Traditional Conjoint (or Trade-off) Analysis and Market Simulation

Hande is a student in GE 461 class and wants to buy a new laptop in order to use it in GE 461 (and later) to learn better how to apply data science to business problems. She goes online and glimpse at the alternative laptops in Table 1 offered for sale at http://www.hepsiburada.com.

Alternative	Brand	Processor	Speed	Memory	HardDisk	Screen	Price	RANK
1	Toshiba	Intel Core i7	2.4 GHz	8 GB	750 GB	17.3 in	3500 USD	6
2	Toshiba	Intel Core i5	2.6 GHz	4 GB	750 GB	15.6 in	2000 USD	2
3	Casper	Intel Core i7	2.2 GHz	8 GB	750 GB	15.6 in	1700 USD	1
4	Casper	Intel Core i5	2.4 GHz	4 GB	500 GB	15.6 in	1500 USD	3
5	Sony	Intel Core i5	1.8 GHz	4 GB	500 GB	15.6 in	1700 USD	5
6	HP	Intel Core i5	1.8 GHz	4 GB	500 GB	15.6 in	1500 USD	4

Table 1: Hande's rankings for the laptops listed in hepsiburada.com

The last column (RANK) in Table 1 shows how attractive she finds each laptop relative to the others in the same list (1 = most attractive, 6 = least attractive).

In the meantime, Casper is interested in increasing its laptop market share. The company wants to introduce to the market a new laptop that will hopefully increase its sales. To help with the design of the new laptop, the company would like firstly to explore the importance of each laptop feature (brand, processor, speed, memory size, hard disk capacity, screen size, and price) to the Turkish customers. The company also wants to know how a potential customer trades off between various features of a laptop.

The company decides to run a market survey to understand the customer behavior. It restricts its survey to the following features of a laptop and their values:

Brand	Processor	Speed	Memory	HardDisk	Screen	Price
Casper	Intel Core i5	1.8 GHz	4 GB	500 GB	15.6 in	1500 USD
HP	Intel Core i7	2.2 GHz	8 GB	750 GB	17.3 in	1700 USD
Sony		2.4 GHz				2000 USD
Toshiba		2.6 GHz				3500 USD

To assess how the customers solve the trade-offs between superior and inferior levels of different features, the company carefully designs the experiment in Table 2, in which each potential customer is asked to rank 16 laptop configurations from the most preferred (rank 1) to the least preferred (rank 16). Hande happens to be one of the participants in the survey, and the last column (named RANK) shows her preferences.

It is quite natural to think that $-\texttt{RANK}_i$ reflects the utility that a survey taker expects to receive from the ith alternative. We may even assume that for the survey taker

- total utility $\texttt{U}_i$ derived from the ith alternative is proportional to $-\texttt{RANK}_i$,

- total utility $\texttt{U}_i$ is the sum of the partial utilities $\texttt{UBrand}_i, \texttt{UProcessor}_i, \ldots, \texttt{UPrice}_i$ derived from Brand, Processor, ..., Price features of the ith alternative plus utility ε_i derived from other unobserved features of the same alternative; namely,

$$-\texttt{RANK}_i \propto \texttt{U}_i = \texttt{UBrand}_i + \texttt{UProcessor}_i + \ldots + \texttt{UPrice}_i + \varepsilon_i. \tag{1}$$

Let $\widetilde{\beta}_{\texttt{Casper}}, \widetilde{\beta}_{\texttt{HP}}, \widetilde{\beta}_{\texttt{Sony}}, \widetilde{\beta}_{\texttt{Toshiba}}$ be the utilities that this survey taker receives from brand names Casper, HP, Sony, and Toshiba, respectively. If $\texttt{Brand}_i$ shows the brand of the ith alternative, then

$$\texttt{UBrand}_i = \widetilde{\beta}_{\texttt{Casper}}1_{\{\texttt{Brand}_i=\texttt{Casper}\}} + \widetilde{\beta}_{\texttt{HP}}1_{\{\texttt{Brand}_i=\texttt{HP}\}} + \widetilde{\beta}_{\texttt{Sony}}1_{\{\texttt{Brand}_i=\texttt{Sony}\}} + \widetilde{\beta}_{\texttt{Toshiba}}1_{\{\texttt{Brand}_i=\texttt{Toshiba}\}} \tag{2}$$

Alternative	Brand	Processor	Speed	Memory	HardDisk	Screen	Price	RANK
1	HP	Intel Core i5	1.8 GHz	4 GB	500 GB	15.6 in	1500 USD	8
2	Sony	Intel Core i5	2.4 GHz	8 GB	750 GB	17.3 in	1500 USD	1
3	Casper	Intel Core i5	2.6 GHz	4 GB	500 GB	17.3 in	1700 USD	6
4	Sony	Intel Core i7	2.6 GHz	8 GB	500 GB	15.6 in	3500 USD	13
5	Sony	Intel Core i5	2.2 GHz	4 GB	500 GB	17.3 in	2000 USD	10
6	Sony	Intel Core i7	1.8 GHz	4 GB	750 GB	15.6 in	1700 USD	11
7	Casper	Intel Core i7	2.4 GHz	4 GB	750 GB	15.6 in	2000 USD	12
8	HP	Intel Core i5	2.6 GHz	8 GB	750 GB	15.6 in	2000 USD	3
9	HP	Intel Core i7	2.2 GHz	4 GB	750 GB	17.3 in	3500 USD	14
10	Toshiba	Intel Core i7	2.6 GHz	4 GB	750 GB	17.3 in	1500 USD	9
11	Casper	Intel Core i5	1.8 GHz	8 GB	750 GB	17.3 in	3500 USD	16
12	Toshiba	Intel Core i5	2.2 GHz	8 GB	750 GB	15.6 in	1700 USD	2
13	Toshiba	Intel Core i7	1.8 GHz	8 GB	500 GB	17.3 in	2000 USD	7
14	Casper	Intel Core i7	2.2 GHz	8 GB	500 GB	15.6 in	1500 USD	5
15	Toshiba	Intel Core i5	2.4 GHz	4 GB	500 GB	15.6 in	3500 USD	15
16	HP	Intel Core i7	2.4 GHz	8 GB	500 GB	17.3 in	1700 USD	4

Table 2: Experimental design and Hande's rankings

is how we calculate the utility $\texttt{UBrand}_i$ derived from Brand for the same survey taker. Because all alternatives on our survey belongs to one of those four brands, the sum of four indicators always equals one. For example, if we know that the ith alternative is none of HP, Sony, or Toshiba, then its brand should definitely be Casper. In other words,

$$1_{\{\texttt{Brand}_i=\texttt{Casper}\}} = 1 - 1_{\{\texttt{Brand}_i=\texttt{HP}\}} - 1_{\{\texttt{Brand}_i=\texttt{Sony}\}} - 1_{\{\texttt{Brand}_i=\texttt{Toshiba}\}}.$$

If we substitute this into (2), then we get

$$
\begin{aligned}
\texttt{UBrand}_i &= \widetilde{\beta}_{\texttt{Casper}}\Big(1 - 1_{\{\texttt{Brand}_i=\texttt{HP}\}} - 1_{\{\texttt{Brand}_i=\texttt{Sony}\}} - 1_{\{\texttt{Brand}_i=\texttt{Toshiba}\}}\Big) + \\
&\qquad \widetilde{\beta}_{\texttt{HP}}1_{\{\texttt{Brand}_i=\texttt{HP}\}} + \widetilde{\beta}_{\texttt{Sony}}1_{\{\texttt{Brand}_i=\texttt{Sony}\}} + \widetilde{\beta}_{\texttt{Toshiba}}1_{\{\texttt{Brand}_i=\texttt{Toshiba}\}} \\
&= \widetilde{\beta}_{\texttt{Casper}} + (\widetilde{\beta}_{\texttt{HP}} - \widetilde{\beta}_{\texttt{Casper}})1_{\{\texttt{Brand}_i=\texttt{HP}\}} + (\widetilde{\beta}_{\texttt{Sony}} - \widetilde{\beta}_{\texttt{Casper}})1_{\{\texttt{Brand}_i=\texttt{Sony}\}} + (\widetilde{\beta}_{\texttt{Toshiba}} - \widetilde{\beta}_{\texttt{Casper}})1_{\{\texttt{Brand}_i=\texttt{Toshiba}\}} \\
&= \widetilde{\beta}_{\texttt{Casper}} + \widehat{\beta}_{\texttt{HP}}1_{\{\texttt{Brand}_i=\texttt{HP}\}} + \widehat{\beta}_{\texttt{Sony}}1_{\{\texttt{Brand}_i=\texttt{Sony}\}} + \widehat{\beta}_{\texttt{Toshiba}}1_{\{\texttt{Brand}_i=\texttt{Toshiba}\}},
\end{aligned}
$$

where $\widehat{\beta}_{\texttt{HP}}, \widehat{\beta}_{\texttt{Sony}}, \widehat{\beta}_{\texttt{Toshiba}}$ are now the *excess* utilities derived, respectively, from HP, Sony, and Toshiba over the the utility of Casper. Because they are differences of the brand utilities (effects, in general), they are also called *contrasts* between the brand utilities or effects. Here, Casper became the baseline or reference Brand level. In statistics, the indicator functions $1_{\{\texttt{Brand}_i=\texttt{HP}\}}, 1_{\{\texttt{Brand}_i=\texttt{Sony}\}}, 1_{\{\texttt{Brand}_i=\texttt{Toshiba}\}}$ are often called the dummy variables; and typically expressed/defined as

$$\texttt{BrandHP}_i := \begin{cases} 1, & \text{if } \texttt{Brand}_i = \text{HP}, \\ 0, & \text{otherwise}, \end{cases} \quad \texttt{BrandSony}_i := \begin{cases} 1, & \text{if } \texttt{Brand}_i = \text{Sony}, \\ 0, & \text{otherwise}, \end{cases}$$

$$\texttt{BrandToshiba}_i := \begin{cases} 1, & \text{if } \texttt{Brand}_i = \text{Toshiba}, \\ 0, & \text{otherwise}, \end{cases}$$

and the utility derived from brand of ith alternative becomes

$$\texttt{UBrand}_i = \widetilde{\beta}_{\texttt{Casper}} + \widehat{\beta}_{\texttt{HP}} \times \texttt{BrandHP}_i + \widehat{\beta}_{\texttt{Sony}} \times \texttt{BrandSony}_i + \widehat{\beta}_{\texttt{Toshiba}} \times \texttt{BrandToshiba}_i. \tag{3}$$

Note that one of the brands (here, Casper) was chosen as the baseline (reference), and then we define one 0-1 dummy variable for each of the remaining three brands. This is known as *treatment-contrast* representation. One undesirable property of this representation is that contrasts are measured relative to Casper **Brand**, but this

choice of baseline or reference level was totally arbitrary. For example, other analysts may choose to work with different baseline brands and miscommunication may take place.

That shortcoming of treatment-contrast motivates an alternative representation, so-called *sum-contrast* representation of the same utility. Let us *not* take the utility of any specific brand, but the average of utilities of all brands,

$$\beta_{\text{average}}^{\text{brand}} := \frac{\widetilde{\beta}_{\text{HP}} + \widetilde{\beta}_{\text{HP}} + \widetilde{\beta}_{\text{Sony}} + \widetilde{\beta}_{\text{Toshiba}}}{4}$$

as reference/baseline. After adding to and subtraction from (2) the average $\beta_{\text{average}}^{\text{brand}}$, (2) becomes

$$\text{UBrand}_i = \beta_{\text{average}}^{\text{brand}} + (\widetilde{\beta}_{\text{Casper}} - \beta_{\text{average}}^{\text{brand}})1_{\{\text{Brand}_i = \text{Casper}\}} + (\widetilde{\beta}_{\text{HP}} - \beta_{\text{average}}^{\text{brand}})1_{\{\text{Brand}_i = \text{HP}\}} +$$
$$(\widetilde{\beta}_{\text{Sony}} - \beta_{\text{average}}^{\text{brand}})1_{\{\text{Brand}_i = \text{Sony}\}} + (\widetilde{\beta}_{\text{Toshiba}} - \beta_{\text{average}}^{\text{brand}})1_{\{\text{Brand}_i = \text{Toshiba}\}}$$
$$= \beta_{\text{average}}^{\text{brand}} + \beta_{\text{Casper}}1_{\{\text{Brand}_i = \text{Casper}\}} + \beta_{\text{HP}}1_{\{\text{Brand}_i = \text{HP}\}} + \beta_{\text{Sony}}1_{\{\text{Brand}_i = \text{Sony}\}} + \beta_{\text{Toshiba}}1_{\{\text{Brand}_i = \text{Toshiba}\}}, \quad (4)$$

where $\beta_{\text{Casper}}, \beta_{\text{HP}}, \beta_{\text{Sony}}, \beta_{\text{Toshiba}}$ are the *excess* utilities of all brands *over average utility of all brands*. The reference/baseline is, indeed, the average utility $\beta_{\text{average}}^{\text{brand}}$, and we also observe that

$$\beta_{\text{Casper}} + \beta_{\text{HP}} + \beta_{\text{Sony}} + \beta_{\text{Toshiba}} = 0,$$

which implies, for example,

$$\beta_{\text{Toshiba}} = -\beta_{\text{Casper}} - \beta_{\text{HP}} - \beta_{\text{Sony}}. \quad (5)$$

Namely, $\beta_{\text{Casper}}, \beta_{\text{HP}}, \beta_{\text{Sony}}$ can change freely, but once we find their values, β_{Toshiba} is uniquely determined by (5). Therefore, we should eliminate β_{Toshiba} from (4) by substituting into it the equality in (5). We get

$$\text{UBrand}_i = \beta_{\text{average}}^{\text{brand}} + \beta_{\text{Casper}}1_{\{\text{Brand}_i = \text{Casper}\}} + \beta_{\text{HP}}1_{\{\text{Brand}_i = \text{HP}\}} + \beta_{\text{Sony}}1_{\{\text{Brand}_i = \text{Sony}\}}$$
$$+ (-\beta_{\text{Casper}} - \beta_{\text{HP}} - \beta_{\text{Sony}})1_{\{\text{Brand}_i = \text{Toshiba}\}}$$
$$= \beta_{\text{average}}^{\text{brand}} + \beta_{\text{Casper}}(1_{\{\text{Brand}_i = \text{Casper}\}} - 1_{\{\text{Brand}_i = \text{Toshiba}\}}) + \beta_{\text{HP}}(1_{\{\text{Brand}_i = \text{HP}\}} - 1_{\{\text{Brand}_i = \text{Toshiba}\}})$$
$$+ \beta_{\text{Sony}}(1_{\{\text{Brand}_i = \text{Sony}\}} - 1_{\{\text{Brand}_i = \text{Toshiba}\}}).$$

If we define ± 1-dummy variables

$$\text{Brand1}_i = \begin{cases} 1, & i\text{th alt. is a Casper} \\ 0, & i\text{th alt. is a HP} \\ 0, & i\text{th alt. is a Sony} \\ -1, & i\text{th alt. is a Toshiba} \end{cases}, \qquad \text{Brand2}_i = \begin{cases} 0, & i\text{th alt. is a Casper} \\ 1, & i\text{th alt. is a HP} \\ 0, & i\text{th alt. is a Sony} \\ -1, & i\text{th alt. is a Toshiba} \end{cases}$$

$$\text{Brand3}_i = \begin{cases} 0, & i\text{th alt. is a Casper} \\ 0, & i\text{th alt. is a HP} \\ 1, & i\text{th alt. is a Sony} \\ -1, & i\text{th alt. is a Toshiba} \end{cases},$$

and rename $\beta_{\text{Casper}}, \beta_{\text{HP}}, \beta_{\text{Sony}}, \beta_{\text{Toshiba}}$ as $\beta_{\text{Brand1}}, \beta_{\text{Brand2}}, \beta_{\text{Brand3}}, \beta_{\text{Brand4}}$, respectively, then the sum-contrast representation of the utility received from brand feature of the $i$the alternative becomes

$$\text{UBrand}_i = \beta_{\text{average}}^{\text{brand}} + \beta_{\text{Brand1}} \times \text{Brand1}_i + \beta_{\text{Brand2}} \times \text{Brand2}_i + \beta_{\text{Brand3}} \times \text{Brand3}_i, \quad (6)$$

where, let us recall, $\beta_{\text{average}}^{\text{brand}}$ is the average of utilities from all brands, and $\beta_{\text{Brand1}}, \beta_{\text{Brand2}}, \beta_{\text{Brand3}}$ are the excess utilities received from brands Casper (=Brand1), HP (=Brand2), Sony (=Brand3) over the average of utilities, $\beta_{\text{average}}^{\text{brand}}$, received from all brands. Finally, the excess utility from Toshiba (=Brand4) over the average $\beta_{\text{average}}^{\text{brand}}$ is later found from (5). In conjoint analysis, we prefer the sum-contrast representation of (6) rather than the treatment-contrast representation of (3).

Likewise, we introduce only one ± 1 dummy variable for `Processor` (because there are only two kinds of processors)

$$\texttt{Processor1}_i = \left\{ \begin{array}{ll} 1, & i\text{th alt. has Inter Core i5} \\ -1, & i\text{th alt. has Intel Core i7} \end{array} \right\},$$

three ± 1 dummy variables ($\texttt{Speed1}_i, \texttt{Speed2}_i, \texttt{Speed3}_i$) for `Speed` (because there are four different levels of speed), etc., and express utilities from processor, speed, etc. of the ith alternative in the sum-contrast format

$$\texttt{UProcessor}_i = \beta_{\text{average}}^{\text{processor}} + \beta_{\texttt{Processor1}} \times \texttt{Processor1}_i,$$
$$\texttt{USpeed}_i = \beta_{\text{average}}^{\text{speed}} + \beta_{\texttt{Speed1}} \times \texttt{Speed1}_i + \beta_{\texttt{Speed2}} \times \texttt{Speed2}_i + \beta_{\texttt{Speed3}} \times \texttt{Speed3}_i,$$

and so on. Finally, the total utility in (1) derived by the survey taker from the ith alternative becomes the *multiple regression model*

$$\begin{aligned} -\texttt{RANK}_i = \beta_0 &+ \beta_{\texttt{Brand1}} \times \texttt{Brand1}_i + \beta_{\texttt{Brand2}} \times \texttt{Brand2}_i + \beta_{\texttt{Brand3}} \times \texttt{Brand3}_i \\ &+ \beta_{\texttt{Processor1}} \times \texttt{Processor1}_i \\ &+ \beta_{\texttt{Speed1}} \times \texttt{Speed1}_i + \beta_{\texttt{Speed2}} \times \texttt{Speed2}_i + \beta_{\texttt{Speed3}} \times \texttt{Speed3}_i \\ &+ \beta_{\texttt{Memory1}} \times \texttt{Memory1}_i \\ &+ \beta_{\texttt{HardDisk1}} \times \texttt{HardDisk1}_i \\ &+ \beta_{\texttt{Screen1}} \times \texttt{Screen1}_i \\ &+ \beta_{\texttt{Price1}} \times \texttt{Price1}_i + \beta_{\texttt{Price2}} \times \texttt{Price2}_i + \beta_{\texttt{Price3}} \times \texttt{Price3}_i + \varepsilon_i, \end{aligned} \tag{7}$$

where $\beta_0 \equiv \beta_{\text{average}}^{\text{brand}} + \beta_{\text{average}}^{\text{processor}} + \ldots + \beta_{\text{average}}^{\text{price}}$, and we have

$$\beta_{\texttt{Brand4}} = -\beta_{\texttt{Brand1}} - \beta_{\texttt{Brand2}} - \beta_{\texttt{Brand3}},$$
$$\beta_{\texttt{Processor2}} = -\beta_{\texttt{Processor1}},$$
$$\beta_{\texttt{Speed4}} = -\beta_{\texttt{Speed1}} - \beta_{\texttt{Speed2}} - \beta_{\texttt{Speed3}},$$
$$\beta_{\texttt{Memory2}} = -\beta_{\texttt{Memory1}},$$
$$\beta_{\texttt{HardDisk2}} = -\beta_{\texttt{HardDisk1}},$$
$$\beta_{\texttt{Screen2}} = -\beta_{\texttt{Screen1}},$$
$$\beta_{\texttt{Price4}} = -\beta_{\texttt{Price1}} - \beta_{\texttt{Price2}} - \beta_{\texttt{Price3}}.$$

Part-worths: The regression coefficients represent conjoint measures of utility called part-worths. Part-worths reflect the strength of individual consumer preferences for each level of each laptop feature. For example,

$$\beta_{\texttt{Brand1}}, \quad \beta_{\texttt{Brand2}}, \quad \beta_{\texttt{Brand3}}, \quad \text{and} \quad \beta_{\texttt{Brand4}}$$

are the part worths for the brands `Casper`, `HP`, `Sony`, and `Toshiba`, respectively, and each quantifies how much the corresponding brand adds to the total utility of the laptop produced by that brand.

Part importance: Using the part worths, we can quantify the importance of each laptop feature in Hande's laptop preferences. Each feature's importance is thought to be proportional to the range of its part worths. If a feature has a wide range of part worths, then the total utility of each alternative can start from a very low or very high value depending on the level of that feature for that alternative. If, however, a feature has a narrow range of part worths, then the contribution of that feature to the total utilities of the alternatives will be in similar amounts and play an unimportant role in the preference rankings of the alternatives. For example, the importance of the laptop brand in Hande's preferences is proportional to the range of the brand part worths $\beta_{\texttt{Brand1}}, \beta_{\texttt{Brand2}}, \beta_{\texttt{Brand3}}$, and $\beta_{\texttt{Brand4}}$ and is calculated by

$$\texttt{Range}_{\texttt{Brand}} := \max\left\{\beta_{\texttt{Brand1}}, \beta_{\texttt{Brand2}}, \beta_{\texttt{Brand3}}, \beta_{\texttt{Brand4}}\right\} - \min\left\{\beta_{\texttt{Brand1}}, \beta_{\texttt{Brand2}}, \beta_{\texttt{Brand3}}, \beta_{\texttt{Brand4}}\right\}.$$

The part-worth ranges $\texttt{Range}_{\texttt{Processor}}, \texttt{Range}_{\texttt{Speed}}, \texttt{Range}_{\texttt{Memory}}, \texttt{Range}_{\texttt{HardDisk}}, \texttt{Range}_{\texttt{Screen}}$, and $\texttt{Range}_{\texttt{Price}}$ are calculated similarly. The part importance of every laptop feature is then defined as the part worth range of that

feature normalized by the sum of the part worth ranges of all laptop features. For example, the part importance of `Brand` is defined as

$$I_{\text{Brand}} := \frac{\text{Range}_{\text{Brand}}}{\text{Range}_{\text{Brand}} + \text{Range}_{\text{Processor}} + \text{Range}_{\text{Speed}} + \text{Range}_{\text{Memory}} + \text{Range}_{\text{HardDisk}} + \text{Range}_{\text{Screen}} + \text{Range}_{\text{Price}}}.$$

The part importances of the features are between zero and one and add up to one. The closer is the part importance of a feature to one, the higher is the impact of that feature to the preference rankings of that individual.

The conjoint analysis of Hande's preference rankings in `R` yields the following results:

```
> response = read.table("laptops-hande.csv", head=TRUE, sep=",")
> head(response)
  Alternative    Brand        Processor    Speed Memory HardDisk   Screen    Price RANK
1           1       HP Intel Core i5  1.8 GHz   4 GB   500 GB 15.6 in 1500 USD    8
2           2     Sony Intel Core i5  2.4 GHz   8 GB   750 GB 17.3 in 1500 USD    1
3           3   Casper Intel Core i5  2.6 GHz   4 GB   500 GB 17.3 in 1700 USD    6
4           4     Sony Intel Core i7  2.6 GHz   8 GB   500 GB 15.6 in 3500 USD   13
5           5     Sony Intel Core i5  2.2 GHz   4 GB   500 GB 17.3 in 2000 USD   10
6           6     Sony Intel Core i7  1.8 GHz   4 GB   750 GB 15.6 in 1700 USD   11
> response = response[,-1] ## Remove Alternative column
> response$RANK = - response$RANK ## more natural for more attractive alternative to have
      positive part worths
> options(contrasts=c('contr.sum','contr.poly')) ## switch to sum contrast
> (res = lm(RANK ~ ., response)) ## calculate the part worths

Call:
lm.default(formula = RANK ~ ., data = response)

Coefficients:
(Intercept)          Brand1          Brand2          Brand3       Processor1          Speed1
 -8.500e+00      -1.250e+00       1.250e+00      -2.500e-01       8.750e-01      -2.000e+00
     Speed2          Speed3         Memory1        HardDisk1          Screen1          Price1
  7.500e-01       5.000e-01      -2.125e+00      -8.327e-17      -1.250e-01       2.750e+00
     Price2          Price3
  2.750e+00       5.000e-01
```

Figure 1 displays the part importance and part worths for Hande. `Price` is the most important laptop feature accounting for 43% of her utility range. The second most important feature is `Memory` (21%) and then `Speed` (14%), `Brand`(12%), `Processor` (9%), `Screen` (1%) follow. `HardDisk` is the least important laptop feature. Bear, however, in mind that the part importance and part worth of feature depend on the range of feature values in the data set. If the levels of a feature cover a very narrow range of the feature values, then the utility derived from that feature will not change drastically with changing feature level.

It will be interesting to check how good the part worths obtained by the conjoint analysis will predict Hande's preference ratings for the laptops on sale in Table 1. First of all, we predict the utility of each alternative by summing up the part worths for each feature level. For example, we predict that Hande obtains from Alternative 1 in Table 1 a total utility of

$$\widehat{\beta}_{\text{Brand4}} + \widehat{\beta}_{\text{Processor2}} + \widehat{\beta}_{\text{Speed3}} + \widehat{\beta}_{\text{Memory2}} + \widehat{\beta}_{\text{HardDisk2}} + \widehat{\beta}_{\text{Screen2}} + \widehat{\beta}_{\text{Price4}}.$$

In `R`, we can calculate the predicted utilities and sort them to predict the preference rankings as follows:

```
> dd = read.table("laptops-hepsiburada-hande.csv", header=TRUE, sep=",")
> (RANK.pred = -predict(res, newdata=dd[,-c(1,9)])) ## estimate scores
     1       2       3       4       5       6
12.375   8.375   5.125   7.875   9.375   7.875
> RANK.pred[order(RANK.pred)] = seq_along(RANK.pred) ## rank the scores
> (RANKS = data.frame(Alternative=dd$Alternative, RANK.obs = dd$RANK, RANK.pred = RANK.pred))
  Alternative RANK.obs RANK.pred
1           1        6         6
2           2        2         4
3           3        1         1
```
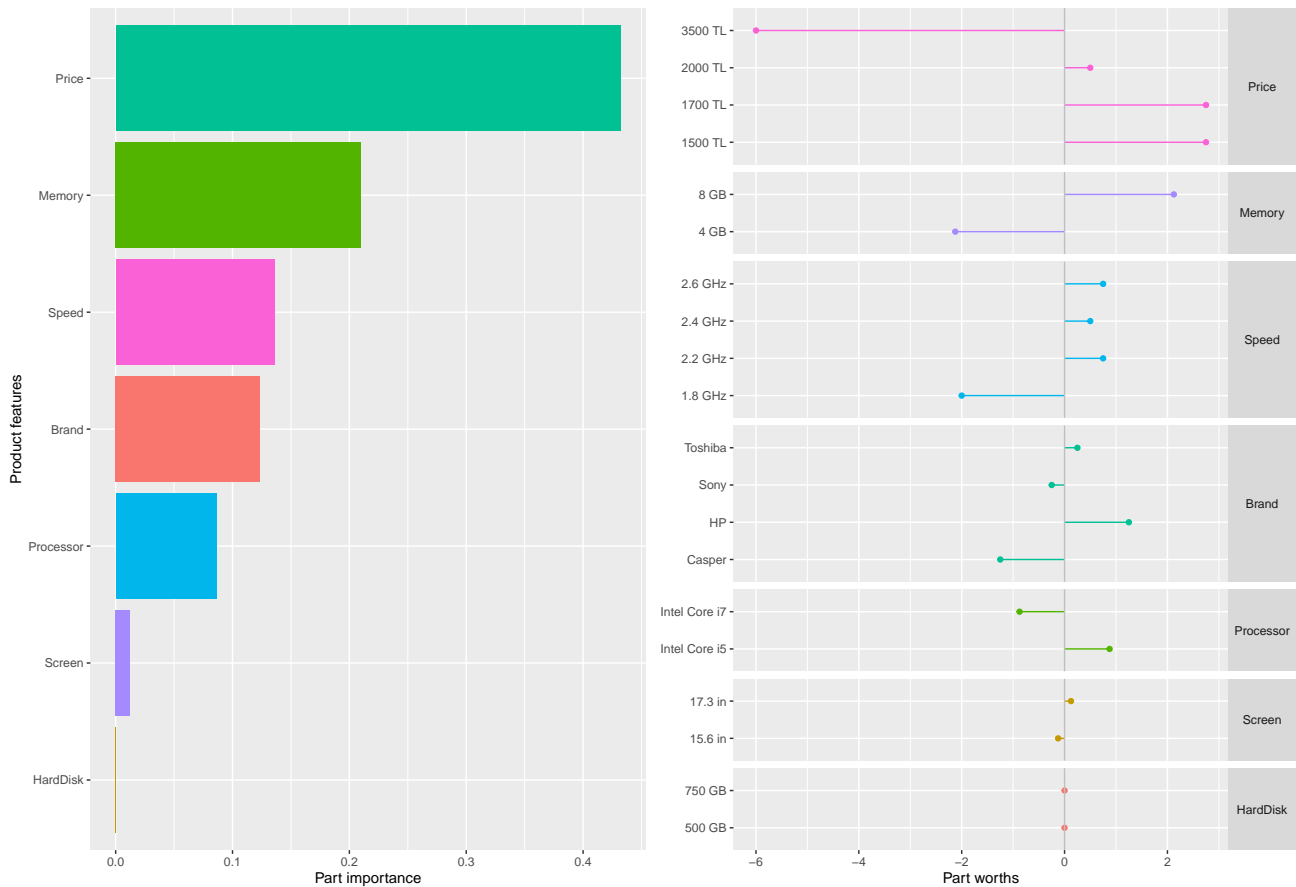
Figure 1: Part importance (left), and part worths (right) for Hande

4	4	3	2
5	5	5	5
6	6	4	3

The plot of `RANK.obs` versus `RANK.pred` in Figure 2 shows that they are in quite good agreement. If the conjoint analysis can be applied to 600-800 customers, then using their part worths one can calculate the market share for a new product, the cannibalization effect, and market shares of the competitors that the new product attracts, and so on.

1. To begin, rank the laptops listed in Tables 2 and 1.

   (a) Run a conjoint analysis on your survey data. Determine the part importance and part worths. Plot them and print in a *single* `pdf` file.

   See Part importance for each participant in the survey.

   (b) Predict your preference ratings for the laptops in Table 1 using the part worths calculated in (1a). Plot observed ranks against predicted ranks. Print the plots into a *single* `pdf` file. Are they in good agreement? Comment.

   (c) Casper considers upgrading its Alternative 4 laptop. The company wants to run some what-if analysis for this purpose.

      i. What is Casper's market share? What are the market shares of each Casper laptop in Table 1? See Figures 3 and 4.

      ii. What is Casper's substitution rate for each of the other laptop brands in Table 1? (If, for example, Sony is unavailable, what is the percentage of customers who choose Casper laptop?) See Figure 5.
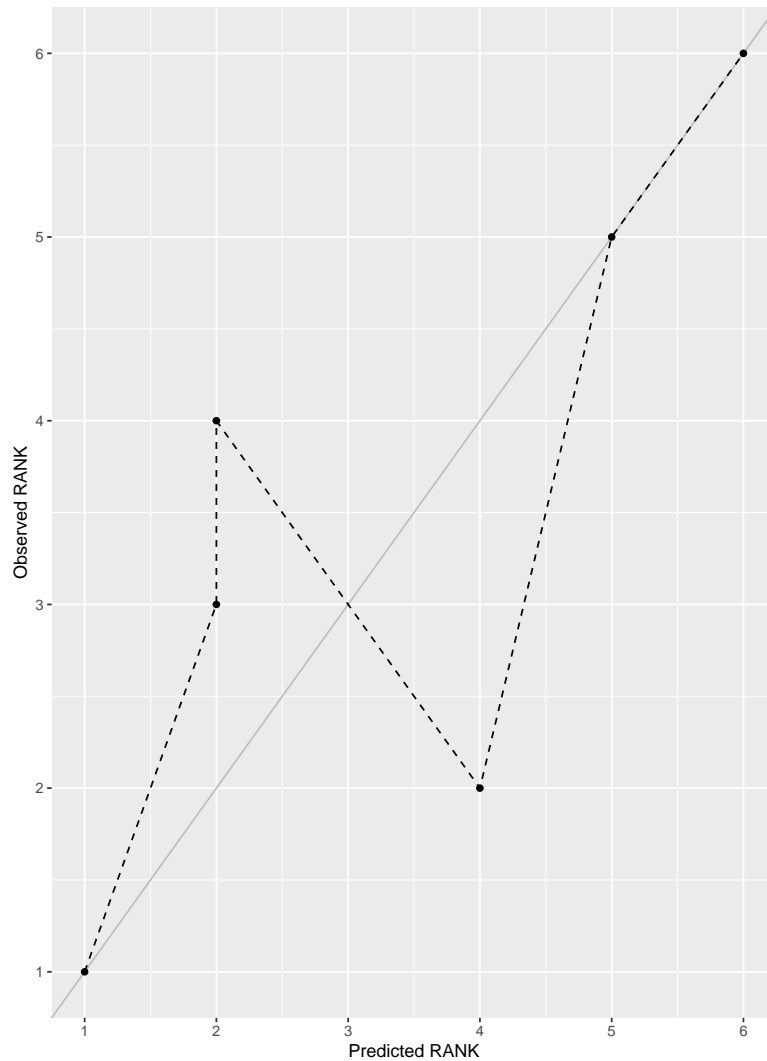
6

Figure 2: Observed versus predicted preference rankings of the laptops in Table 1

iii. How much does Casper Alternative 4 laptop's overall market share increase if Casper upgrades a feature one level while keeping other features unchanged? Answer this question for each feature of the laptop.

See Figure 6. On the left hand side, the largest increase in the new product's market share is $0.16 - 0.03 = 0.12$ if the existing product's processor is upgraded from `Intel Core i5` to `Intel Core i7`. This, however, cannibalizes Casper's other product's market share in part; see the figure on the right-hand side. Even though Casper 3's market share drops by 7%, the total market share of Casper products increases by 6%.

The upgrade will most likely come along with an increase in the price of the new Casper product. The simulation has to be run one more time with possible price increases, and the profits should be calculated and compared. An interesting observation on the left-hand picture in Figure 6 is that, even though the market share of Casper 4 is wiped out when its price increases to 1700 or 2000 USDs, the market share is restored when it is increased even more to 3500 USD. This can be spurious, or it may indicate some irrational shopping habit.

iv. What percentage of demand for Casper Alternative 3 model will be cannibalized by each of new models in 1(c)iii?
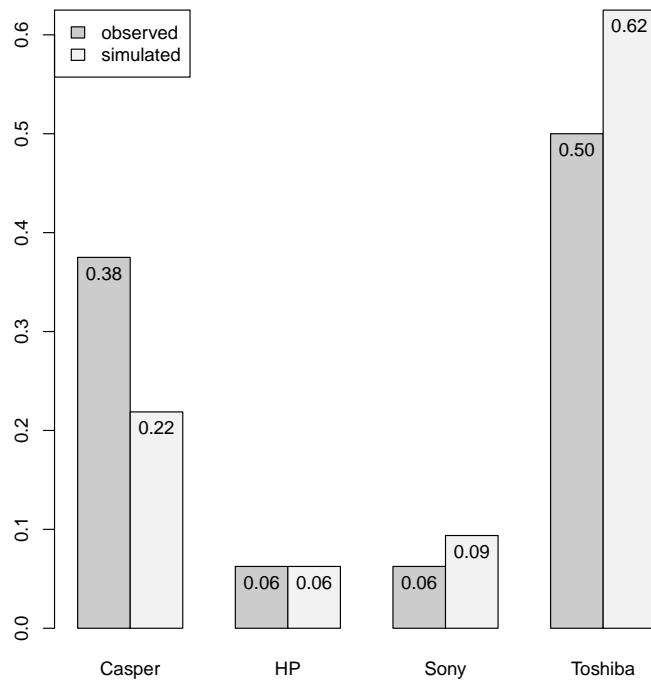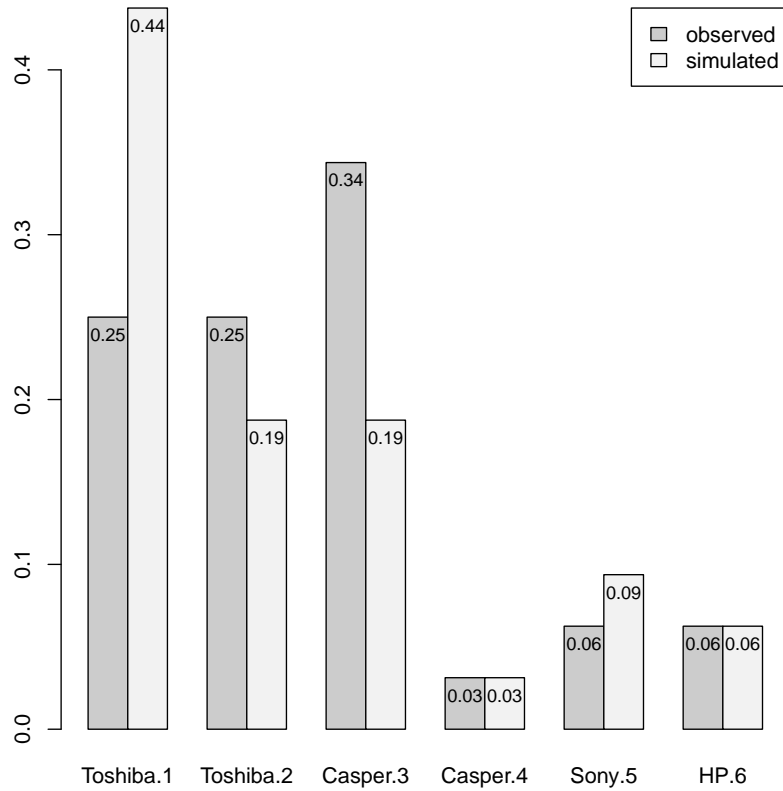
7

Figure 3: Market shares

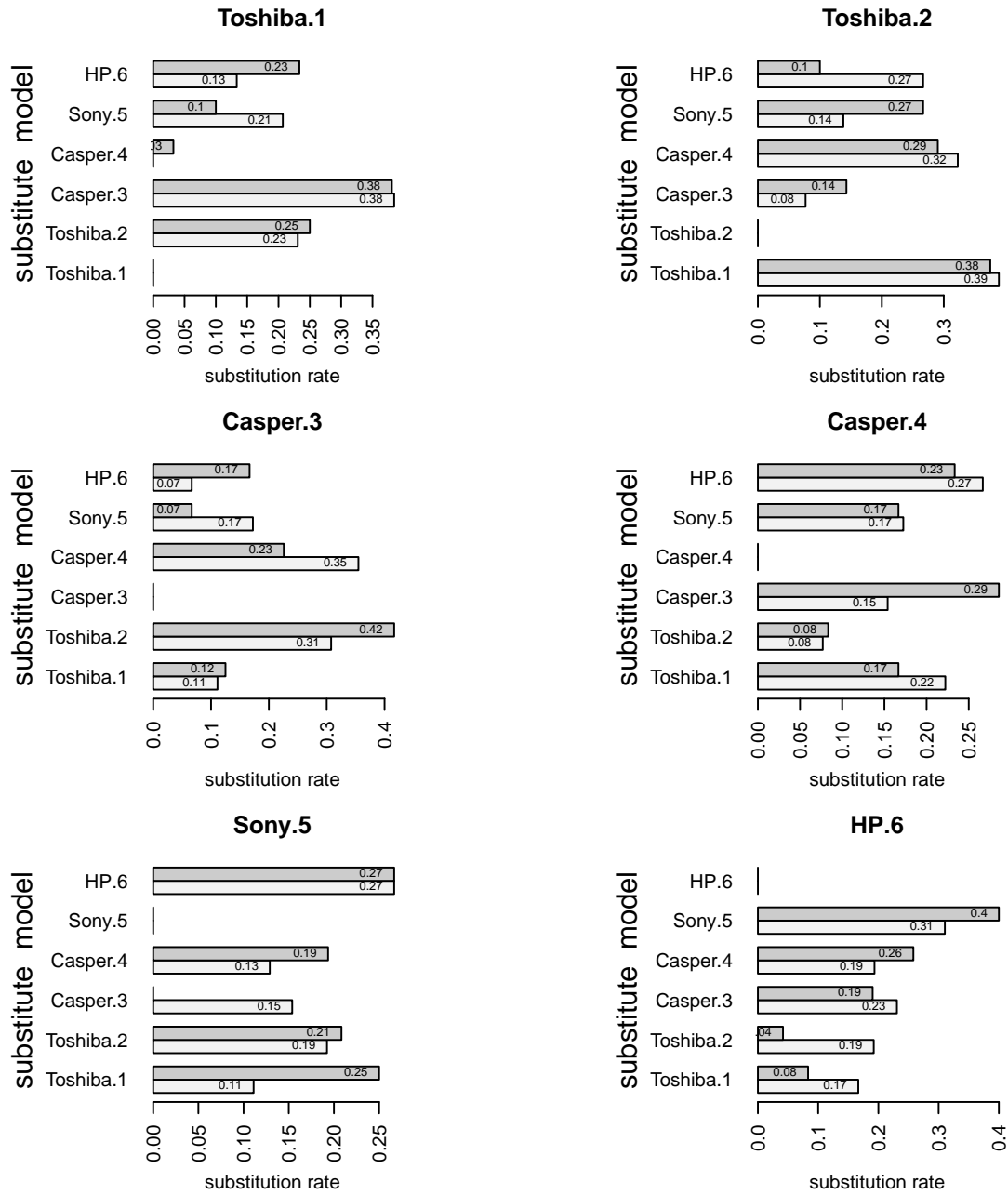Figure 4: Market shares for each brand's each model

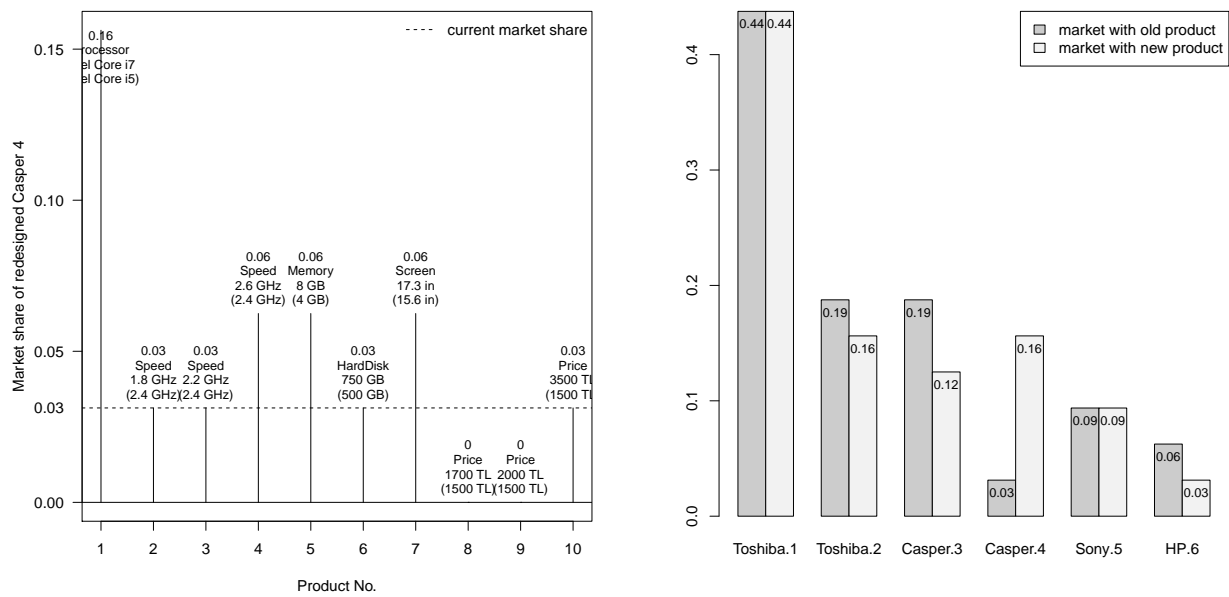Figure 5: Observed (dark) and simulated substitution rates of models

Figure 6: On the left: market rates of new Casper 4 products. For each of ten possible new configurations, each vertical bar shows the simulated new market share, the new product feature, and the old feature in parenthesis. On the right: simulated the new market shares of all products after Casper 4 is upgraded.