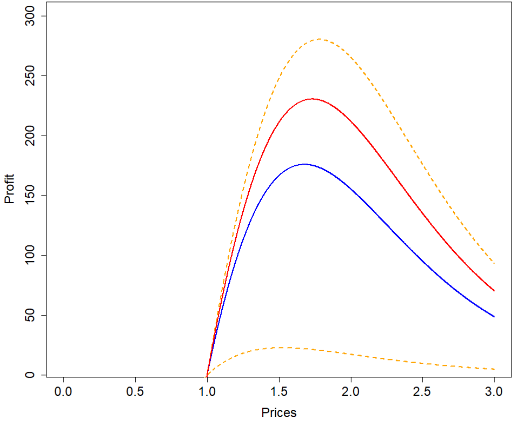
Exercise 4

- Suppose that the rival ran an intensive promotion AND increases its price (because they are now more popular). $\beta_0^{KR} = 5$ and $P^{KR} = 3$. Do we need to react by changing our price? If yes, what is the new price?

Demand shift



Profit shift

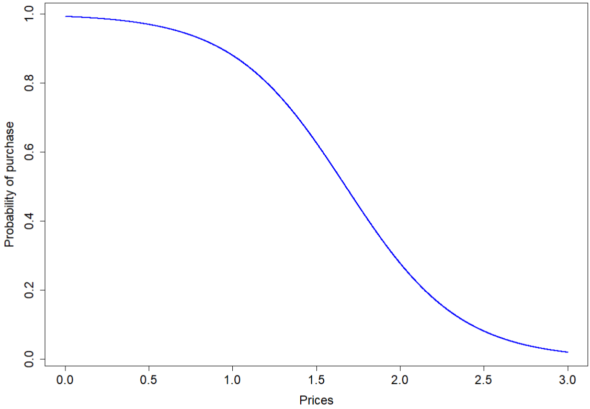


- The optimal price is now 1.73.

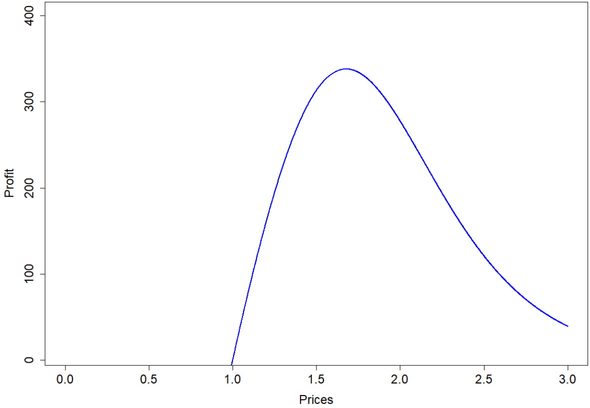
Degree of product substitution

- You may say that optimal prices are not really responsive to rival's prices. But of course the degree of competitive reaction depends on model parameters.
- To see this, let's do the same exercise at different degree of product substitution: $\beta_0^{KB} = 10$, $\beta_0^{KR} = 8$, and $\beta_1 = -2.95$.
- The rival's price is $P^{KR} = 1$ and your unit cost is 1. What is the price you should set?

Choice probability of KB



Profit of KB

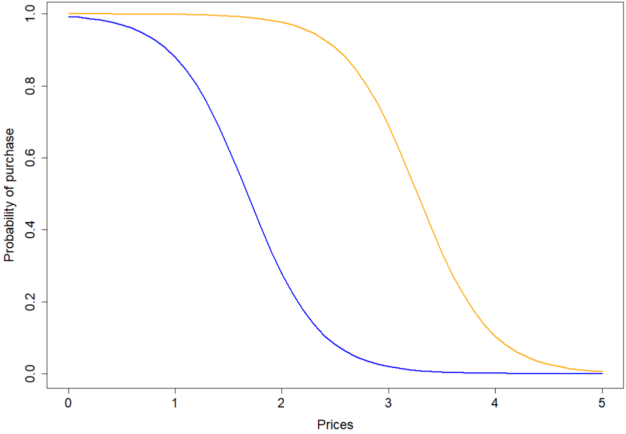


- The optimal price is 1.68 (same as before).

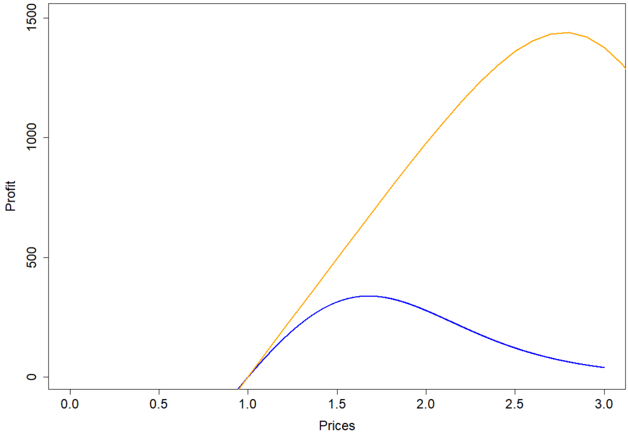
Exercise 2

- Suppose that the rival increased its price to $P^{KR} = 3$. Should you change your price too? If yes, by how much?

Demand shift



Profit shift



- The optimal price is now 2.8.

Multinomial logit models

- As such, the model allows us to consider pricing when competitive pressure exists. In particular, we can study how we should react to rival actions (price change, advertising, etc).
- Moreover, if we own both KB and KR, then this model informs us the magnitude of demand cannibalization (substitution) one product creates on the other. Hence we can optimize product-line pricing.

Elasticity

- Even with multinomial options, price elasticity of demand is still easy to define.
- Own-price elasticity of KB remains exactly identical as in the binary case.

$$\frac{\frac{\partial Pr(y=KB)}{Pr(y=KB)}}{\frac{\partial P^{KB}}{P^{KB}}} = -\beta_1 P^{KB}(1 - Pr(y = KB \mid P)).$$

- Cross-price elasticity is similarly defined in a simple form. This is "percent change in choice probability of KB when the price of KR moves by one percent".

$$\frac{\frac{\partial Pr(y=KB)}{Pr(y=KB)}}{\frac{\partial P^{KR}}{P^{KR}}} = -\beta_1 P^{KR} Pr(y = KR \mid P).$$

Elasticity

- More generally, own- and cross-elasticity in J products cases are given as follows (just notational change from previous example).
- Own-elasticity:

$$\frac{\frac{\partial Pr(y=j)}{Pr(y=j)}}{\frac{\partial P^j}{P^j}} = -\beta_1 P^j (1 - Pr(y = j \mid P)).$$

- Cross-elasticity from the price change of product j' to demand shift of product j:

$$\frac{\frac{\partial Pr(y=j)}{Pr(y=j)}}{\frac{\partial P^{j'}}{P^{j'}}} = -\beta_1 P^{j'} Pr(y = j' \mid P).$$

Estimation of multinomial logit models

- Let's estimate a multinomial logit model with the data we have. The data contain three options: KB, KR and MB (and zero).
- Just like binary logit case, we use "gmnl" and "mlogit".

Full choice data

id	week	trip	price.0	price.KB	price.KR	price.MB	choice
1	96	1	0	1.43	1.43	1.43	0
2	14	1	0	1.43	1.43	1.65	0
2	25	2	0	1.43	1.43	1.65	0
2	26	3	0	1.43	1.43	1.65	0
2	31	4	0	1.43	0.88	1.65	KB
2	89	5	0	1.43	1.43	1.32	KB
2	94	6	0	0.9	0.89	1.43	KR
2	96	7	0	1.43	1.43	1.43	0
2	114	8	0	1.43	1.43	1.34	MB
2	126	9	0	1.43	1.43	1.26	MB
2	146	10	0	1.47	1.43	1.33	KR

Estimation of multinomial logit models with "gmnl"

```
#Now columns 4 through 7 contains "Price.something" info.
mlogitdata=mlogit.data(data,id="id",varying=4:7,choice="choice",shape="wide")

#Run MLE.
mle= gmnl(choice ~  price, data = mlogitdata)
summary(mle)
```

- The code is almost identical to the binary case, except that in "mlogit.data", we need to specify all columns that include price info (4 columns in total).

Intuition on parameter estimates

- Intuitions on how we estimate all the β_0^j's and β_1 is a pure extension of binary logit case.
- Because of MLE, we estimate β to match model prediction the best with the observed choices of consumers.
- Consider a likelihood function that corresponds to "N observations, KB selected K times, KR selected L times, and MB selected M times".

$$
\begin{aligned}
& L(K, L, M, N-K-L-M \mid P) \\
& \quad = Pr(y = KB \mid P)^K \times Pr(y = KR \mid P)^L \\
& \qquad \times Pr(y = MB \mid P)^M \times Pr(y = 0 \mid P)^{N-K-L-M}
\end{aligned}
$$

Intuition on parameter estimates

$$
\begin{aligned}
&L(K, L, M, N-K-L-M \mid P) \\
&= \left(\frac{\exp(\beta_0^{KB} + \beta_1 P^{KB})}{1 + \exp(\beta_0^{KB} + \beta_1 P^{KB}) + \exp(\beta_0^{KR} + \beta_1 P^{KR}) + \exp(\beta_0^{MB} + \beta_1 P^{MB})} \right)^K \times \ldots
\end{aligned}
$$

- If KB gets selected more often (compared to others), β_0^{KB} is higher than $\beta_0^{j'}$ of the other products.
- If a price increase by a product significantly decreases the market share of that product and increases the market share of its rivals, we estimate a large negative β_1.

Estimation of multinomial logit models with "gmnl"

```
Coefficients:
                  Estimate Std. Error  z-value  Pr(>|z|)    
KB:(intercept)  6.15202    0.60211  10.2174 < 2.2e-16 ***
KR:(intercept)  6.01682    0.60275   9.9823 < 2.2e-16 ***
MB:(intercept)  5.82222    0.56675  10.2730 < 2.2e-16 ***
price          -4.93438    0.43220 -11.4170 < 2.2e-16 ***
```

- What do we conclude from here?

Estimation of multinomial logit models with "gmnl"

```
Coefficients:
                  Estimate Std. Error  z-value  Pr(>|z|)    
KB:(intercept)    6.15202    0.60211  10.2174 < 2.2e-16 ***
KR:(intercept)    6.01682    0.60275   9.9823 < 2.2e-16 ***
MB:(intercept)    5.82222    0.56675  10.2730 < 2.2e-16 ***
price            -4.93438    0.43220 -11.4170 < 2.2e-16 ***
```

- Note that the baseline popularity of each product is almost identical. This is because all products are selected roughly at the same rate (KB 117 times, KR 105 times and MB 108 times).
- Assuming homogeneous demand, this is saying that consumers do not particularly prefer any of the products - it that true? We will explore this further in the following weeks.

Summary of current status + looking ahead

- We started from binary logit models. It has two major problems - (1) it doesn't capture competition, and (2) it doesn't allow for consumer heterogeneity.
- Today, we studied multinomial logit models. Now that we are fully utilizing pieces of information available in the data. In particular, we can analyze competitive effect / product-line pricing with the data.

Summary of current status + looking ahead

- In the next deck of slides, we move into finding segments in the market.
- We first study how to segment market according to observed demographic variables. This is a simpler form of segmentation - we observe cues based on which we group consumers together.
- We then move on segmentation based on latent type of consumers. This is harder (data-wise. Actually, coding-wise this may be easier than demographic-based segmentation), but it does capture underlying demand structure better.