Marginal Effects in Logit & Probit Models and Maximum Likelihood Tests

CLABE 2025/2026

Marco Rosso

27 November 2025

Learning Goals

- **Understand** why coefficients $\neq$ marginal effects in non-linear models
- **Compute** marginal effects for continuous variables:

$$\frac{\partial P}{\partial x_k} = f(X\beta) \cdot \beta_k$$

  - This formula holds only when the index is linear in parameters and covariates, and only for continuous regressors.
- **Compute** marginal effects for discrete variables:

$$\text{ME} = P(Y = 1|D = 1) - P(Y = 1|D = 0)$$

  - For discrete (dummy) variables, marginal effects must be computed as finite differences.
- **Distinguish** between AME (Average Marginal Effect) and MEM (at Mean)
- **Understand** the three ML tests: LR, Wald, and Lagrange Multiplier
- **Compute** the Wald test and LR statistic, its p-value, and the critical value **manually** from Stata's log-likelihood output

The Core Problem: Non-Linearity

In OLS:

$$y = \beta_0 + \beta_1 x + u$$

The marginal effect is **constant**: $\frac{dy}{dx} = \beta_1$ for all x

In Logit/Probit:

$$P(Y = 1|X) = F(X\beta)$$

where $F(\cdot)$ is CDF (logistic or normal) The marginal effect **varies**:

$$\frac{\partial P}{\partial x_k} = f(X\beta) \cdot \beta_k$$

$f(X\beta)$ depends on X for each observation!

Marginal Effects Formula

$$\boxed{\frac{\partial P}{\partial x_k} = f(X\beta) \times \beta_k} \tag{1}$$

Note: This formula applies only to models with a linear index and only when x_k is a continuous regressor. For dummy variables, marginal effects must be computed using differences in predicted probabilities.

Two components:

β_k**:** Coefficient

- Can be estimated from data
- **Constant** for all observations

$f(X\beta)$**:** Density

- Derivative of CDF
- **Varies** across individuals

Key insight: Same regressor $\Rightarrow$ **different effects** for different people

Continuous Variables: Worked Example

Setup: Logit model of low birth weight $\rightarrow \beta_{\text{age}} = 0.05$, $\beta_0 = -2.0$

Person A: age=25

$$X_A\beta = -2.0 + 0.05(25) = -0.75$$
$$f(-0.75) \approx 0.218$$
$$\frac{\partial P}{\partial \text{age}} = 0.218 \times 0.05 = \boxed{0.0109 \text{ pp}}$$

Person B: age=35

$$X_B\beta = -2.0 + 0.05(35) = -0.25$$
$$f(-0.25) \approx 0.375 \quad \text{(steeper part of curve!)}$$
$$\frac{\partial P}{\partial \text{age}} = 0.375 \times 0.05 = \boxed{0.0188 \text{ pp}}$$

Result: Same regressor, different effects! Age 25: 1.09 pp vs Age 35: 1.88 pp

Discrete Variables: Finite Differences

For dummy variable D, use finite difference:

$$\boxed{\text{ME}_D = P(Y=1|D=1, X_-) - P(Y=1|D=0, X_-)} \tag{2}$$

where X_- are all the regressors expect D

Example: Smoking (age=35)

$$P(Y=1|\text{smoke}=0) = \frac{e^{-0.25}}{1+e^{-0.25}} \approx 0.406$$
$$P(Y=1|\text{smoke}=1) = \frac{e^{0.95}}{1+e^{0.95}} \approx 0.721$$
$$\text{ME}_{\text{smoke}} = 0.721 - 0.406 = \boxed{0.315 \text{ pp}}$$

Interpretation: Being a smoker increases probability of low birth weight by 31.5 percentage points.

AME vs MEM: Computing Average Effects

Problem: Marginal effects are heterogeneous. How to summarize?

MEM (at Mean):

- Set all $X = \bar{X}$
- Compute $f(\bar{X}\beta)$
- One effect per variable
- **Problem**: "mean person" may not exist

AME (Average):

- For each i: compute $f(X_i\beta)$
- Average across all n
- Realistic interpretation
- **PREFERRED** ✓

$$\boxed{\text{AME} = \frac{1}{n}\sum_{i=1}^{n} f(X_i\beta)\cdot\beta_k} \qquad (3)$$

Stata default: `margins, dydx(*)` gives AME

The "Trinity" of ML Tests

Testing $H_0 : R\beta = q \quad \text{vs} \quad H_1 : R\beta \neq q$

- **Example 1 (joint test):**

$$H_0 : \beta_1 = 0 \text{ and } \beta_2 = 0 \qquad H_1 : \text{at least one is non-zero}$$

- **Example 2 (linear relation):**

$$H_0 : \beta_1 = 2\beta_2 \qquad H_1 : \beta_1 \neq 2\beta_2$$

Three asymptotically equivalent approaches:

Test	Idea	Requires	Best when...
LR	Fit difference	Both models	When comparing models or coefficients
Wald	Distance from H_0	Unrestricted only	When unrestricted model is easiest to estimate
LM	Slope at H_0	Restricted only	When restricted model is much simpler (rare)

All three test statistics follow

$$\chi^2_q \quad \text{under } H_0,$$

where q **is the number of restrictions** in H_0.

Likelihood Ratio (LR) Test (1)

Formula:

$$\boxed{\mathrm{LR} = -2(\ell_R - \ell_U) = -2\log\left(\frac{L_R}{L_U}\right)} \tag{4}$$

where ℓ_R, ℓ_U = log-likelihood of restricted and unrestricted models

Logic:

- If H_0 true: $\ell_R \approx \ell_U \Rightarrow \mathrm{LR} \approx 0 \Rightarrow$ **don't reject**
- If H_0 false: $\ell_U \gg \ell_R \Rightarrow \mathrm{LR}$ large $\Rightarrow$ **reject**

Likelihood Ratio (LR) Test (2)

Example:

- $\ell_{\text{unrestricted}} = -65.4$
- $\ell_{\text{restricted}} = -68.2$
- $\text{LR} = -2(-68.2 + 65.4) = 5.6$
- At $\alpha = 0.05$: $\chi^2_{2,0.05} = 5.99 \Rightarrow$ **Do not reject (but borderline)**

What "borderline" means:

- $\text{LR} = 5.6$ is very close to the 5% critical value (5.99)
- At $\alpha = 0.05 \rightarrow$ **do not reject** H_0
- At $\alpha = 0.10 \rightarrow$ critical value $\chi^2_{2,0.10} = 4.61 \rightarrow$ **would reject** H_0

Where is the Log-Likelihood in Stata?

After estimating, for example:

```
probit low age lwt i.smoke
```

Stata prints at the top:

```
Iteration 0:  log likelihood = -120.345
Iteration 1:  log likelihood = -110.234
...
Iteration 4:  log likelihood =  -65.432
```

- The **last** line `log likelihood = -65.432` is the value at convergence.
- In the full (unrestricted) model this is ℓ_U.
- In the restricted model (with some coefficients set to zero) this is ℓ_R.
- These two numbers are all we need to compute:

$$LR = -2(\ell_R - \ell_U)$$

LR Test in Stata: Manual Computation (1)

Step 1: estimate unrestricted and restricted models

```
probit low age lwt i.smoke
scalar ll_full = e(ll)

probit low i.smoke
scalar ll_rest = e(ll)
```

Step 2: compute LR statistic

```
scalar LR = -2*(ll_rest - ll_full)
display "LR statistic = " LR
```

LR Test in Stata: Manual Computation (2)

Step 3: compute critical value and p-value

```
scalar df   = 2                    // 2 restrictions: age and lwt
scalar crit = invchi2(df, 0.95)
scalar pval = chi2tail(df, LR)

display "Critical value (5%): " crit
display "p-value: " pval
```

Step 4: compare with built-in command

```
lrtest full_model restricted_model
```

Wald Test

Formula (single coefficient):

$$\boxed{W = \left(\frac{\hat{\beta} - \beta_0}{\mathrm{SE}(\hat{\beta})} \right)^2 = t^2} \tag{5}$$

Logic:

- How many standard errors is $\hat{\beta}$ away from the value imposed by H_0?
- If far: reject H_0. If close: do not reject.
- Only requires the **unrestricted** model (easy!).

Note: After `logit` or `probit`, Stata reports **Wald tests** (based on the asymptotic normal distribution of the estimator). After `regress`, Stata instead reports **t-tests and F-tests** with their *exact finite-sample distributions* (not Wald tests).

Wald Test in Stata: Manual Computation (1)

Example: test $H_0 : \beta_{\text{age}} = 0$ in a probit model

$$\text{low} = \beta_0 + \beta_{\text{age}}\text{age} + \beta_{\text{lwt}}\text{lwt} + \beta_{\text{smoke}}\text{smoke} + u$$

Step 1: estimate the unrestricted model

```
probit low age lwt i.smoke
```

From the output, note:

- $\hat{\beta}_{\text{age}}$ (coefficient of age)
- $\text{SE}(\hat{\beta}_{\text{age}})$ (standard error)

Wald Test in Stata: Manual Computation (2)

Step 2: compute the t and Wald statistic by hand

$$t = \frac{\hat{\beta}_{\text{age}} - 0}{\text{SE}(\hat{\beta}_{\text{age}})}, \qquad W = t^2$$

```
scalar b_age  = _b[age]
scalar se_age = _se[age]

scalar t_age = b_age / se_age
scalar W_age = t_age^2
```

Numerical example:

- $\hat{\beta}_{\text{age}} = 0.0487$, $\text{SE} = 0.0156$
- $t = 0.0487/0.0156 = 3.12$
- $W = (3.12)^2 = 9.75$

Wald Test in Stata: Manual Computation (3)

Step 3: critical value and p-value

$$W \sim \chi^2_1 \quad \text{under } H_0$$

```
scalar crit_5 = invchi2(1, 0.95)
scalar p_wald = chi2tail(1, W_age)
```

Numerical example:

- Critical value at $\alpha = 0.05$: $\chi^2_{1,0.95} = 3.84$
- Since $W = 9.75 > 3.84$, we **reject** H_0

Interpretation:

- If $W > \chi^2_{1,0.95}$ or $p\_wald < 0.05 \Rightarrow$ reject H_0
- This matches the z-test and p-value shown by `probit`.

Wald Test: Additional Notes

Example recap:

- $\hat{\beta}_{\text{age}} = 0.0487$, $\text{SE} = 0.0156$
- $t = 3.12, \quad W = 9.75$
- $\chi^2_{1,0.05} = 3.84 \Rightarrow$ **Reject**

Important:

- After `logit` or `probit`, Stata reports **Wald tests** (asymptotic normal approximation).
- After `regress`, Stata reports **t-tests and F-tests** with their *exact finite-sample distributions*, not Wald tests.

LR and Wald Tests: p-value vs. critical value (1)

Setup: Both LR and Wald tests rely on a test statistic

$$T_{\text{obs}} \sim \chi^2_q \quad \text{under } H_0,$$

where q is the number of restrictions (degrees of freedom) and T_{obs} is the observed value of the test statistic.

Two equivalent approaches to conduct the test:

1. **Critical-value approach**
   - Fix a significance level α (e.g. 5%).
   - Obtain the critical value from the χ^2_q distribution:

     $$c_\alpha = \chi^2_{q,\,1-\alpha}.$$

   - **Decision rule:**
     - If $T_{\text{obs}} > c_\alpha \Rightarrow$ **Reject** H_0.
     - If $T_{\text{obs}} \leq c_\alpha \Rightarrow$ **Do not reject** H_0.

LR and Wald Tests: p-value vs. critical value (2)

2. **p-value approach**

   - Compute the p-value as the upper-tail probability:

   $$\text{p-value} = P(\chi^2_q \geq T_{\text{obs}}).$$

   - **Decision rule:**
     - ★ If p-value $< \alpha \Rightarrow$ **Reject** H_0.
     - ★ If p-value $\geq \alpha \Rightarrow$ **Do not reject** H_0.

Key point: For both LR and Wald tests, the **critical value** and **p-value** approaches are equivalent ways to reach the same decision about H_0.

In other words, the **critical-value** rule and the **p-value** rule always give the **same decision**: they are two equivalent ways of comparing T_{obs} with the same χ^2_q reference distribution.

Summary: Three Tests Compared

Test	Formula	Requires
LR	$-2(\ell_R - \ell_U)$	Both models
Wald	$(t)^2$	Unrestricted only
LM	$s_0^\top I_0^{-1} s_0$	Restricted only

Practical guide:

- **Not interested in estimates under H_0:** use the Wald test (requires only the unrestricted model)
- **Model comparison:** use the LR test (requires both models; useful for discussing sensitivity)
- **Restricted model much simpler:** use the LM test (rare for logit/probit; not covered in practice)

Notes:

- If the exam asks you to *choose* a test, any is fine as long as justified.
- If the exam requires a *specific* test, using a different one is an error.

Stata: Key Commands

Estimate model:

```
probit low age lwt i.smoke
```

Marginal effects:

```
margins, dydx(*)                // AME
margins, dydx(*) atmeans        // MEM
margins, at(age=(20(5)40))      // Predicted P at different ages
marginsplot                     // Plot results
```

Tests:

```
test age lwt                    // Wald test
probit low age lwt i.smoke
estimates store full
probit low i.smoke
estimates store restricted
lrtest full restricted          // LR test
```

Key Takeaways

✓ Logit/Probit coefficients are **NOT** marginal effects

✓ Marginal effects **vary across individuals** (heterogeneity)

✓ For interpretation, prefer **AME** over MEM

✓ Three asymptotically equivalent tests:

LR (benchmark), **Wald** (easy to compute), **LM** (rare in practice)

✓ Stata's `margins` command computes everything for you...
...*BUT in the exam you must know the math behind marginal effects formulas for both continuous and discrete regressors in logit and probit models.*

✓ You must be able to read ℓ_R and ℓ_U from Stata output and compute the LR statistic, critical value, and p-value using the χ^2 distribution.

✓ You must also know how to compute LR and Wald tests **by hand**.

Questions?

Let's apply this in Stata