DegreeLive

Bachelor of Commerce (B2001) | Level 3 • Part B Major | Monash University

Applied econometrics

Course Code: ETC3410 • Module 4 Notes

Module 4: Non-Linear Microeconometrics, Count Data & Quantile Regression

4.1 Binary Choice Models: From Linear Probability to Latent Index Models

Microeconometric outcomes frequently take discrete binary values $y_i \in \{0, 1\}$—for example, whether an Australian household holds private health insurance, whether a university graduate enters full-time employment within four months, or whether an ASX-listed firm defaults on debt obligations. Modeling the conditional response probability $p_i = P(y_i = 1 \mid \mathbf{x}_i) = E[y_i \mid \mathbf{x}_i]$ via OLS yields the **Linear Probability Model (LPM)**: $y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i$. Although LPM offers fast computation and directly interpretable linear slopes, it suffers three structural defects:

1. **Unbounded Predictions:** Fitted probabilities $\hat{y}_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}}$ routinely fall outside $[0, 1]$, violating the axioms of probability.
2. **Built-In Heteroskedasticity:** Since $y_i \mid \mathbf{x}_i \sim \text{Bernoulli}(p_i)$, its conditional variance is $\text{Var}(u_i \mid \mathbf{x}_i) = p_i(1 - p_i) = \mathbf{x}_i'\boldsymbol{\beta}(1 - \mathbf{x}_i'\boldsymbol{\beta})$, invalidating homoskedastic OLS standard errors.
3. **Constant Marginal Effects:** LPM forces a unit rise in x_{ki} to shift $P(y_i = 1 \mid \mathbf{x}_i)$ by the exact same amount β_k whether baseline probability is 0.02 or 0.50, whereas realistic economic probabilities compress near 0 and 1.

The Latent Variable Index Model

To bound predicted probabilities strictly within $(0, 1)$ while anchoring binary choices in utility maximisation, econometricians specify a continuous unobserved **latent variable** y_i^* (representing net utility or net expected return from choosing $y_i = 1$ over $y_i = 0$):

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \qquad y_i = \mathbf{1}(y_i^* > 0)$$

Where:

- y_i^* is the continuous latent propensity/net utility for observation i.
- $\mathbf{x}_i$ is the $K \times 1$ covariate vector (including a constant 1).
- $\boldsymbol{\beta}$ is the $K \times 1$ index parameter vector.
- ε_i is a mean-zero unobserved disturbance independent of $\mathbf{x}_i$ with symmetric cumulative distribution function (CDF) $F(\cdot)$ and probability density function (PDF) $f(\cdot)$.

Because $y_i^* > 0 \iff \varepsilon_i > -\mathbf{x}_i'\boldsymbol{\beta}$, symmetry of $f(\varepsilon_i)$ $(1 - F(-z) = F(z))$ gives the exact conditional response probability:

$$P(y_i = 1 \mid \mathbf{x}_i) = F(\mathbf{x}_i'\boldsymbol{\beta})$$

Degree Live

Where:

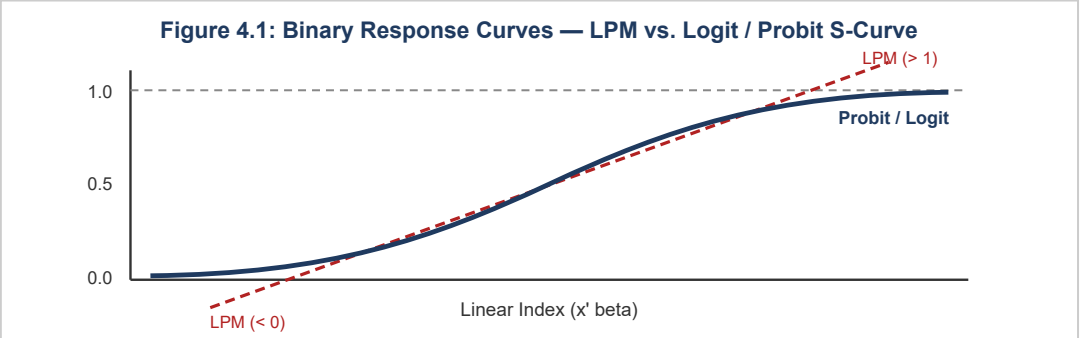
- **Probit Model:** Assumes $\varepsilon_i \sim \mathrm{N}(0,1)$, giving the standard normal CDF $\Phi(\mathbf{x}_i'\boldsymbol{\beta})$:

$$P(y_i = 1 \mid \mathbf{x}_i) = \Phi(\mathbf{x}_i'\boldsymbol{\beta}) = \int_{-\infty}^{\mathbf{x}_i'\boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

- **Logit Model:** Assumes ε_i follows the standard Logistic distribution with mean 0 and variance $\pi^2/3 \approx 3.29$, yielding the closed-form sigmoid CDF:

$$P(y_i = 1 \mid \mathbf{x}_i) = \Lambda(\mathbf{x}_i'\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i'\boldsymbol{\beta})} = \frac{1}{1 + \exp(-\mathbf{x}_i'\boldsymbol{\beta})}$$

Notice the **scale normalisation**: because multiplying $y_i^* > 0$ by any positive constant $c > 0$ leaves the binary indicator $y_i = \mathbf{1}(cy_i^* > 0)$ unchanged, $\boldsymbol{\beta}$ and σ_ε are not separately identified; only the ratio $\boldsymbol{\beta}/\sigma_\varepsilon$ is identified. Fixing $\sigma_\varepsilon = 1$ in Probit versus $\sigma_\varepsilon = \pi/\sqrt{3} \approx 1.814$ in Logit means raw Logit coefficients are roughly 1.6 to 1.8 times larger in magnitude than Probit coefficients for identical marginal effects ($\hat{\beta}_{\text{Logit}} \approx 1.6\hat{\beta}_{\text{Probit}}$).



4.2 Maximum Likelihood Estimation, Marginal Effects, and Hypothesis Testing

Because $F(\mathbf{x}_i'\boldsymbol{\beta})$ is non-linear in $\boldsymbol{\beta}$, binary choice models are estimated by **Maximum Likelihood Estimation (MLE)**. Given an independent random sample $\{(y_i, \mathbf{x}_i)\}_{i=1}^N$, each observation contributes $F(\mathbf{x}_i'\boldsymbol{\beta})^{y_i}[1 - F(\mathbf{x}_i'\boldsymbol{\beta})]^{1-y_i}$ to the joint likelihood. Taking logarithms gives the strictly concave **log-likelihood function**:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{N} \{y_i \ln F(\mathbf{x}_i'\boldsymbol{\beta}) + (1 - y_i) \ln [1 - F(\mathbf{x}_i'\boldsymbol{\beta})]\}$$

Where:

- $\ell(\boldsymbol{\beta})$ is maximised numerically via Newton-Raphson or Fisher Scoring iterations until the gradient (score vector) $\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \mathbf{0}$.
- For the Logit model specifically, the first-order condition simplifies to $\sum_{i=1}^{N}[y_i - \Lambda(\mathbf{x}_i'\hat{\boldsymbol{\beta}})]\mathbf{x}_i = \mathbf{0}$, meaning Logit residuals are orthogonal to every regressor and the sample average of predicted probabilities equals the sample proportion $\bar{y}$ identically.

Goodness of Fit and Classical Likelihood Tests

Conventional R^2 is inappropriate for binary outcomes. Instead, applied econometricians report **McFadden's Pseudo-R^2**, comparing the maximised log-likelihood $\ell(\hat{\boldsymbol{\beta}})$ against the intercept-only log-likelihood ℓ_0:

$$R^2_{\text{McFadden}} = 1 - \frac{\ell(\hat{\boldsymbol{\beta}})}{\ell_0}$$

Where:

- $\ell_0 = N[\bar{y}\ln\bar{y} + (1-\bar{y})\ln(1-\bar{y})]$ is the log-likelihood with only an intercept. Values between 0.20 and 0.40 denote strong empirical fit in cross-sectional microdata.

Testing q joint parameter restrictions $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ uses three asymptotically equivalent χ^2_q statistics: the **Likelihood Ratio (LR)** test $LR = 2[\ell(\hat{\boldsymbol{\beta}}_{\text{unres}}) - \ell(\hat{\boldsymbol{\beta}}_{\text{res}})]$, the **Lagrange Multiplier (Score)** test, and the **Wald test**:

$$W = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'[\mathbf{R}\widehat{\text{Var}}(\hat{\boldsymbol{\beta}})\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \quad \overset{d}{} \quad \chi^2_q$$

Where:

- $\mathbf{R}$ is the $q \times K$ linear restriction matrix and $\mathbf{r}$ is the $q \times 1$ vector of hypothesised values.

Marginal Effects: MEM, AME, and Odds Ratios

Raw Probit and Logit coefficients $\hat{\beta}_k$ measure effects on the unobservable latent index y_i^*, not directly on the probability $P(y_i = 1 \mid \mathbf{x}_i)$. For a continuous covariate x_{ki}, the **marginal effect (partial derivative)** is:

$$\frac{\partial P(y_i = 1 \mid \mathbf{x}_i)}{\partial x_{ki}} = f(\mathbf{x}_i'\boldsymbol{\beta}) \cdot \beta_k$$

Where:

- $f(z) = \phi(z) = \frac{1}{2\pi} e^{-z^2/2}$ for Probit, and $f(z) = \Lambda(z)[1 - \Lambda(z)]$ for Logit.
- Because $f(\mathbf{x}_i'\boldsymbol{\beta}) > 0$ everywhere, the sign of β_k always matches the direction of the probability effect, but its magnitude varies across individuals.

Econometricians summarise this distribution in two ways:

- **Marginal Effect at the Mean (MEM):** Evaluates $f(\bar{\mathbf{x}}'\hat{\boldsymbol{\beta}}) \cdot \hat{\beta}_k$ at the sample mean vector $\bar{\mathbf{x}}$ (criticised when $\bar{\mathbf{x}}$ averages over binary dummies like 0.48 female, representing no real individual).
- **Average Marginal Effect (AME):** Computes the sample average of individual marginal effects across all N observations:

$$\widehat{\text{AME}}_k = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i'\hat{\boldsymbol{\beta}}) \cdot \hat{\beta}_k$$

- **Discrete Change for Binary Dummies ($D_i \in \{0, 1\}$):** Rather than taking derivatives, compute the average counterfactual probability difference across all N observations:

$$\widehat{\text{AME}}_D = \frac{1}{N} \sum_{i=1}^{N} [F(\hat{\beta}_D + \mathbf{x}_{-D,i}'\hat{\boldsymbol{\beta}}_{-D}) - F(\mathbf{x}_{-D,i}'\hat{\boldsymbol{\beta}}_{-D})]$$

  Standard errors for MEM and AME are calculated via the **Delta Method** or parametric bootstrap.

In Logit models, exponentiating $\hat{\beta}_k$ yields the **Odds Ratio (OR)**. Define odds as $\Omega(\mathbf{x}_i) = \frac{P(y_i=1|\mathbf{x}_i)}{1-P(y_i=1|\mathbf{x}_i)} = \exp(\mathbf{x}_i'\boldsymbol{\beta})$. Increasing x_{ki} by 1 unit multiplies the odds by $\exp(\beta_k)$.

Worked Numerical Example 4.1: Probit & Logit Probabilities, AME, and Odds Ratio

Using HILDA data on working-age Australians, a Logit regression for labour force participation ($y_i = 1$) on years of post-school education (x_{1i}) and a regional postcode dummy ($x_{2i} \in \{0, 1\}$) estimates: $\hat{\beta}_0 = -1.20, \hat{\beta}_1 = 0.25$, and $\hat{\beta}_2 = -0.40$.

Step 1 — Compute Predicted Probability and Marginal Effect for an Urban Worker ($x_2 = 0$) with $x_1 = 4$ years:

$$\hat{z}_A = -1.20 + 0.25(4) - 0.40(0) = -0.20$$

$$\hat{p}_A = \Lambda(-0.20) = \frac{1}{1 + e^{0.20}} = \frac{1}{1 + 1.2214} \approx 0.4502$$

DegreeLive

$$\frac{\partial \hat{p}_A}{\partial x_1} = \hat{p}_A(1-\hat{p}_A)\hat{\beta}_1 = (0.4502)(0.5498)(0.25) \approx 0.0619$$

Where:

- At $x_1 = 4, x_2 = 0$, the predicted participation probability is 45.02%, and an additional year of post-school education raises participation probability by 6.19 percentage points.

Step 2 — Compute the Discrete Effect of Regional Residence ($x_2 : 0 \to 1$) at $x_1 = 4$ and the Odds Ratio:

$$\hat{z}_B = -1.20 + 0.25(4) - 0.40(1) = -0.60, \qquad \hat{p}_B = \frac{1}{1+e^{0.60}} = \frac{1}{1+1.8221} \approx 0.3543$$

$$\Delta \hat{p}_{x_2} = \hat{p}_B - \hat{p}_A = 0.3543 - 0.4502 = -0.0959 \ (-9.59 \text{ percentage points})$$

$$\text{OR}_{x_2} = e^{-0.40} \approx 0.6703$$

Where:

- Living in a regional postcode reduces the odds of participation by 32.97% $(1 - 0.6703)$, lowering this individual's participation probability from 45.02% to 35.43%.

4.3 Multinomial and Ordered Choice Models

Many microeconomic decisions involve more than two discrete alternatives. These fall into two distinct structures: **unordered multinomial choices** (e.g., Melbourne commuter mode choice: train, car, tram, or bicycle) and **ordered discrete choices** (e.g., HILDA self-reported health rated $1 = \text{Poor}$ to $5 = \text{Excellent}$, or S&P sovereign/corporate credit ratings).

Random Utility Model and Multinomial Logit (MNL)

Following McFadden (1974), suppose individual i chooses among J mutually exclusive unordered alternatives $j \in \{1, \ldots, J\}$, deriving indirect utility $U_{ij} = V_{ij} + \varepsilon_{ij}$. If the stochastic taste shocks ε_{ij} are independent and identically distributed (i.i.d.) Type I Extreme Value (Gumbel), the choice probability takes the **Multinomial Logit (MNL)** form (normalising $\boldsymbol{\beta}_1 = \mathbf{0}$ for base category $j = 1$):

DegreeLive

$$P(y_i = j \mid \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}{1 + \sum_{m=2}^{J} \exp(\mathbf{x}_i'\boldsymbol{\beta}_m)}, \qquad j = 2, \ldots, J$$

Where:

- $\boldsymbol{\beta}_j$ measures the effect of individual characteristics $\mathbf{x}_i$ on the log-odds of choosing alternative j relative to base category 1: $\ln \frac{P(y_i=j|\mathbf{x}_i)}{P(y_i=1|\mathbf{x}_i)} = \mathbf{x}_i'\boldsymbol{\beta}_j$.
- **Caution:** Even if $\beta_{jk} > 0$, the marginal effect $\frac{\partial P_{ij}}{\partial x_{ik}} = P_{ij}(\beta_{jk} - \sum_{m=1}^{J} P_{im}\beta_{mk})$ can be negative if another category m has an even larger positive coefficient $\beta_{mk} > \beta_{jk}$.

Independence of Irrelevant Alternatives (IIA) and Mixed Logit

Notice from the log-odds ratio above that $\frac{P(y_i=j|\mathbf{x}_i)}{P(y_i=m|\mathbf{x}_i)} = \exp[\mathbf{x}_i'(\boldsymbol{\beta}_j - \boldsymbol{\beta}_m)]$ depends *solely* on categories j and m, independent of all other alternatives in the choice set. This strong property—**Independence of Irrelevant Alternatives (IIA)**—fails when alternatives share unobserved attributes (the classic "red bus / blue bus" paradox: introducing a slightly re-coloured tram line draws riders disproportionately from existing trams rather than cars).

Analysts test IIA via the **Hausman-McFadden test** (comparing $\hat{\boldsymbol{\beta}}$ from the full J-alternative MNL against a restricted subset dropping one alternative). When IIA fails, applied econometricians deploy **Nested Logit** (grouping correlated alternatives into nests) or **Mixed (Random Parameters) Logit**, which allows coefficients $\boldsymbol{\beta}_i \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to vary randomly across consumers and eliminates IIA.

Ordered Logit and Ordered Probit Models

When discrete categories $y_i \in \{1, 2, \ldots, J\}$ possess a natural ranking, an **Ordered Choice Model** maps a single continuous latent variable $y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$ (with no intercept when all $J-1$ thresholds are free) onto observed categories via ascending threshold cut-points $\mu_1 < \mu_2 < \cdots < \mu_{J-1}$:

$$y_i = j \iff \mu_{j-1} < y_i^* \leq \mu_j, \qquad (\mu_0 = -\infty,\ \mu_J = +\infty)$$

Where:

- Category probabilities are $P(y_i = j \mid \mathbf{x}_i) = F(\mu_j - \mathbf{x}_i'\boldsymbol{\beta}) - F(\mu_{j-1} - \mathbf{x}_i'\boldsymbol{\beta})$, with $F = \Phi$ (Ordered Probit) or $F = \Lambda$ (Ordered Logit).
- **Middle-Category Sign Ambiguity:** While $\beta_k > 0$ unambiguously decreases $P(y_i = 1 \mid \mathbf{x}_i)$ and increases $P(y_i = J \mid \mathbf{x}_i)$, its marginal effect on interior categories $(1 < j < J)$, $\frac{\partial P(y_i=j)}{\partial x_k} = [f(\mu_{j-1} - \mathbf{x}_i'\boldsymbol{\beta}) - f(\mu_j - \mathbf{x}_i'\boldsymbol{\beta})]\beta_k$, can be positive, zero, or negative depending on where $\mathbf{x}_i'\boldsymbol{\beta}$ lies relative to μ_{j-1} and μ_j.

Degreelive

Worked Numerical Example 4.2: Ordered Logit Probabilities and Middle-Category Marginal Effect

An analyst estimates an Ordered Logit model of financial stress ($y_i \in \{1 = \text{Comfortable},\ 2 = \text{Managing},\ 3 = \text{Distressed}\}$) on household debt-to-income ratio x_i, obtaining $\hat{\beta} = 0.80$, $\hat{\mu}_1 = 1.00$, and $\hat{\mu}_2 = 2.60$. Evaluate a Melbourne household with $x_i = 2.0$, so $\mathbf{x}_i'\hat{\beta} = 2.0(0.80) = 1.60$.

Step 1 — Compute Predicted Category Probabilities:

$$P(y_i = 1) = \Lambda(1.00 - 1.60) = \Lambda(-0.60) = \frac{1}{1 + e^{0.60}} = 0.3543$$

$$P(y_i \leq 2) = \Lambda(2.60 - 1.60) = \Lambda(1.00) = \frac{1}{1 + e^{-1.00}} = 0.7311$$

$$P(y_i = 2) = 0.7311 - 0.3543 = 0.3768, \qquad P(y_i = 3) = 1 - 0.7311 = 0.2689$$

Where:

- The probabilities sum to $0.3543 + 0.3768 + 0.2689 = 1.0000$.

Step 2 — Compute the Marginal Effect of Debt-to-Income Ratio on the Middle Category ($y_i = 2$):

$$\lambda(-0.60) = 0.3543(1 - 0.3543) = 0.2288, \qquad \lambda(1.00) = 0.7311(1 - 0.7311) = 0.1966$$

$$\frac{\partial P(y_i = 2)}{\partial x_i} = [\lambda(-0.60) - \lambda(1.00)] \cdot \hat{\beta} = (0.2288 - 0.1966)(0.80) = +0.0258$$

Where:

- At $x_i = 2.0$, higher debt reduces $P(y = 1)$ by $-0.2288(0.80) = -0.1830$, raises $P(y = 3)$ by $+0.1966(0.80) = +0.1573$, and net raises the middle category $P(y = 2)$ by $+0.0258$ ($-0.1830 + 0.0258 + 0.1573 = 0$).

DegreeLive

4.4 Count Data Models: Poisson, Overdispersion, and Negative Binomial

Count outcomes take non-negative integer values $y_i \in \{0, 1, 2, \dots\}$—such as annual GP consultations billed to Medicare, workplace injury claims filed under WorkSafe Victoria, or firm patent registrations. Modeling counts via OLS can predict negative counts and fails to reflect discrete skewness and heteroskedasticity.

Poisson Regression and QMLE Robustness

The baseline count model specifies $y_i \mid \mathbf{x}_i \sim \text{Poisson}(\lambda_i)$ with exponential conditional mean ensuring $\lambda_i > 0$:

$$P(Y_i = y_i \mid \mathbf{x}_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \qquad \lambda_i = E[y_i \mid \mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$$

Where:

- $\beta_k = \frac{\partial \ln E[y_i|\mathbf{x}_i]}{\partial x_{ki}}$ is the semi-elasticity (a 1-unit rise in x_{ki} shifts expected counts by roughly $100 \cdot \beta_k\%$, or exact factor $\text{IRR}_k = \exp(\beta_k)$, the **Incidence Rate Ratio**).
- The Poisson distribution enforces strict **equidispersion**: $\text{Var}(y_i \mid \mathbf{x}_i) = E[y_i \mid \mathbf{x}_i] = \lambda_i$.

In empirical microdata, count outcomes almost universally exhibit **overdispersion** ($extVar(y_i \mid \mathbf{x}_i) > E[y_i \mid \mathbf{x}_i]$) due to unobserved heterogeneity. Remarkably, Gourieroux, Monfort, and Trognon (1984) proved that the **Poisson Quasi-MLE (QMLE)** estimator remains $\sqrt{N}$-consistent for $\boldsymbol{\beta}$ whenever the conditional mean $E[y_i \mid \mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$ is correctly specified—even if the data are *not* Poisson distributed (or even continuous non-negative outcomes like trade flows or health spending!). Under overdispersion, however, standard Poisson MLE variance estimates are too small, requiring **Huber-White sandwich (robust) standard errors**.

Cameron-Trivedi Overdispersion Test & Negative Binomial (NB2)

To test equidispersion ($H_0 : \text{Var}(y_i \mid \mathbf{x}_i) = \lambda_i$) against overdispersion ($\text{Var}(y_i \mid \mathbf{x}_i) = \lambda_i + \alpha\lambda_i^2$ with $\alpha > 0$), Cameron and Trivedi (1990) run an auxiliary OLS regression without an intercept using fitted Poisson values $\hat{\lambda}_i = \exp(\mathbf{x}_i'\hat{\boldsymbol{\beta}})$:

$$\frac{(y_i - \hat{\lambda}_i)^2 - y_i}{\hat{\lambda}_i} = \alpha\hat{\lambda}_i + \text{error}_i$$

Where:

- A statistically significant positive t-statistic on $\hat{\alpha}$ rejects equidispersion in favour of quadratic overdispersion.

To model overdispersion parametrically and improve efficiency, the **Negative Binomial (NB2)** model introduces a multiplicative unobserved heterogeneity term $v_i \sim \text{Gamma}(1/\alpha, \alpha)$ with $E[v_i] = 1$ and $\text{Var}(v_i) = \alpha$, yielding $y_i \mid \mathbf{x}_i, v_i \sim \text{Poisson}(\lambda_i v_i)$. Integrating out v_i leaves the same conditional mean $E[y_i \mid \mathbf{x}_i] = \lambda_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$ with quadratic variance:

$$\text{Var}(y_i \mid \mathbf{x}_i) = \lambda_i + \alpha\lambda_i^2$$

Where:

- $\alpha \geq 0$ is the overdispersion parameter (as $\alpha \to 0$, NB2 collapses cleanly to the Poisson model).

Excess Zeros: Zero-Inflated (ZIP) and Hurdle Models

In datasets where an unusually large share of observations equal zero (for example, 65% of households making zero charitable donations or zero hospital admissions), two mechanisms accommodate excess zeros:

- **Hurdle (Two-Part) Model:** Separates the participation decision ($y_i > 0$ vs. $y_i = 0$, estimated by Logit/Probit) from positive intensity ($y_i \mid y_i > 0$, estimated by a zero-truncated Poisson or zero-truncated Negative Binomial). Every zero arises from a single corner-solution decision.
- **Zero-Inflated Poisson / NB (ZIP / ZINB):** Assumes two latent subpopulations: "structural zeros" (probability $\pi_i = \Lambda(\mathbf{z}_i'\boldsymbol{\gamma})$ who never participate) and "at-risk count units" (probability $1 - \pi_i$ whose counts follow Poisson(λ_i), which may also produce sampling zeros).

Worked Numerical Example 4.3: Poisson IRR, Overdispersion Test, and NB2 Variance

A health economist models annual Medicare GP visits (y_i) among $N = 2,500$ Australian adults as a function of age and Bulk-Billing Clinic access ($D_i \in \{0, 1\}$), obtaining Poisson QMLE estimate $\hat{\beta}_D = 0.223$ (robust $\text{SE} = 0.055$) and $\hat{\lambda}_i = 4.50$ for a benchmark patient with $D_i = 0$.

Step 1 — Compute the Incidence Rate Ratio (IRR) and Marginal Effect:

$$\text{IRR}_D = e^{0.223} \approx 1.250$$

$$\hat{\lambda}(D = 1) = 4.50 \times 1.250 = 5.625, \qquad \Delta\hat{\lambda} = 5.625 - 4.50 = +1.125 \text{ GP visits/year}$$

Where:

- Residing near a bulk-billing clinic is associated with a 25.0% higher expected annual GP visit rate ($+1.125$ visits from a baseline of 4.50).

Degreelive

Step 2 — Evaluate Overdispersion Under NB2: Suppose the Cameron-Trivedi auxiliary regression estimates $\hat{\alpha} = 0.40$ ($t = 9.8, p < 0.001$). Under Poisson equidispersion, predicted variance at $\hat{\lambda}_i = 4.50$ is 4.50; under NB2, true conditional variance is:

$$\widehat{\mathrm{Var}}_{\mathrm{NB2}}(y_i \mid \mathbf{x}_i) = 4.50 + 0.40(4.50)^2 = 4.50 + 8.10 = 12.60$$

Where:

- Conditional variance (12.60) is 2.8 times larger than the Poisson mean (4.50), explaining why naive Poisson MLE standard errors would understate true uncertainty by roughly $\sqrt{2.8} \approx 1.67\times$ unless robust QMLE or NB2 estimation is applied.

4.5 Quantile Regression (Koenker & Bassett, 1978)

Standard OLS models exclusively the conditional mean $E[y_i \mid \mathbf{x}_i]$. Yet microeconomic policy questions often target inequality, tail risks, or distributional heterogeneity: does a minimum-wage rise lift the 10th percentile of Australian wages more than the median? Does tertiary education compress or widen wage dispersion across the earnings distribution? **Quantile Regression** (Koenker & Bassett, 1978) estimates the τ-th conditional quantile $Q_\tau(y_i \mid \mathbf{x}_i)$ for any percentile $\tau \in (0, 1)$:

$$Q_\tau(y_i \mid \mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}(\tau)$$

Where:

- $\tau \in (0, 1)$ specifies the quantile index (e.g., $\tau = 0.10$ for the 10th percentile, $\tau = 0.50$ for the conditional median, $\tau = 0.90$ for the 90th percentile).
- $\boldsymbol{\beta}(\tau)$ is a quantile-specific slope vector measuring how covariates shift the τ-th conditional quantile of y_i.

Asymmetric Check (Pinball) Loss Function

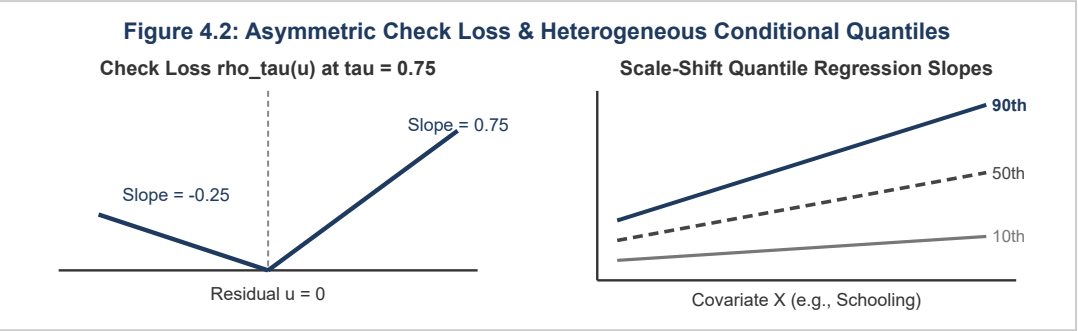
Just as OLS solves a quadratic loss minimisation to estimate the conditional mean, Quantile Regression solves a linear programming problem minimising the sum of asymmetrically weighted absolute residuals via the **check (pinball) loss function** $\rho_\tau(u)$:

$$\hat{\boldsymbol{\beta}}(\tau) = \arg\min_{\mathbf{b} \in \mathbb{R}^K} \sum_{i=1}^{N} \rho_\tau(y_i - \mathbf{x}_i'\mathbf{b}), \qquad \rho_\tau(u) = u \cdot (\tau - \mathbf{1}(u < 0)) = \begin{cases} \tau u & \text{if } u \geq 0 \\ (\tau - 1)u & \text{if } u < 0 \end{cases}$$

Where:

- $u_i = y_i - \mathbf{x}_i'\mathbf{b}$ is the residual for observation i.

- Positive residuals ($y_i \geq \mathbf{x}_i'\mathbf{b}$) receive weight τ, while negative residuals ($y_i < \mathbf{x}_i'\mathbf{b}$) receive weight $1-\tau$. At $\tau = 0.50, \rho_{0.5}(u) = 0.5|u|$, recovering **Least Absolute Deviations (LAD / Median Regression)**.
- **Properties:** Quantile regression is robust to extreme vertical outliers in y_i (because only the sign of residuals matters in the first-order condition), is equivariant to monotone transformations ($Q_\tau(g(y) \mid \mathbf{x}) = g(Q_\tau(y \mid \mathbf{x}))$ for increasing $g(\cdot)$, unlike $E[\ln y] \neq \ln E[y]$!), and separates pure *location shifts* (identical $\beta(\tau)$ across all τ) from *scale shifts* (varying $\beta(\tau)$ across τ).



Worked Numerical Example 4.4: Check Loss Minimisation & HILDA Wage Dispersion

Suppose five Australian workers share identical covariates except for log hourly wage observations $y = \{3.10,\ 3.25,\ 3.40,\ 3.60,\ 4.20\}$. Evaluate the total asymmetric check loss for the $\tau = 0.80$ conditional quantile at candidate parameter $b = 3.60$ versus $b = 3.40$:

Step 1 — Evaluate Check Loss at $b = 3.60$:

- $u_1 = 3.10 - 3.60 = -0.50 < 0 \implies \rho_{0.8}(-0.50) = (0.80-1)(-0.50) = 0.100$
- $u_2 = 3.25 - 3.60 = -0.35 < 0 \implies \rho_{0.8}(-0.35) = 0.20(0.35) = 0.070$
- $u_3 = 3.40 - 3.60 = -0.20 < 0 \implies \rho_{0.8}(-0.20) = 0.20(0.20) = 0.040$
- $u_4 = 3.60 - 3.60 = 0.00 \geq 0 \implies \rho_{0.8}(0.00) = 0.000$
- $u_5 = 4.20 - 3.60 = +0.60 \geq 0 \implies \rho_{0.8}(0.60) = 0.80(0.60) = 0.480$

Total check loss at $b = 3.60$: $L(3.60) = 0.100 + 0.070 + 0.040 + 0 + 0.480 = 0.690$.

Step 2 — Compare Against $b = 3.40$: At $b = 3.40$, residuals are $\{-0.30, -0.15, 0, +0.20, +0.80\}$, giving losses $0.20(0.30 + 0.15) + 0.80(0.20 + 0.80) = 0.090 + 0.800 = 0.890 > 0.690$. Because 80% of observations lie at or below 3.60, $b = 3.60$ uniquely minimises the $\tau = 0.80$ check loss, even if the top earner rose from 4.20 to 9.20 (illustrating outlier robustness).

Empirical Interpretation: If a HILDA quantile wage regression on years of university education produces $\hat{\beta}(0.10) = 0.055$, $\hat{\beta}(0.50) = 0.082$, and $\hat{\beta}(0.90) = 0.124$, higher education raises wages across the entire distribution while *widening* conditional within-group wage inequality ($0.124 - 0.055 = +0.069$ log-point spread between the 90th and 10th percentiles).

4.6 Endogeneity in Non-Linear Models: The Control Function Approach

In linear models, endogeneity is corrected via 2SLS by plugging Stage 1 fitted values $\hat{D}_i$ into Stage 2. In non-linear models such as Probit, Logit, or Poisson, naive "plug-in 2SLS" ($F(\mathbf{w}_i'\boldsymbol{\gamma} + \beta\hat{D}_i)$) is **inconsistent** because $E[F(\cdot)] \neq F(E[\cdot])$ ("the forbidden regression"). Instead, econometricians employ the **Control Function (Rivers-Vuong) Two-Stage Residual Inclusion (2SRI)** method.

Consider a continuous endogenous regressor D_i in a Probit model with valid excluded instrument Z_i:

$$y_i^* = \beta D_i + \mathbf{w}_i'\boldsymbol{\gamma} + u_i, \qquad D_i = \pi Z_i + \mathbf{w}_i'\boldsymbol{\delta} + v_i$$

Assuming (u_i, v_i) are jointly normal with $u_i = \rho v_i + e_i$ where $e_i \perp (D_i, \mathbf{w}_i, Z_i)$ and $e_i \sim \mathrm{N}(0, \sigma_e^2)$:

1. **Stage 1:** Estimate the reduced form of D_i on $(Z_i, \mathbf{w}_i)$ by OLS and save the first-stage residuals $\hat{v}_i = D_i - \hat{\pi} Z_i - \mathbf{w}_i'\hat{\boldsymbol{\delta}}$.
2. **Stage 2:** Estimate a standard Probit (or Logit/Poisson) of y_i on $(D_i, \mathbf{w}_i, \hat{v}_i)$—retaining actual D_i and adding $\hat{v}_i$ as a control function!

Testing $H_0 : \rho = 0$ (the coefficient on $\hat{v}_i$) provides a direct **Hausman test of exogeneity** inside non-linear models. Because $\hat{v}_i$ is an estimated regressor, Stage 2 standard errors must be corrected via pairs bootstrap across both stages. Similarly, in Poisson count models with endogenous D_i, including $\hat{v}_i$ inside $\exp(\beta D_i + \mathbf{w}_i'\boldsymbol{\gamma} + \rho\hat{v}_i)$ consistently recovers structural semi-elasticities when u_i enters multiplicatively.

4.7 Model Selection, Classification Diagnostics, and Applied Empirical Workflows

In applied econometrics across R, Stata, and Python, rigorous empirical workflows follow a structured diagnostic sequence tailored to model class:

Model Class	Primary Estimation Method	Key Fit / Diagnostic Metrics	Inference & Robustness Best Practice
Binary Choice (Probit / Logit)	Maximum Likelihood (MLE)	McFadden Pseudo-R^2, AIC/BIC, ROC-AUC, Calibration plot, Link test	Report AMEs (not raw $\hat{\beta}$) with Delta-method / clustered SEs; check LPM consistency
Multinomial / Ordered	MLE (MNL, Ordered Logit)	Hausman-McFadden IIA test; Brant parallel-odds test (Ordered Logit)	Compute full category AME matrix; use Mixed Logit if IIA fails
Count Data (Poisson / NB2 / ZIP)	Poisson QMLE; NB2 MLE; Hurdle MLE	Cameron-Trivedi overdispersion test; Vuong / AIC comparison for excess zeros	Always use Huber-White robust SEs with Poisson QMLE; verify zero-share fit

Model Class	Primary Estimation Method	Key Fit / Diagnostic Metrics	Inference & Robustness Best Practice
Quantile Regression	Linear Programming (Check Loss)	Pseudo-$R^1(\tau) = 1 - L_{\text{model}}(\tau)/L_{\text{null}}(\tau)$ across τ	Report quantile process $\hat{\beta}(\tau)$ for $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$ with pairs-bootstrapped SEs

Information Criteria and ROC-AUC Evaluation

When comparing non-nested non-linear specifications (such as NB2 versus ZIP, or Probit versus complementary log-log), analysts balance log-likelihood fit against parameter parsimony using the **Akaike Information Criterion (AIC)** and **Bayesian Information Criterion (BIC)**:

$$\text{AIC} = -2\ell(\hat{\boldsymbol{\theta}}) + 2p, \qquad \text{BIC} = -2\ell(\hat{\boldsymbol{\theta}}) + p\ln(N)$$

Where:

- p is the number of estimated parameters and N is the sample size. Lower AIC/BIC values indicate superior penalised fit (BIC penalises extra parameters more heavily when $N \geq 8$).

For binary classification performance in credit scoring or default prediction (e.g., APRA prudential stress testing), analysts plot the **Receiver Operating Characteristic (ROC) curve** tracing True Positive Rate (Sensitivity) against False Positive Rate $(1 - \text{Specificity})$ across all probability cutoffs $c \in (0, 1)$. The **Area Under the Curve (AUC)** equals 0.50 for random guessing and 1.00 for perfect separation, evaluated via K-fold out-of-sample cross-validation to prevent in-sample overfitting.

4.8 Corner Solutions & Sample Selection: Tobit and Heckman Models

Non-negative continuous economic variables frequently exhibit a point mass at zero alongside a continuous right tail—for instance, annual voluntary household superannuation contributions, household expenditure on alcohol, or weekly paid hours worked. When zeros represent optimal corner solutions rather than missing observations, OLS suffers downward attenuation bias toward zero.

The Type I Tobit Model

Tobin (1958) specified a latent linear index $y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + u_i$ with $u_i \mid \mathbf{x}_i \sim \text{N}(0, \sigma^2)$, where the observed outcome is censored from below at zero:

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

Where:

- The probability of a positive observation is $P(y_i > 0 \mid \mathbf{x}_i) = \Phi(\mathbf{x}_i'\boldsymbol{\beta}/\sigma)$.

Degree Live

- The conditional expectation among positive observations ($y_i > 0$) incorporates the **Inverse Mills Ratio (IMR)** $\lambda(z) = \phi(z)/\Phi(z)$:

$$E[y_i \mid y_i > 0, \mathbf{x}_i] = \mathbf{x}_i'\boldsymbol{\beta} + \sigma\lambda\left(\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)$$

- And the unconditional expectation across all observations (including zeros) is:

$$E[y_i \mid \mathbf{x}_i] = \Phi\left(\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)\mathbf{x}_i'\boldsymbol{\beta} + \sigma\phi\left(\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)$$

Applying the McDonald-Moffitt decomposition, the marginal effect on the unconditional mean equals the participation probability times the latent slope: $\frac{\partial E[y_i|\mathbf{x}_i]}{\partial x_{ki}} = \Phi(\mathbf{x}_i'\boldsymbol{\beta}/\sigma)\beta_k$. Unlike OLS, Tobit MLE relies heavily on exact homoskedasticity and normality of u_i.

Heckman Two-Step Sample Selection Correction

When outcomes are missing due to non-random selection—such as observing market hourly wages w_i exclusively for workers who choose to enter the labour force ($s_i = 1$)—Heckman (1979) separated the **selection equation** from the **outcome equation**:

$$s_i = \mathbf{1}(\mathbf{z}_i'\boldsymbol{\gamma} + v_i > 0), \qquad y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i \ \text{ observed iff } s_i = 1$$

Assuming bivariate normality $(u_i, v_i) \sim \mathrm{N}(\mathbf{0}, \Sigma)$ with $\mathrm{Corr}(u_i, v_i) = \rho$, the conditional expectation in the selected sample becomes:

$$E[y_i \mid \mathbf{x}_i, \mathbf{z}_i, s_i = 1] = \mathbf{x}_i'\boldsymbol{\beta} + \rho\sigma_u\lambda(\mathbf{z}_i'\boldsymbol{\gamma})$$

Where:

- $\lambda(\mathbf{z}_i'\boldsymbol{\gamma}) = \phi(\mathbf{z}_i'\boldsymbol{\gamma})/\Phi(\mathbf{z}_i'\boldsymbol{\gamma})$ is the Inverse Mills Ratio estimated from a Stage 1 Probit of s_i on $\mathbf{z}_i$.
- Stage 2 regresses y_i on $\mathbf{x}_i$ and $\hat{\lambda}_i$ among $s_i = 1$ observations. Testing $H_0 : \rho\sigma_u = 0$ tests for sample selection bias.
- **Exclusion Restriction Requirement:** Although non-linearity of $\lambda(\cdot)$ technically identifies $\boldsymbol{\beta}$ even when $\mathbf{z}_i = \mathbf{x}_i$, credible empirical identification requires at least one valid exclusion restriction in $\mathbf{z}_i$ (e.g., Childcare Subsidy entitlement or non-labour household income) that shifts participation s_i without directly affecting market wage productivity u_i.

DegreeLive

Worked Numerical Example 4.5: Tobit Unconditional vs. Conditional Marginal Effects

A Monash economist estimates a Type I Tobit model for annual voluntary superannuation contributions y_i (in A$1,000s) among Victorian workers aged 30–55, obtaining $\hat{\beta}_{\text{inc}} = 0.080$ (per A$1,000 of disposable income) and $\hat{\sigma} = 5.0$. For a worker whose linear index is $\mathbf{x}_i'\hat{\boldsymbol{\beta}} = 2.0$, compute $\mathbf{x}_i'\hat{\boldsymbol{\beta}}/\hat{\sigma} = 2.0/5.0 = 0.40$, where $\Phi(0.40) = 0.6554$ and $\phi(0.40) = 0.3683$.

Step 1 — Compute Participation Probability, IMR, and Expected Contributions:

$$P(y_i > 0 \mid \mathbf{x}_i) = \Phi(0.40) = 0.6554, \qquad \lambda(0.40) = \frac{0.3683}{0.6554} \approx 0.5619$$

$$E[y_i \mid y_i > 0, \mathbf{x}_i] = 2.0 + 5.0(0.5619) = 4.81 \ (\$4,810)$$

$$E[y_i \mid \mathbf{x}_i] = 0.6554(2.0) + 5.0(0.3683) = 1.3108 + 1.8415 = 3.15 \ (\$3,152)$$

Step 2 — Compute the Unconditional Marginal Effect of A$1,000 Higher Income:

$$\frac{\partial E[y_i \mid \mathbf{x}_i]}{\partial \text{inc}_i} = \Phi(0.40) \cdot \hat{\beta}_{\text{inc}} = 0.6554 \times 0.080 = 0.0524 \ (\$52.40 \text{ per } \$1,000 \text{ income})$$

Where:

- Of the latent A$80 propensity increase per A$1,000 income, 65.54% translates into higher expected unconditional contributions (A$52.40) after accounting for the 34.46% corner mass at zero.

4.9 Unconditional Quantile Regression (RIF-OLS) and Empirical Workflows

A subtle interpretation trap arises in standard Conditional Quantile Regression (Section 4.5): because $Q_\tau(y_i \mid X_i)$ conditions on covariates, a unit at the 90th percentile of wages *conditional on holding a postgraduate degree* is not necessarily at the 90th percentile of the *overall national wage distribution*. When policymakers ask how raising the Fair Work Commission national minimum wage or expanding union coverage shifts the 10th, 50th, or 90th percentile of the **unconditional marginal wage distribution**, Firpo, Fortin, and Lemieux (2009) introduced **Unconditional Quantile Regression via Recentred Influence Functions (RIF-OLS)**.

For target quantile q_τ of the unconditional distribution $F_Y(y)$, the Recentred Influence Function is defined as:

$$\text{RIF}(y_i; q_\tau, F_Y) = q_\tau + \frac{\tau - \mathbf{1}(y_i \leq q_\tau)}{f_Y(q_\tau)}$$

Where:

- q_τ is the sample τ-th quantile of y across all observations.
- $f_Y(q_\tau)$ is the kernel density estimate of y evaluated at q_τ.
- By construction, $E[\text{RIF}(y_i; q_\tau, F_Y)] = q_\tau$. Regressing $\text{RIF}(y_i; q_\tau, F_Y)$ on covariates $\mathbf{x}_i$ via OLS estimates the **Unconditional Quantile Partial Effect (UQPE)**—the marginal effect of a small shift in the population distribution of x_k on the overall τ-th quantile q_τ.

Reproducible Microeconometric Workflow Checklist

Executing rigorous empirical research in ETC3410 and applied policy settings requires integrating identification, specification, and computational reproducibility across four stages:

1. **Identification Design First:** Specify the causal estimand (ATE, ATT, LATE, or CATE) and justify identification (RCT randomisation, Parallel Trends, Continuity at c, IV Exclusion, or Conditional Independence) before selecting an estimator.
2. **Functional Form Matching Outcome Support:** Match the model to outcome support—Linear/FE/DiD for continuous outcomes, Probit/Logit (reporting AMEs) for binary outcomes, Poisson QMLE/NB2 (reporting IRRs and robust SEs) for counts, and Check-Loss/RIF-OLS for distributional quantiles.
3. **Variance & Dependence Alignment:** Match standard error structure to sampling and assignment correlation—HC2/cluster-robust at the unit of randomisation/policy, Wild Cluster Bootstrap for few clusters, pairs bootstrap for two-step Control Function and Heckman models, and Anderson-Rubin sets for weak instruments.
4. **Falsification & Sensitivity Reporting:** Report covariate balance, pre-trends, McCrary density checks, overdispersion tests, bandwidth stability, and alternative estimators side-by-side in transparent regression tables.

DegreeLive