Advanced Econometrics
06 Limited Dependent Variable Models

Eduard Brüll

Fall 2025

Lecture 6

Advanced Econometrics

6. Limited Dependent Variable Models
   - 6.1 The Linear Probability Model
   - 6.2 Logit
   - 6.3 Random Utility & Multinomial Logit
   - 6.4 Other Limited Dependent Variable Models

Literature: Greene Chapter 17

6.1: Linear Probability Model

The Linear Probability Model

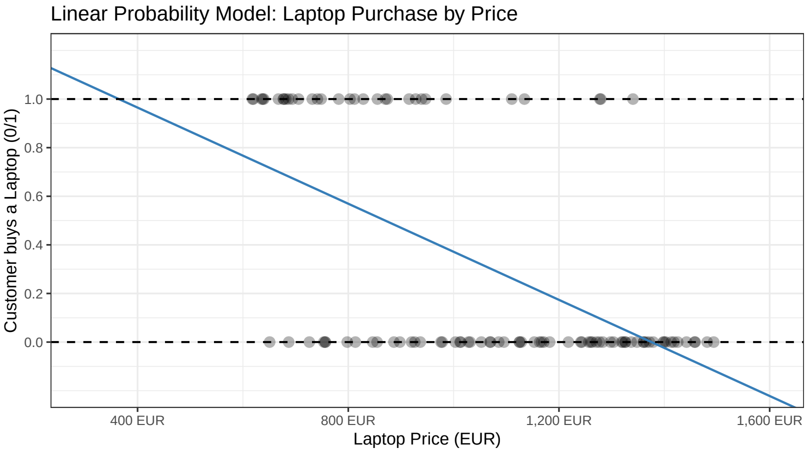
Example: A Model of Laptop PC Demand

- Outcome: $Y_i = 1$ if customer buys, 0 otherwise.
- Explanatory variable: laptop price (Price_i).
- Model:

$$Y_i = \beta_1 + \beta_2 \,\text{Price}_i + \varepsilon_i$$

- Predictions $\hat{Y}_i$ are interpreted as purchase probabilities.

The Linear Probability Model in a Scatterplot



Interpreting the LPM (Binary Outcome)

Key idea

For a binary $Y_i \in \{0, 1\}$,

$$\mathbf{E}[Y_i \mid \text{Price}_i] \;=\; P(Y_i = 1 \mid \text{Price}_i) \;=\; \beta_1 + \beta_2 \, \text{Price}_i.$$

Thus fitted values $\hat{Y}_i$ are purchase probabilities.

- **Marginal effect:** β_2 is the change in purchase probability from a 1€ price change.
- **Example (generic):** A price increase of Δp lowers the purchase probability by $\beta_2 \cdot \Delta p$ percentage points (if $\beta_2 < 0$).
- **Multiple regressors:** With more features X_i (e.g., specs, promos), each β_k is the c.p. change in probability per unit of X_{ik}.

- **Conveniences**
  - Easy to estimate with standard OLS
  - Coefficients have a **direct interpretation as changes in probabilities**.
  - Standard tools (tests, confidence intervals) apply.
- **Caveats**
  - **Heteroskedasticity** is inherent: $\text{var}(Y_i \mid X_i) = p_i(1 - p_i)$.
    $\Rightarrow$ Use heteroskedasticity-robust standard errors.
  - Predicted values can lie **outside the $[0, 1]$ range**.
  - R^2 **is not informative** with binary outcomes.
  - **But does our linearity assumption for OLS apply?**

6.2: Linear Logit Regression

From Probability to Odds

- The probability of a purchase is $p_i = P(y_i = 1 \mid X_i)$.
- Probabilities are bounded: $p_i \in (0, 1)$. Hard to model linearly.
- **Econometric trick:** work with **odds**.
- Define the **odds** of purchase:

$$\text{odds}(p_i) = \frac{p_i}{1 - p_i}.$$

- Interpretation:
  - If $\text{odds}(p_i) = 2$, purchase is twice as likely as non-purchase.
  - If $\text{odds}(p_i) = 1$, equally likely.
  - If $\text{odds}(p_i) = 0.5$, purchase is half as likely.
- Odds take values in $(0, \infty)$: still not convenient for linear models!

From Odds to Log-Odds (the Logit)

- The odds $\frac{p_i}{1-p_i}$ take values in $(0, \infty)$:

$$p_i \in (0,1) \;\Rightarrow\; \frac{p_i}{1-p_i} \in (0, \infty).$$

- Taking logs removes the positive-only restriction, since

$$\log(x) \to -\infty \text{ as } x \to 0^+, \quad \text{and } \log(x) \to \infty \text{ as } x \to \infty,$$

  the transformed variable now spans the whole real line.

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) \in (-\infty, \infty).$$

- This makes it possible to model the log-odds as a linear function of covariates:

$$\log\left(\frac{p_i}{1-p_i}\right) = X_i\beta.$$

Linear Model for Log-Odds

- Start from the linear relationship in log-odds:

$$\log\left(\frac{p_i}{1-p_i}\right) = X_i\beta.$$

- Exponentiate both sides to remove the log:

$$\frac{p_i}{1-p_i} = \exp(X_i\beta).$$

- Solve for p_i:

$$p_i = (1-p_i)\exp(X_i\beta)$$
$$p_i + p_i\exp(X_i\beta) = \exp(X_i\beta)$$
$$p_i = \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)}.$$

Binary Choice: Laptop Demand Example

- In our laptop demand example, we observe for each customer *i*:

$$y_i = \begin{cases} 1 & \text{if laptop purchased} \\ 0 & \text{otherwise.} \end{cases}$$

- Purchase probability:

$$P(y_i = 1 \mid X_i) = p_i = \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}.$$

- Non-purchase probability:

$$P(y_i = 0 \mid X_i) = 1 - p_i = \frac{1}{1 + \exp(X_i\beta)}.$$

- Each observation follows a **Bernoulli** distribution, just like in our coin example last week!

$$P(y_i \mid X_i) = p_i^{y_i}(1 - p_i)^{1-y_i}.$$

From Coins to Consumers

- Recall our first MLE example: flipping a coin n times, each flip $Y_i \sim \text{Bernoulli}(p)$.

$$L(p \mid Y_1, \ldots, Y_n) = \prod_i p^{Y_i}(1-p)^{1-Y_i}$$

- Each observation Y_i was binary, either "success" (*head*) or "failure" (*tail*).
- Now, our **laptop purchase** example works the same way:
  - Each customer either buys ($Y_i = 1$) or doesn't ($Y_i = 0$).
  - The probability of success (p_i) is no longer constant, but depends on covariates X_i.
- **Key idea:** Binary choice models (logit/probit) generalize the Bernoulli coin model by letting p_i vary with $X_i\beta$.

Step 1: Likelihood for the Logit Model

- Each observation $y_i \in \{0, 1\}$ follows a Bernoulli distribution:

$$P(y_i \mid X_i) = p_i^{y_i}(1 - p_i)^{1-y_i}.$$

- Assuming independence across customers:

$$L(\beta) = \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{1-y_i}.$$

- Substitute the logit probability to get a **likelihood expressed in terms of β.**

$$p_i = \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)} :$$

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)} \right)^{y_i} \left(\frac{1}{1 + \exp(X_i\beta)} \right)^{1-y_i} .$$

Step 2: Log-Likelihood Derivation

- Start from the likelihood (from Step 1):

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(X_i\beta)}{1+\exp(X_i\beta)} \right)^{y_i} \left(\frac{1}{1+\exp(X_i\beta)} \right)^{1-y_i}.$$

- Take logs (turn product into sum):

$$\ell(\beta) = \sum_{i=1}^{n} \left[y_i \log\left(\frac{\exp(X_i\beta)}{1+\exp(X_i\beta)} \right) + (1-y_i)\log\left(\frac{1}{1+\exp(X_i\beta)} \right) \right].$$

- Simplify each term:

$$\begin{aligned} \log\left(\frac{\exp(X_i\beta)}{1+\exp(X_i\beta)} \right) &= \log[\exp(X_i\beta)] - \log[1+\exp(X_i\beta)] \\ &= X_i\beta - \log(1+\exp(X_i\beta)), \\ \log\left(\frac{1}{1+\exp(X_i\beta)} \right) &= -\log(1+\exp(X_i\beta)). \end{aligned}$$

- Plug back in and collect terms:

$$\ell(\beta) = \sum_{i=1}^{n} \Big[y_i X_i\beta - \log(1+\exp(X_i\beta)) \Big].$$

Step 3: First-Order Condition for β

- Start from the log-likelihood:

$$\ell(\beta) = \sum_{i=1}^{n} \Big[y_i X_i \beta - \log(1 + \exp(X_i \beta)) \Big].$$

- Differentiate term by term, applying the chain rule to the second part:

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{n} \Bigg[y_i X_i - \underbrace{\frac{1}{1 + \exp(X_i \beta)}}_{\text{outer derivative of } \log(\cdot)} \underbrace{\exp(X_i \beta)}_{\text{derivative of } \exp(\cdot)} \underbrace{X_i}_{\text{derivative of } X_i \beta} \Bigg].$$

- Collecting terms, the gradient (score) of the log-likelihood is:

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{n} \left[y_i X_i - \frac{\exp(X_i \beta)}{1 + \exp(X_i \beta)} X_i \right]$$

Step 4: The Score (First-Order Condition)

- Start from the score:

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{n} X_i \left[y_i - \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)} \right].$$

- Recall p_i:

$$p_i = \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)}.$$

- Then the **score function** can be written compactly as:

$$s(\beta) = \sum_{i=1}^{n} X_i(y_i - p_i).$$

- The **first-order condition (FOC)** for the MLE is obtained by setting the score to zero:

$$s(\beta) = 0.$$

Why We Can't Solve the Logit MLE Analytically

- The first-order condition is

$$s(\beta) = \sum_{i=1}^{n} X_i(y_i - p_i) = 0, \qquad p_i = \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}.$$

- Substituting p_i gives

$$\sum_{i=1}^{n} X_i \left[y_i - \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)} \right] = 0.$$

- The equation is **nonlinear in** β because β appears both inside and outside the nonlinear function $\exp(X_i\beta)$.
- Rearranging terms does not isolate β: it enters in a sum of logistic functions across observations, so there is no algebraic way to express β as a finite combination of elementary functions.
- Therefore, $\hat{\beta}$ must be found by **numerical optimization**.

Step 5: Solving for $\hat{\beta}$ (Numerical Optimization)

- The first-order condition

$$s(\beta) = \sum_i X_i(y_i - p_i) = 0$$

  cannot be solved in closed form.
- We must **iterate numerically** until the score is close to zero.
- The usual algorithm is the **Newton–Raphson method**:

$$\beta^{(t+1)} = \beta^{(t)} + H^{-1}(\beta^{(t)})\, s(\beta^{(t)}),$$

  where $H(\beta)$ is the **Hessian (matrix of second derivatives)** of the log-likelihood.
- Intuition: move in the direction of the gradient (score) scaled by the local curvature.
- In practice, software (e.g. ‘glm(family = binomial)‘ in R) does this internally.

Step 6: Fisher Information for the Logit Model

- Recall the log-likelihood:

$$\ell(\beta) = \sum_i \left[y_i X_i' \beta - \log(1 + e^{X_i' \beta}) \right].$$

- Score (gradient):

$$s(\beta) = \sum_i X_i (y_i - p_i), \qquad p_i = \frac{e^{X_i' \beta}}{1 + e^{X_i' \beta}}.$$

- Hessian (matrix of 2nd derivatives):

$$H(\beta) = \frac{\partial^2 \ell(\beta)}{\partial \beta \, \partial \beta'} = -\sum_i p_i (1 - p_i) \, X_i X_i'.$$

- The **Observed Information** is $J(\beta) = -H(\beta)$.
- Taking expectations yields the **Fisher Information**:

$$I(\beta) = E[J(\beta)] = \sum_i p_i (1 - p_i) \, X_i X_i'.$$

Step 7: Asymptotic Properties of $\hat{\beta}$ (MLE Theory Link)

- From Lecture 5: for any regular MLE $\hat{\theta}$,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\big(0, I(\theta_0)^{-1}\big).$$

- Applying this to the logit model:

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N\big(0, I(\beta_0)^{-1}\big), \quad I(\beta_0) = \sum_i p_i(1 - p_i)X_i X_i'.$$

- The estimated covariance matrix:

$$\widehat{\text{Var}}(\hat{\beta}) = I(\hat{\beta})^{-1}.$$

- All standard MLE properties (consistency, asymptotic normality, efficiency) apply.

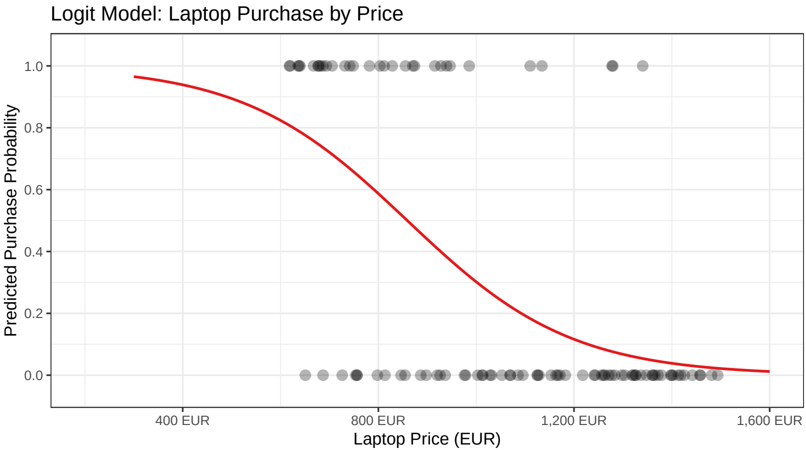
Interpreting the Variance Structure

- $p_i(1-p_i)$ is the **variance of a Bernoulli(p_i)** outcome, analogous to the constant error variance σ^2, but varying across observations.
- $X_i X_i'$ is the familiar **outer product of regressors**.
- Therefore, the Fisher information

$$I(\beta_0) = \sum_i p_i(1-p_i)X_i X_i'$$

  can be viewed as a **weighted OLS error variance matrix**, where each observation's contribution depends on its outcome uncertainty.
- Observations with $p_i \approx 0.5$ provide the most information; those with p_i near 0 or 1 contribute little.

The Logit Regression Line for our Example



Interpreting Logit Coefficients

- **Linear effect on log-odds:**

$$\frac{\partial}{\partial x_{ij}} \log\left(\frac{p_i}{1-p_i}\right) = \beta_j$$

Each β_j shifts the log-odds linearly.

- **Effect on odds:**

$$\frac{\text{odds}(x_{ij}+1)}{\text{odds}(x_{ij})} = e^{\beta_j}$$

A one-unit increase in x_{ij} multiplies the odds by $\exp(\beta_j)$.

Interpreting Logit Coefficients

- **Marginal effect on probability:**

$$\frac{\partial p_i}{\partial x_{ij}} = p_i(1 - p_i)\beta_j$$

  The effect on p_i depends on its current level.

- **In practice:** To compute marginal effects for your sample, plug in each observation's predicted $\hat{p}_i$ and take the average:

$$\text{AME}_j = \frac{1}{N}\sum_i \hat{p}_i(1 - \hat{p}_i)\beta_j$$

- **Interpretation:** Effects are largest around $\hat{p}_i = 0.5$ and shrink near 0 or 1.

Sidenote: Probit

- The **probit model** assumes that the latent index $Y_i^* = X_i\beta + \varepsilon_i$ has **normally distributed errors:** $\varepsilon_i \sim N(0,1)$.
- The observed binary outcome is

$$Y_i = \begin{cases} 1 & \text{if } Y_i^* > 0, \\ 0 & \text{otherwise.} \end{cases}$$

- Hence, the choice probability is

$$P(Y_i = 1 \mid X_i) = P(\varepsilon_i > -X_i\beta) = \Phi(X_i\beta),$$

  where $\Phi(\cdot)$ is the standard normal CDF.
- **Interpretation:** Similar to logit, but with a different link function:

$$\text{Logit: } p_i = \frac{e^{X_i\beta}}{1 + e^{X_i\beta}} \qquad \text{vs.} \qquad \text{Probit: } p_i = \Phi(X_i\beta).$$

- Both yield very similar fitted probabilities in practice; the probit has **slightly thinner tails.**

6.3: Random Utility Model & Multinomial Logit

Random Utility Model (RUM)

- Individual i chooses one option $j \in \{1, \ldots, J\}$.
- Utility is decomposed into

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \quad V_{ij} = X_{ij}'\beta + \alpha_j.$$

- Choice rule:

$$P(y_i = j) = \Pr\big(U_{ij} \geq U_{ik},\ \forall k\big)\ .$$

- Binary choice ($J = 2$) collapses to our logit/probit model.

From Random Utility to Choice Probabilities

- Each individual i faces J alternatives, each with utility

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \quad V_{ij} = X_{ij}'\beta + \alpha_j.$$

- The decision rule:

$$y_i = \arg\max_j U_{ij}.$$

- Hence, the probability of choosing option j is

$$P(y_i = j) = P(U_{ij} \geq U_{ik},\ \forall k \neq j) = P(\varepsilon_{ik} - \varepsilon_{ij} \leq V_{ij} - V_{ik},\ \forall k).$$

- The choice probabilities depend entirely on the joint distribution of the error vector $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})$.

From Random Utility to the Multinomial Logit

- Each individual i faces J alternatives, each with utility

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \quad V_{ij} = X_{ij}'\beta + \alpha_j.$$

- The probability of choosing j is the chance that its utility exceeds all others:

$$P(y_i = j) = P(U_{ij} \geq U_{ik}, \ \forall k \neq j).$$

- To compute this probability, we must assume a distribution for the unobserved parts ε_{ij}.
- If the ε_{ij} are **i.i.d. Type I Extreme Value (Gumbel)**, the resulting choice probabilities have a closed form:

$$P(y_i = j) = \frac{\exp(V_{ij})}{\sum_{k=1}^{J} \exp(V_{ik})}.$$

Key idea

The familiar multinomial logit model is the Random Utility Model with Gumbel-distributed unobserved utility components.

Why the Gumbel Assumption?

- We assume the unobserved utilities ε_{ij} are **i.i.d. Type I Extreme Value (Gumbel)**:

$$F(\varepsilon) = \exp[-\exp(-\varepsilon)].$$

- **Independence across** j means unobserved shocks to one alternative do not affect others.
- The Gumbel assumption yields a closed form for the choice probability:

$$P(y_i = j) = \frac{\exp(V_{ij})}{\sum_{k=1}^{J} \exp(V_{ik})}.$$

- **Key property:** differences of i.i.d. Gumbel errors are **logistically distributed**:

$$\varepsilon_{ik} - \varepsilon_{ij} \sim \text{Logistic}(0, 1),$$

  which gives the **logit form**.

Sidenote: The Softmax Function

$$P(y_i = j) = \frac{\exp(V_{ij})}{\sum_{k=1}^{J} \exp(V_{ik})}$$

- **Computer scientists and statisticians** call this function a softmax.
- It converts any set of real-valued scores $\{V_{i1}, \ldots, V_{iJ}\}$ into probabilities that sum to 1.
- It's used **everywhere in classifiers**, including as the **final layer of large language models (LLMs)**.

Different language - Same math

In econometrics, it's the **multinomial logit**; in machine learning, it's the **softmax function**.

Likelihood Function for MNL

- Write data in long format: one row per (i, j).
- Indicator $y_{ij} = 1$ if i chose j, else 0.
- Log-likelihood:

$$\ell(\theta) = \sum_{i=1}^{n} \left[\sum_{j=1}^{J} y_{ij}(\alpha_j + X_{ij}'\beta) - \log\left(\sum_{k=1}^{J} e^{\alpha_k + X_{ik}'\beta} \right) \right].$$

- Estimated by maximum likelihood (same principle as binary logit).

Interpreting Coefficients in the Multinomial Logit

- The MNL expresses the probability of choosing j as

$$P(y_i = j) = \frac{\exp(V_{ij})}{\sum_{k=1}^{J} \exp(V_{ik})}, \quad V_{ij} = X_{ij}'\beta + \alpha_j.$$

- The **relative odds** of choosing j over k are:

$$\frac{P(y_i = j)}{P(y_i = k)} = \exp\big((X_{ij} - X_{ik})'\beta + (\alpha_j - \alpha_k)\big).$$

- **Interpretation:**
  - A one-unit increase in attribute x raises the odds of choosing j over k by a factor of e^{β_x}.
  - If $\beta_x > 0$, alternative j becomes relatively more attractive.
  - For continuous variables (e.g. price, travel time), the marginal effect depends on current probabilities.
- **Identification:**
  - One alternative is chosen as the **baseline**, setting $\alpha_j = 0$.
  - All other coefficients are interpreted **relative to that base choice**.

- A key property of the **multinomial logit (MNL)** model:

$$\frac{P(y_i = j)}{P(y_i = k)} = \frac{\exp(V_{ij})}{\exp(V_{ik})} = \exp(V_{ij} - V_{ik}),$$

  which **depends only on** (j, k), not on other alternatives.
- This property is called the **Independence of Irrelevant Alternatives (IIA)**.
- **Implication:** The relative odds between any two options remain unchanged if a new alternative is added or an existing one is removed.
- **Example: Red Bus / Blue Bus Problem**
  - Suppose options are *Car* and *Red Bus*.
  - Adding a very similar *Blue Bus* splits the bus probability share, but also (unrealistically) reduces $P(\text{Car})$.
- **Consequence:** The IIA assumption can be restrictive when alternatives are close substitutes.

Beyond IIA: Relaxing the Independence Assumption

- The **Independence of Irrelevant Alternatives (IIA)** assumption is often unrealistic when some choices are close substitutes.
- More flexible models relax IIA by allowing **correlated unobserved utility components** across alternatives.
- **Common extensions:**
  - **Nested Logit:** Groups similar alternatives into “nests” (e.g. public vs. private transport), allowing correlation within nests.
  - **Mixed Logit (Random Parameters Logit):** Allows individual-level taste heterogeneity via random coefficients β_i.
  - **Multinomial Probit:** Replaces the logistic with a multivariate normal error structure, fully relaxing IIA.

Market Shares and WTP

- **Market shares:** predicted choice probabilities can be aggregated across individuals:

$$s_j = \frac{1}{n}\sum_{i=1}^{n} P(y_i = j \mid X_i).$$

These s_j give model-based market shares for each alternative.

- **Counterfactuals:** change an attribute (e.g. price of brand m), recompute shares $\{s_j\}$.

- **Willingness to Pay (WTP):** if price enters with coefficient $\beta_{\text{price}} < 0$, the implicit WTP for attribute a is

$$\text{WTP}_a = -\frac{\beta_a}{\beta_{\text{price}}}.$$

Interpretation: amount of money consumers are willing to pay for a one-unit increase in a.

- Widely used in marketing and IO for pricing, demand estimation and policy analysis.

- **Choice experiments** (or conjoint / vignette experiments) present respondents with hypothetical choice sets among alternatives differing in attributes.
- Each alternative is described by levels of attributes (e.g. price, brand, features, warranty).
- Respondents **choose their preferred alternative in each choice set**; repeated across many sets and individuals.
- By estimating a discrete choice model (e.g. MNL) on that data, we can infer:
  - Preference weights (part-worths) for attribute levels,
  - Willingness to Pay for attribute changes,
  - Predicted market shares under new product configurations or pricing strategies.
- Widely used in marketing, public policy, political science (different names in different disciplines)

Example Screenshot from a Choice Experiment

Choice 1/7

Please read the description of the two job offers carefully and make your personal decision. We ask you, even if you are unsure, to choose one of the two cities.

Which of the two job offers would you prefer?

	Job Offer in City A	**Job Offer in City B**
Wage	A 5% higher wage	Your current/previous wage
Cultural offerings	Medium →	Low ↓
Social Diversity	Medium →	Medium →
Ecologic quality	High ↑	High ↑
Quality of the infrastructure	High ↑	High ↑
Economic dynamism	Low ↓	Medium →
Family Friendliness	Low ↓	Medium →
	○	○

Continue

From: Arntz, Brüll and Lipowski (2023) "*Do preferences for urban amenities differ by skill*", Journal of Economic Geography, Volume 23, Issue 3, May 2023, Pages 541–576

- In **behavioural microsimulation models** (e.g. *ZEW-EviSTA*), discrete choice models are used to simulate household labour supply decisions.
- Each household i faces a set of discrete labour supply alternatives j.
- For each alternative, disposable income is simulated from the tax-benefit system.
- Estimated MNL (or nested logit / mixed logit) models deliver behavioural elasticities and are used for policy counterfactuals (e.g. reform of tax rates or child benefits).

Example:
Hebsaker, M., Stichnoth, H., Bühlmann, F., Kreuz, T., Schmidhäuser, J. & Siegloch, S. (2022). *A Microsimulation Model of the German Tax and Transfer System (ZEW-EviSTA)*. ZEW Discussion Paper No. 22–026, Mannheim.

Example Data Structure for a Labour Supply Microsimulation

Household ID	Alternative j	Hours Partner A	Hours Partner B	Net income (EUR)	y_{ij}
1	1	0	0	1,800	0
1	2	0	20	2,200	0
1	3	20	0	2,400	0
1	4	20	20	2,650	1
1	5	30	30	2,800	0
1	6	40	30	3,000	0
1	7	40	40	3,100	0

- Each row represents one simulated alternative j for the household
- Both partners' hours enter the tax-benefit simulation of disposable income
- The observed (chosen) configuration has $y_{ij} = 1$.
- Predicted choice probabilities $P(y_i = j)$ allow behavioural response simulations under tax or benefit reforms.

Application 3: Trust and Stock Market Participation

- **Research question:** Why do many households avoid investing in stocks despite positive expected returns?
- **Key idea:** Low **social trust** reduces participation in financial markets.
- **Data:** Dutch Household Survey, Italian bank customers, and cross-country data on stockholding and interpersonal trust.
- **Model:** Binary choice model (**probit**)

  $$P(\text{Stock}_i = 1) = \Phi(\alpha + \beta_1 \,\text{Trust}_i + \beta_2 \,X_i),$$

  where $\Phi(\cdot)$ is the standard normal CDF.
- **Interpretation:**
  - $\beta_1 > 0$: higher trust $\Rightarrow$ greater likelihood of stock market participation.
  - Trust acts as a **non-financial barrier** to market entry, shaping household portfolio choices.

Source: Guiso, L., Sapienza, P. & Zingales, L. (2008). *Trusting the Stock Market*. Journal of Finance, 63(6), 2557–2600.

6.4: Other Limited Dependent Variable Models

Beyond Binary Choice Models

- Many dependent variables are not continuous and not just binary.
- General principle: specify a data generating process (DGP) and estimate by **Maximum Likelihood (ML)**.
- Typical features that break standard regression:
  - **Censoring:** Outcomes observed only above/below a threshold
  - **Truncation:** Sample selection into the data
  - **Counts:** Nonnegative integers: Poisson, Neg. Binomial
  - **Durations:** Time-to-event, hazard models
  - **Rare events:** Low-probability binary outcomes

The Tobit Model: Setup

- Latent linear model:

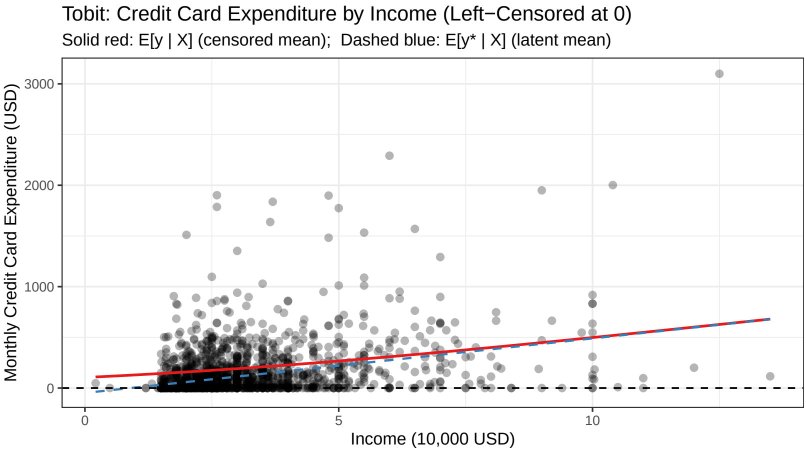
$$y_i^* = X_i\beta + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2).$$

- Observed outcome:

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0, \\ 0 & \text{if } y_i^* \leq 0. \end{cases}$$

- **Application:** household expenditure with many zeros (e.g. spending on durable goods, where many households report no purchase).
- Idea: regression structure still holds, but part of the distribution is **censored at 0**.

Tobit: Censoring at Zero (CreditCard Expenditures)



Tobit Likelihood: Two Cases

- For **censored observations** ($y_i = 0$):

$$\begin{aligned}\Pr(y_i = 0 \mid X_i) &= \Pr(y_i^* \leq 0 \mid X_i) \\ &= \Pr\left(\frac{y_i^* - X_i\beta}{\sigma} \leq \frac{0 - X_i\beta}{\sigma}\right) \\ &= \Phi\left(-\frac{X_i\beta}{\sigma}\right)\end{aligned}$$

  the area under the standard normal curve below $-X_i\beta/\sigma$.

- For **uncensored observations** ($y_i > 0$):

$$\begin{aligned}f(y_i \mid X_i, y_i > 0) &= f(y_i^* = y_i \mid X_i) \\ &= \frac{1}{\sigma}\,\phi\left(\frac{y_i - X_i\beta}{\sigma}\right)\end{aligned}$$

  the normal density of the latent outcome at the observed value.

Tobit Likelihood: Two Cases

- Each observation contributes one of these terms to the full likelihood:

$$L(\beta, \sigma) = \prod_{y_i=0} \Phi\left(-\frac{X_i\beta}{\sigma}\right) \prod_{y_i>0} \frac{1}{\sigma} \phi\left(\frac{y_i - X_i\beta}{\sigma}\right)$$

- Taking logs yields the Tobit log-likelihood:

$$\ell(\beta, \sigma) = \sum_{y_i=0} \log \Phi\left(-\frac{X_i\beta}{\sigma}\right) + \sum_{y_i>0} \log\left[\frac{1}{\sigma} \phi\left(\frac{y_i - X_i\beta}{\sigma}\right)\right]$$

Other Variants at a Glance

- **Truncated regression:** Data below cutoff not observed at all.
- **Heckman selection:** Outcome observed only if a selection equation holds.
- **Count models:** Outcomes are integers $(0, 1, 2, \ldots)$; Poisson, Neg. Binomial, zero-inflated.
- **Duration models:** Outcomes are times to event; hazard rate models.
- **Rare events logit/probit:** Adjustments for highly imbalanced data.

Key Message

All are **MLE-based variants**, linking the observed *y* to an underlying latent process.