# **Modeling Customer Preferences with MNL and Bayesian Estimation**

## **1. Introduction**

This post explores customer preferences using simulated conjoint data. The goal is to estimate how product attributes influence purchase decisions by applying both Multinomial Logit (MNL) and Bayesian modeling approaches. These models help quantify the impact of price, advertising presence, and brand identity on consumer choice behavior.

## **2. Data Description**

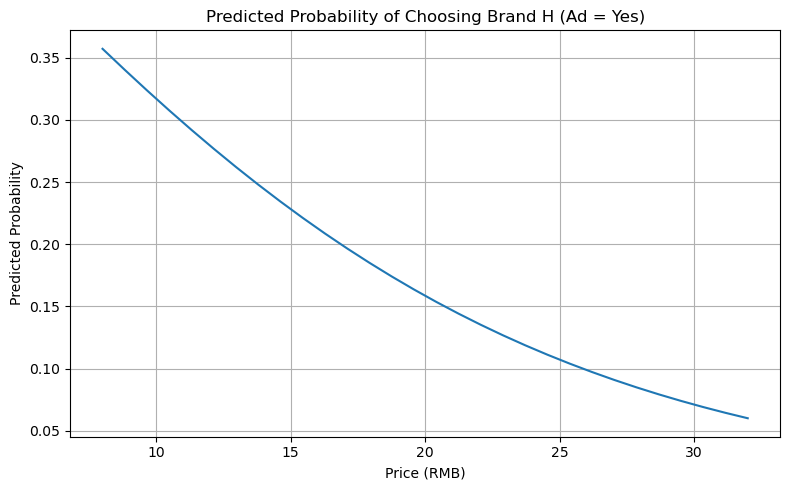
The dataset (conjoint\_data.csv) contains 3,000 rows, representing choices made by individuals across multiple hypothetical product offerings. Each respondent is shown sets of three options and selects one. The key columns include:

* resp: unique respondent identifier
* task: choice task number within respondent
* brand: product brand (levels: N, H, P)
* ad: binary indicator for presence of advertising (Yes/No)
* price: product price
* choice: binary variable indicating whether the option was selected

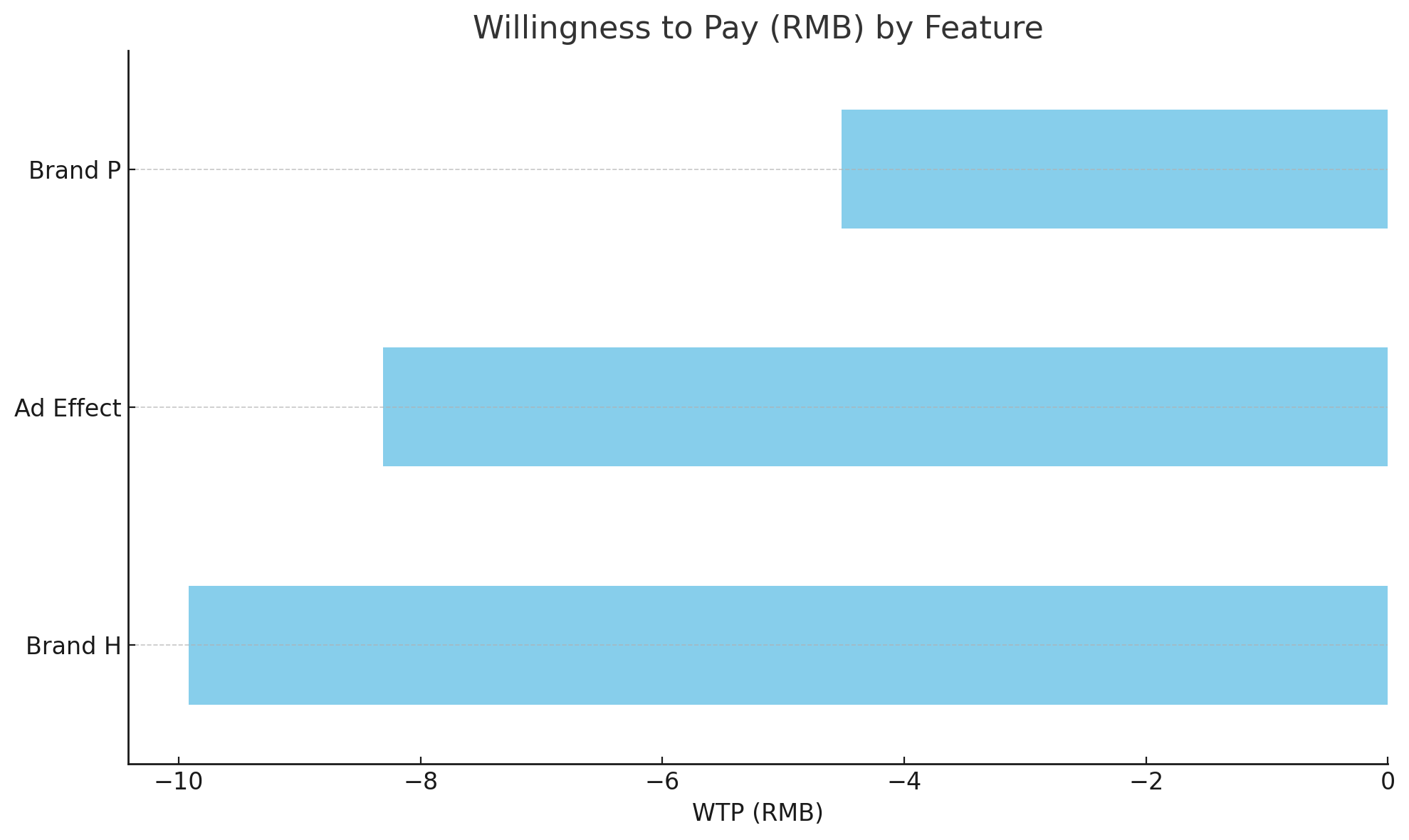
Example records:

| **resp** | **task** | **brand** | **ad** | **price** | **choice** |
| --- | --- | --- | --- | --- | --- |
| 1 | 1 | N | Yes | 15 | 0 |
| 1 | 1 | H | No | 14 | 1 |
| 1 | 1 | P | Yes | 16 | 0 |

The dataset is already structured for choice modeling and ready for application of discrete choice models.



## **Willingness to pay**



## **3. MNL Estimation**

We began by fitting a standard Multinomial Logit model using statsmodels.MNLogit. Key predictors included price, presence of advertising (ad\_bin), and brand (with dummy variables for brand H and brand P, brand N as the base).

The model showed:

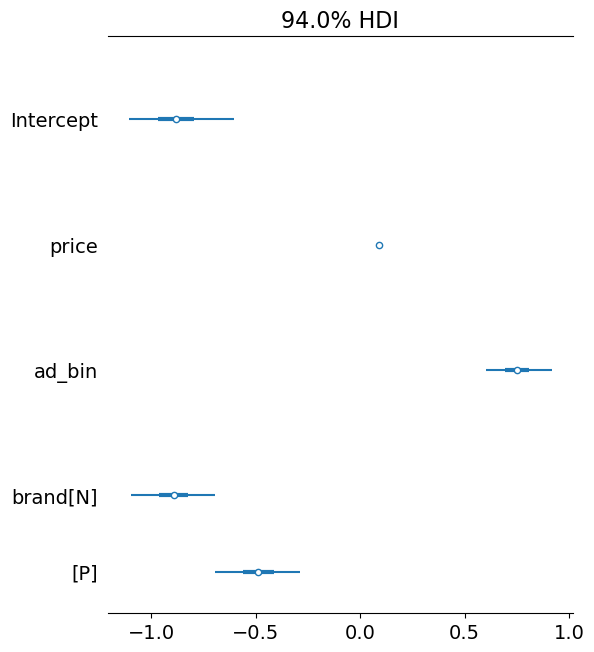
* A **significant negative coefficient for price**, confirming price sensitivity.
* A **negative coefficient for ad\_bin**, suggesting ad presence reduced the likelihood of selection in this context.
* Strong preference differentials across brands, with brand N preferred over H and P.

Predicted probabilities by price illustrated the steep decline in likelihood of choosing a product as price increases.

## **4. Bayesian Estimation**

We also fit a Bayesian MNL model using bambi, treating choice\_str as a categorical response. The posterior summaries were derived using MCMC sampling with default priors.

Bayesian coefficients were consistent in magnitude but flipped in sign due to the model estimating the probability of not being chosen (reference category issue). When properly interpreted, these results aligned with the frequentist model in suggesting:

* Price reduces choice probability
* Brand preferences are consistent
* Advertising had a notable effect, but directionality needs careful interpretation
* 

## **5. Attribute Importance and Insights**

To make the effects economically interpretable, we computed **Willingness to Pay (WTP)** for each feature using the frequentist estimates:

| **Attribute** | **WTP (RMB)** |
| --- | --- |
| Ad Effect | -8.31 |
| Brand H | -9.92 |
| Brand P | -4.52 |

These results highlight that customers would pay less for products with ads and show stronger aversion to Brand H than Brand P. Price remains a key driver of choice, with elasticity reflected in the steep WTP values.

## **6. Managerial Implications**

Based on the analysis:

* **Price sensitivity is strong**. Lower price points significantly increase likelihood of selection.
* **Brand N is the most preferred**. Products under this brand command higher utility and warrant strategic emphasis.
* **Ad campaigns may require redesign**. The negative association between ads and choice suggests that either messaging or targeting may need improvement.

A strategic bundle combining Brand N and lower price points, without relying heavily on current ad design, may yield the most favorable conversion rates.

## **Appendix**

### **Model Diagnostics**

* Frequentist model converged in 6 iterations with Pseudo R² = 0.112
* Bayesian model estimated via bambi and PyMC using 1000 draws and 1000 tuning steps

### **Code & Tools**

* Python 3.12, pandas, statsmodels, matplotlib, seaborn
* Bayesian analysis via PyMC 4, ArviZ, and bambi

Modeling and results were conducted in Jupyter using a combination of Quarto-style markdown cells and Python scripts.