Empirical Economics Summary

1 Causal Effects, Experiments, and Regression Analysis

Potential-Outcome Model

The treatment variable D is a binary indicator:

$$D_i = \begin{cases} 1, & \text{if treated} \\ 0, & \text{if not treated} \end{cases} \tag{1}$$

The outcome variable Y_i (e.g. self-assessed health) depends on treatment status:

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i \tag{2}$$

where:

- Y_{1i} is the potential outcome if treated,
- Y_{0i} is the potential outcome if not treated.

The causal effect, which I denote as τ for unit i, is defined as:

$$\tau_i = Y_{1i} - Y_{0i} \tag{3}$$

Since one of the potential outcomes is always unobserved, empirical research focuses on expected values.

Average Treatment Effect (ATE) and Treatment Effect on the Treated (ATT)

The average treatment effect (ATE) is given by:

$$\mathbb{E}[Y_{1i} - Y_{0i}] \tag{4}$$

The average treatment effect on the treated (ATT) is:

$$\mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1] \tag{5}$$

The observed difference in means is:

$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] = ATT + \text{Selection Bias} \tag{6}$$

where selection bias is:

$$\mathbb{E}[Y_{0i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 0] \tag{7}$$

Randomized Experiments

If treatment is randomly assigned, selection bias disappears:

$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] = \mathbb{E}[Y_{1i} - Y_{0i}] \tag{8}$$

Under random assignment:

$$\mathbb{E}[Y_{0i}|D_i = 1] = \mathbb{E}[Y_{0i}|D_i = 0] \tag{9}$$

This allows for unbiased estimation of the treatment effect.

Regression Analysis and Treatment Effects

A simple regression model can estimate treatment effects:

$$Y_i = \alpha + \beta D_i + \varepsilon_i \tag{10}$$

where:

- $\alpha = \mathbb{E}[Y_{0i}]$, the expected outcome for non-treated individuals.
- $\beta = \mathbb{E}[Y_{1i} - Y_{0i}]$, the treatment effect.
- $\varepsilon_i = Y_{0i} - \mathbb{E}[Y_{0i}]$, the error term.

If treatment assignment is correlated with ε_i, selection bias occurs. To address this, control variables X_i are added:

$$Y_i = \alpha + \beta D_i + \gamma X_i + \varepsilon_i \tag{11}$$

which helps reduce omitted variable bias.

Heterogeneous Treatment Effects

If treatment effects vary across individuals:

$$Y_{1i} - Y_{0i} = \beta + \nu_i, \quad \mathbb{E}[\nu_i] = 0 \tag{12}$$

Then the regression model becomes:

$$Y_i = \alpha + \beta D_i + \nu_i D_i + \varepsilon_i \tag{13}$$

which introduces heteroskedasticity.

2 Recap: Ordinary Least Squares (OLS)

The Least Squares Estimator

The linear regression model is given by:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \tag{14}$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are coefficients and u is the error term.

The OLS estimator minimizes the sum of squared residuals:

$$\hat{\beta}_{\text{OLS}} = \arg\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 \tag{15}$$

For bivariate regression ($k = 1$), the solution is:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\widehat{\text{Cov}}(x, y)}{\widehat{\text{Var}}(x)} \tag{16}$$

Assumptions of OLS

1. **Linearity**: The model is linear in parameters (Allowed: βx^2, not allowed: x^β).
2. **Random Sampling**: Observations are independently drawn from the population.
3. **No Perfect Collinearity**: No exact linear relationships among explanatory variables.
4. **Zero Conditional Mean**:
$$E(u|x_1, x_2, ..., x_k) = 0 \tag{17}$$
5. **Homoskedasticity**:
$$\text{Var}(u|x_1, ..., x_k) = \sigma^2 \tag{18}$$
6. **Normality (for inference)**: $u \sim \mathcal{N}(0, \sigma^2)$

Properties of OLS

Under assumptions 1-4, OLS is **unbiased**:

$$\mathbb{E}(\hat{\beta}_j|X) = \beta_j, \quad \forall j \tag{19}$$

Under assumptions 1-5, the variance of $\hat{\beta}_j$ is:

$$\text{Var}(\hat{\beta}_j|X) = \frac{\sigma^2}{SST_j(1 - R_j^2)} \tag{20}$$

where $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ and R_j^2 is the R^2 from regressing x_j on other regressors.

Under assumptions 1-6:
$$\hat{\beta}_j \mid \mathbf{X} \sim \mathcal{N}(\beta_j, \sigma^2_{\beta_j}) \tag{21}$$

By the **Gauss-Markov Theorem**, under assumptions 1-5, OLS is the **Best Linear Unbiased Estimator (BLUE)**.

Large Sample Properties

Under assumptions 1-4, OLS is **consistent**:

$$\text{plim}\, \hat{\beta}_j = \beta_j \tag{22}$$

Under assumptions 1-5, OLS is **asymptotically normal**:

$$\hat{\beta}_j \overset{a}{\sim} \mathcal{N}\left(\beta_j, \frac{\sigma^2}{na_j^2}\right) \tag{23}$$

where:

- $\frac{\sigma^2}{na_j^2}$ is the **asymptotic variance** of $\hat{\beta}_j$, with $j = 1, \ldots, k$.
- $a_j^2 = \text{plim}\left(n^{-1} \sum_{i=1}^{n} \hat{r}_{ij}^2\right)$, where $\hat{r}_{ij}$ are the residuals from regressing x_j on all other explanatory variables.

Threats to Validity of OLS

Violations of assumptions can lead to bias or inefficiency:

- **Endogeneity (Violation of Assumption 4)**: Omitted variable bias, measurement error, simultaneity.
- **Heteroskedasticity (Violation of Assumption 5)**: Leads to inefficient estimates.
- **Multicollinearity (Violation of Assumption 3)**: Large standard errors in estimates.

3 Heteroscedasticity

Consequences for OLS

Homoscedasticity assumption:

$$\mathrm{Var}(u|x_1, \ldots, x_k) = \sigma^2 \tag{24}$$

Heteroscedasticity implies:

$$\mathrm{Var}(u_i|x_i) = \sigma_i^2 \tag{25}$$

Consequences:

- OLS remains unbiased and consistent.
- Variance estimation is biased: $\mathrm{Var}(\hat{\beta}_j)$ incorrectly estimated.
- Standard errors and test statistics (e.g., t, F) are invalid.
- OLS is no longer BLUE.

Heteroscedasticity-Robust Inference

Heteroscedasticity-robust variance estimator:

$$\widehat{\mathrm{Var}}(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} r_{ij}^2 \hat{u}_i^2}{(\mathrm{SSR}_j)^2} \tag{26}$$

Where:

- r_{ij} is the residual from regressing x_j on all other x variables.
- $\hat{u}_i$ is the residual from the original regression model $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$.
- SSR_j is the sum of squared residuals from the regression of x_j on all other x variables.

Testing for Heteroscedasticity

Breusch-Pagan Test:

$$H_0 : \delta_i = 0, \forall i \quad \text{vs.} \quad H_1 : \exists i \text{ such that } \delta_i \neq 0 \tag{27}$$

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + v_i \tag{28}$$

Test statistic:

$$LM = nR^2_{\hat{u}^2} \sim \chi^2_k \tag{29}$$

White Test extends this by including squared and interaction terms of regressors.

Weighted Least Squares (WLS)

If heteroscedasticity is known to follow:

$$\text{Var}(u_i|x_i) = \sigma^2 h(x_i) \tag{30}$$

We transform:

$$y_i^* = \frac{y_i}{\sqrt{h_i}}, \quad x_i^* = \frac{x_i}{\sqrt{h_i}}, \quad u_i^* = \frac{u_i}{\sqrt{h_i}} \tag{31}$$

Thus, the transformed model satisfies homoscedasticity, and OLS applied to it is efficient.

Feasible Generalized Least Squares (FGLS)

If $h(x_i)$ is unknown, estimate it using:

$$\log \hat{u}_i^2 = \delta_0 + \delta_1 x_{i1} + \cdots + \delta_k x_{ik} + v_i \tag{32}$$

Then estimate:

$$\hat{h}_i = \exp(\hat{\delta}_0 + \sum_{j=1}^{k} \hat{\delta}_j x_{ij}) \tag{33}$$

WLS is applied using estimated weights.

Heteroscedasticity in the Linear Probability Model (LPM)

For binary outcome y_i:

$$\text{Var}(y_i|x_i) = p(x_i)(1 - p(x_i)) \tag{34}$$

LPM always exhibits heteroscedasticity. Robust standard errors are required.

4 Specification and Data Issues

Functional Form Misspecification

A regression model is misspecified if the wrong functional form is chosen. For example:

- Using y instead of $\log(y)$ on the left-hand side.
- Omitting interaction terms like $x_k \cdot x_{k-1}$.

Misspecification leads to biased estimation of coefficients, primarily affecting the interpretation of model parameters.

The RESET Test

The regression specification error test (RESET) is used to detect misspecifications:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + \text{error} \tag{35}$$

where $\hat{y}^2$ and $\hat{y}^3$ capture nonlinearities. Testing $H_0 : \delta_1 = \delta_2 = 0$ determines whether the linear model is adequate using an F-test.

Proxy Variables for Unobserved Explanatory Variables

When an important variable is unobserved (e.g., ability in wage equations), a proxy variable x_3 is used:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3^* + u, \tag{36}$$

$$x_3^* = \delta_0 + \delta_3 x_3 + \nu_3. \tag{37}$$

A good proxy should meet the following criteria:

- $\rho(u, x_i) = 0, \quad \forall i \quad \& \;\; \rho(u, x_3^*) = 0$
- $\rho(\nu_3, x_i) = 0, \quad \forall i.$

Measurement Error in Variables

If a dependent variable y^* is measured with error:

$$y = y^* + e_0, \tag{38}$$

OLS remains unbiased if e_0 is uncorrelated with regressors. However, if an explanatory variable x_1^* is mismeasured:

$$x_1 = x_1^* + e_1, \tag{39}$$

OLS suffers from attenuation bias:

$$\text{plim}(\hat{\beta}_1) = \beta_1 \frac{\sigma^2_{x_1^*}}{\sigma^2_{x_1^*} + \sigma^2_{e_1}}. \tag{40}$$

Missing Data and Sample Selection

OLS is consistent if data is missing at random. However, if selection is endogenous (e.g., wage data observed only for employed individuals), OLS is inconsistent due to correlation between the error term and regressors.

Least Absolute Deviations (LAD) Estimation

LAD minimizes absolute deviations instead of squared deviations:

$$\hat{\beta}_{\text{LAD}} = \arg\min_{\beta} \sum_{i=1}^{n} |y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij}|. \tag{41}$$

LAD estimates the median effect and is robust to outliers.

5 Simple Panel Data Methods

Pooled Cross Sections

A pooled cross-section model takes multiple independent cross-sections over time:

$$y_{it} = \beta_0 + \beta_1 x_{it} + u_{it}, \quad t = 1, 2, \ldots, T \tag{42}$$

where y_{it} is the dependent variable, x_{it} is a vector of explanatory variables, and u_{it} is the error term.

Difference-in-Differences (DiD) Estimation

DiD is used to evaluate treatment effects by comparing changes in outcomes before and after treatment:

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 D_i + \delta_1 (d2_t \cdot D_i) + u_{it} \tag{43}$$

where:

- $d2_t$ is a post-treatment indicator ($d2_t = 1$ after treatment, 0 otherwise),
- D_i is a treatment group indicator,
- δ_1 measures the treatment effect.

	Before	After	After–Before
Control	β_0	$\beta_0 + \delta_0$	δ_0
Treatment	$\beta_0 + \beta_1$	$\beta_0 + \delta_0 + \beta_1 + \delta_1$	$\delta_0 + \delta_1$
Treatment–Control	β_1	$\beta_1 + \delta_1$	δ_1

Two-Period Panel Data Model

For panel data with two periods:

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it} \tag{44}$$

where a_i represents time-invariant individual heterogeneity.

First-Differences Estimator

First-differencing eliminates a_i:

$$\Delta y_{it} = \beta_1 \Delta x_{it} + \Delta u_{it} \tag{45}$$

This method controls for unobserved heterogeneity by removing individual fixed effects.

Fixed Effects Model

The fixed effects (FE) estimator accounts for individual heterogeneity:

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it} \tag{46}$$

Using the within transformation:

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it} + \ddot{u}_{it} \tag{47}$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_i$ and $\ddot{x}_{it} = x_{it} - \bar{x}_i$.

Random Effects Model

The random effects (RE) estimator assumes a_i is uncorrelated with x_{it}:

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}, \quad E[a_i|x_{it}] = 0 \tag{48}$$

The variance components structure is given by:

$$\mathrm{Var}(y_{it}) = \sigma_a^2 + \sigma_u^2 \tag{49}$$

The RE model is estimated using Generalized Least Squares (GLS), which optimally combines between-group and within-group variation.

6 Advanced Panel Data Methods

Fixed Effects Estimation

The fixed effects (FE) model accounts for time-invariant unobserved heterogeneity:

$$y_{it} = \beta_1 x_{it} + a_i + u_{it}, \tag{50}$$

where a_i is an individual-specific effect. The within transformation eliminates a_i:

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it} + \ddot{u}_{it}, \tag{51}$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_i$ and similarly for x and u. The within estimator is consistent even if a_i is correlated with x_{it}, but it cannot estimate the effects of time-invariant variables.

Random Effects Estimation

The random effects (RE) model assumes that a_i is uncorrelated with x_{it}:

$$y_{it} = \beta_1 x_{it} + a_i + u_{it}, \quad \text{where } E[a_i|x_{it}] = 0. \tag{52}$$

The GLS transformation is applied:

$$y_{it}^* = (1-\theta)\bar{y}_i + \theta y_{it}, \quad x_{it}^* = (1-\theta)\bar{x}_i + \theta x_{it}, \tag{53}$$

where θ depends on the variance components σ_a^2 and σ_u^2. If $\theta = 0$, RE coincides with pooled OLS; if $\theta = 1$, RE coincides with FE.

Hausman Test for RE vs FE

The Hausman test checks if RE is appropriate:

$$H_0 : E[a_i|x_{it}] = 0, \quad H_A : E[a_i|x_{it}] \neq 0. \tag{54}$$

The test statistic is (not mentioned in lecture slides):

$$H = (\hat{\beta}_{RE} - \hat{\beta}_{FE})'[Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})]^{-1}(\hat{\beta}_{RE} - \hat{\beta}_{FE}). \tag{55}$$

If H is large, RE is inconsistent, and FE is preferred.

Correlated Random Effects (CRE)

The correlated random effects model relaxes the assumption that a_i is uncorrelated with x_{it}:

$$a_i = \alpha + \gamma\bar{x}_i + r_i, \tag{56}$$

where $\bar{x}_i$ is the group mean of x_{it}. Substituting into the model:

$$y_{it} = \alpha + \beta x_{it} + \gamma\bar{x}_i + r_i + u_{it}. \tag{57}$$

This model combines within and between effects, allowing for correlation between a_i and x_{it}.

Application to Clustered Data

Panel methods apply to clustered data structures, such as:

- States (counties within states form clusters),
- Families (siblings form clusters),
- Firms (employees within firms form clusters).

Fixed effects methods can be used to remove group fixed effects in any kind of clustered data. Cluster-robust standard errors account for within-cluster correlations, providing more robust inference than relying solely on random or fixed effects estimation.

7 Instrumental Variables Estimation

Instrumental Variables (IV) estimation is used to address endogeneity in regressors, such as omitted variable bias or measurement error. The basic model is:

$$y = \beta_0 + \beta_1 x + u, \tag{58}$$

where x is endogenous, i.e., $\text{Cov}(x, u) \neq 0$. An instrument z must satisfy:

$$\text{Cov}(z, u) = 0 \quad \text{and} \quad \text{Cov}(z, x) \neq 0. \tag{59}$$

The IV estimator for β_1 is:

$$\hat{\beta}_1^{IV} = \frac{\text{Cov}(z, y)}{\text{Cov}(z, x)}. \tag{60}$$

Two Stage Least Squares (2SLS)

In the presence of multiple instruments, 2SLS is used. The first stage estimates the endogenous regressor x as:

$$x = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \nu, \tag{61}$$

and the second stage estimates the outcome:

$$y = \beta_0 + \beta_1 \hat{x} + u, \tag{62}$$

where $\hat{x}$ is the fitted value from the first stage.

Testing for Endogeneity

To test whether a regressor x is endogenous, the following regression is used:

$$y = \beta_0 + \beta_1 x + \beta_2 z_1 + \beta_3 z_2 + \delta \hat{\nu} + \varepsilon, \tag{63}$$

where $\hat{\nu}$ is the residual from the first stage. If δ is significant, x is endogenous.

Weak Instruments

A weak instrument z has a small correlation with x. The asymptotic bias of the IV estimator is:

$$\text{plim}(\hat{\beta}_1^{IV}) = \beta_1 + \frac{\text{Cor}(z, u)}{\text{Cor}(z, x)} \cdot \frac{\sigma_u}{\sigma_x}. \tag{64}$$

If $\text{Cor}(z, x)$ is small, the bias can be large, and the standard errors may become inflated.

Local Average Treatment Effect (LATE)

IV estimation identifies the LATE, which is the average effect for those affected by the instrument (compliers):

$$\text{LATE} = E[Y_1 - Y_0|\text{compliers}], \tag{65}$$

where compliers are units for whom the instrument affects the treatment status. Monotonicity is required, meaning the instrument affects all units in the same direction.

Errors-in-Variables Problem

IV can address errors-in-variables problems. Suppose x^* is unobserved, and $x = x^* + e$ is observed. The model is:

$$y = \beta_0 + \beta_1 x^* + u, \tag{66}$$

and the IV estimator uses an instrument z that is correlated with x^* but uncorrelated with e.

Testing Overidentification Restrictions

When more instruments than needed are available, the overidentification test checks if the instruments are valid:

$$nR^2 \sim \chi^2_q, \tag{67}$$

where q is the number of overidentifying restrictions. A large test statistic suggests some instruments are invalid.

Heteroskedasticity in 2SLS

Heteroskedasticity-robust standard errors can be calculated for 2SLS. The Breusch-Pagan test for heteroskedasticity is:

$$\hat{u}^2 = \gamma_0 + \gamma_1 z_1 + \cdots + \gamma_m z_m + \varepsilon, \tag{68}$$

where $\hat{u}$ are the 2SLS residuals. A joint test of $\gamma_1 = \cdots = \gamma_m = 0$ is performed. Additionally, 2SLS allows for weighting schemes based on estimated error variance.

8 Regression Discontinuity Designs (RD)

Regression Discontinuity (RD) designs exploit precise knowledge of rules determining treatment assignment. The key idea is that some rules assign treatment quasi-randomly near a cutoff, allowing for causal inference. There are two versions of RD:

- **Sharp RD**: Treatment assignment is deterministic based on a running variable x_i.
- **Fuzzy RD**: Treatment assignment is probabilistic, and instrumental variables (IV) estimation is used.

Sharp RD

In Sharp RD, treatment status D_i is a deterministic function of the running variable x_i:

$$D_i = \begin{cases} 1 & \text{if } x_i \geq x_0, \\ 0 & \text{if } x_i < x_0. \end{cases} \tag{69}$$

The causal effect ρ is estimated by distinguishing the discontinuous treatment assignment from the smooth function of x_i:

$$Y_i = \alpha + \beta x_i + \rho D_i + \eta_i. \tag{70}$$

Non-Linear Specification

When $\mathbb{E}(Y_{0i}|x_i)$ is non-linear, a flexible functional form (e.g., polynomial) is used:

$$Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x_i^p + \rho D_i + \eta_i. \tag{71}$$

Centering x_i at the cutoff x_0 is advisable, as it simplifies interpretation of the treatment effect ρ.

Non-Parametric Estimation

Non-parametric RD focuses on a small neighborhood around the cutoff x_0:

$$\lim_{\Delta \to 0} [\mathbb{E}(Y_i|x_0 \leq x_i < x_0 + \Delta) - \mathbb{E}(Y_i|x_0 - \Delta < x_i < x_0)] = \mathbb{E}(Y_{1i} - Y_{0i}|x_i = x_0). \tag{72}$$

This approach reduces the likelihood of mistaking non-linearity for discontinuity. Bandwidth selection is crucial, and kernel weighting is often used to give more weight to observations closer to the cutoff.

Fuzzy RD

In Fuzzy RD, treatment status D_i is not deterministic but changes discontinuously at the cutoff. The intended treatment status $T_i = 1(x_i \geq x_0)$ is used as an instrument for D_i:

$$\mathbb{E}(D_i|x_i) = g_0(x_i) + [g_1(x_i) - g_0(x_i)]T_i. \tag{73}$$

The causal effect is estimated using 2SLS, where T_i is the instrument for D_i.

Parametric Fuzzy RD

In the parametric version, the first stage estimates D_i as a function of x_i and T_i:

$$D_i = \gamma_{00} + \gamma_{01} x_i + \gamma_{02} x_i^2 + \cdots + \gamma_{0p} x_i^p + \gamma_0^* T_i + \gamma_1^* x_i T_i + \cdots + \gamma_p^* x_i^p T_i + \xi_i. \tag{74}$$

The second stage estimates the outcome:

$$Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x_i^p + \rho D_i + \eta_i. \tag{75}$$

Non-Parametric Fuzzy RD

The non-parametric version uses the Wald estimator to estimate the treatment effect:

$$\hat{\rho} = \left. \frac{\bar{Y}|_{T_i=1} - \bar{Y}|_{T_i=0}}{\bar{D}|_{T_i=1} - \bar{D}|_{T_i=0}} \right|_{x_0 - \Delta < x_i < x_0 + \Delta}. \tag{76}$$

This rescales the outcome discontinuity by the treatment discontinuity near the cutoff.

Checking for Validity

To ensure the validity of RD designs, several checks can be performed:

- **Placebo Tests**: Analyze data known to not exhibit a discontinuity at the cutoff.
- **Bandwidth Sensitivity**: Check if the treatment effect remains stable with different bandwidths.
- **Manipulation Tests**: Ensure no excess mass of observations just above or below the cutoff, which could indicate manipulation.

Example: Government Transfers & Political Support

An RD analysis of a poverty alleviation program in Uruguay found that receiving transfers increased support for the government. The treatment effect was estimated using both parametric and non-parametric RD approaches, yielding similar results. However, the effect was sensitive to bandwidth selection, highlighting the importance of local estimation.

Example: Home Ownership & Veteran Status

A fuzzy RD analysis of the GI Bill's effect on home ownership found that eligibility for veteran home loans significantly increased home ownership rates. The treatment effect was estimated using both local and parametric fuzzy RD approaches. The local Wald estimator yielded a strong and statistically significant effect, while the parametric approach provided a noisier estimate.

9 Limited Dependent Variable Models

Limited dependent variables are those whose range of values is substantively restricted, such as binary outcomes, counts, or censored data. These models are used when the dependent variable does not satisfy the assumptions of linear regression. Common models include:

- **Logit and Probit Models** for binary outcomes.
- **Poisson Model** for count data.
- **Tobit Model** for censored data.
- **Truncated Regression Model** for truncated samples.
- **Sample Selection Corrections** for non-random samples.

Logit and Probit Models

For binary outcomes $y \in \{0, 1\}$, the probability of success is modeled as:

$$P(y = 1|x) = G(x\beta), \tag{77}$$

where $G(\cdot)$ is the link function. For logit, $G(x\beta) = \frac{\exp(x\beta)}{1+\exp(x\beta)}$, and for probit, $G(x\beta) = \Phi(x\beta)$, where $\Phi(\cdot)$ is the standard normal CDF.

Latent Variable Model

The logit and probit models can be motivated by a latent variable model:

$$y^* = x\beta + \varepsilon, \tag{78}$$

where $y = 1(y^* > 0)$. If $\varepsilon \sim \mathcal{N}(0, 1)$, the model is probit; if ε follows a standard logistic distribution, the model is logit.

Partial Effects

The marginal effect of x_j on $P(y=1|x)$ is:

$$\frac{\partial P(y=1|x)}{\partial x_j} = g(x\beta)\beta_j, \tag{79}$$

where $g(\cdot)$ is the density function of $G(\cdot)$. For logit, $g(x\beta) = \frac{\exp(x\beta)}{(1+\exp(x\beta))^2}$, and for probit, $g(x\beta) = \phi(x\beta)$. The partial effects are heterogeneous across units and are often reported as average partial effects (APE).

Maximum Likelihood Estimation

The log-likelihood function for logit and probit models is:

$$\mathcal{L}(\beta) = \sum_{i=1}^{n} \left[y_i \log G(x_i\beta) + (1-y_i)\log(1-G(x_i\beta)) \right]. \tag{80}$$

The MLE $\hat{\beta}$ maximizes $\mathcal{L}(\beta)$. The MLE is consistent, asymptotically normal, and asymptotically efficient under correct model specification.

Testing Multiple Hypotheses

The likelihood ratio test compares the log-likelihood of the unrestricted model L_{ur} to that of the restricted model L_r:

$$LR = 2(L_{ur} - L_r) \sim \chi^2_q, \tag{81}$$

where q is the number of restrictions. The Wald test and Lagrange multiplier test are also commonly used for hypothesis testing in maximum likelihood estimation.

Poisson Model

For count data y, the Poisson model assumes:

$$y \sim \text{Poisson}(\lambda), \quad \lambda = \exp(x\beta). \tag{82}$$

The log-likelihood function is:

$$\mathcal{L}(\beta) = \sum_{i=1}^{n} \left[y_i(x_i\beta) - \exp(x_i\beta) - \log(y_i!) \right]. \tag{83}$$

Extensions such as the negative binomial model are used to address overdispersion.

Tobit Model

The Tobit model is used for censored data:

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0, \\ 0 & \text{otherwise}, \end{cases} \tag{84}$$

where $y_i^* = x_i\beta + \varepsilon_i$ and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. The log-likelihood function accounts for both the censored and uncensored observations.

Truncated Regression Model

For truncated samples, the observed outcome y_i is:

$$y_i = \begin{cases} y_i^* & \text{if } c_{li} < y_i^* < c_{ui}, \\ \text{unobserved} & \text{otherwise.} \end{cases} \tag{85}$$

The log-likelihood function is:

$$\mathcal{L}(\beta, \sigma) = -\log \left[\Phi\left(\frac{c_{ui} - x_i\beta}{\sigma}\right) - \Phi\left(\frac{c_{li} - x_i\beta}{\sigma}\right) \right] - \frac{1}{2}\log(2\pi\sigma^2) + \frac{(y_i - x_i\beta)^2}{2\sigma^2}. \tag{86}$$

Sample Selection Corrections

For non-random samples, Heckman's two-step correction is used:

1. Estimate a probit model for the selection equation:

$$P(s = 1|z) = \Phi(z\gamma). \tag{87}$$

2. Compute the inverse Mills ratio $\hat{\lambda}_i = \frac{\phi(z_i\hat{\gamma})}{\Phi(z_i\hat{\gamma})}$.

3. Estimate the outcome equation using OLS with $\hat{\lambda}_i$ as an additional regressor:

$$y_i = x_i\beta + \rho\hat{\lambda}_i + \varepsilon_i. \tag{88}$$

Exclusion restrictions are crucial to avoid near-collinearity in the second stage.

Example: Weight Loss and Financial Incentives

A study on the effectiveness of financial incentives for weight loss used Heckman's correction to address selection bias. The key result was that financial incentives significantly increased weight loss, with similar estimates from OLS and Heckit, though standard errors were larger for Heckit. The exclusion restriction (living near the weigh-in location) was essential for identifying the selection effect.