Data Scaling, Functional Forms, APE, and Goodness-of-Fit in Logit and Probit

CLABE 2025/2026

Marco Rosso

1 December 2025

Block 1: Learning Outcomes

By the end of this block, you will understand:

- ✓ How scaling (rescaling) variables affects OLS coefficients, SEs, and test statistics
- ✓ Why **some OLS statistics are invariant to scaling** while others are not
- ✓ The motivation for nonlinear functional forms (level-level, log-level, log-log, level-log)
- ✓ How to interpret coefficients correctly: elasticities, semi-elasticities, and percent changes
- ✓ The special case: dummy variables in linear and log models
- ✓ Practical Stata: comparing specifications and interpreting output

Effects of Data Scaling on OLS Statistics

The Setup: Consider the model $y = \beta_0 + \beta_1 x + u$ estimated by OLS.

Now scale the regressor: Define $x^* = c \cdot x$ where c is a constant (e.g., $c = 1000$ converts euros to thousands of euros).

The new regression becomes:

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

where $x^* = c \cdot x$, so $x = \frac{x^*}{c}$.

Substituting:

$$y = \beta_0^* + \beta_1^*(c \cdot x) + u = \beta_0^* + (c \cdot \beta_1^*)x + u$$

Comparing coefficients:

$$\boxed{\beta_1^* = \frac{\beta_1}{c}} \tag{1}$$

Key insight: If you multiply a regressor by c, its coefficient is divided by c.

Scaling Invariance in OLS: What's Invariant?

Statistic	Invariant?	Explanation
Slope coefficient	NO	$\beta_1^* = \beta_1/c$
Standard error (SE)	NO	$SE(\beta_1^*) = SE(\beta_1)/c$
t-statistic	✓YES	$t = \frac{\beta_1^*}{SE(\beta_1^*)} = \frac{\beta_1/c}{SE(\beta_1)/c} = \frac{\beta_1}{SE(\beta_1)}$
p-value	✓YES	Depends only on t-stat
Fitted values $\hat{y}$	✓YES	$\hat{y} = \beta_0^* + \beta_1^* x^*$ unchanged
Residuals	✓YES	$\hat{u} = y - \hat{y}$ unchanged
R^2	✓YES	$R^2 = 1 - \frac{\sum \hat{u}^2}{\sum (y-\bar{y})^2}$ unchanged
F-statistic (overall)	✓YES	Model fit unchanged

Practical implication: Reporting results in different units (euros vs. thousands) changes the magnitude of coefficients and SEs, but does NOT affect inference (t-stats, p-values, confidence intervals).

Understanding Scaling: An Example

Example: Wage model: $\text{wage} = \beta_0 + \beta_1 \cdot \text{education} + u$

Suppose we estimate: $\widehat{\text{wage}} = 2000 + 500 \cdot \text{education} \quad (\text{R}^2 = 0.40)$
Interpretation: Each additional year of education increases wage by \$500.

Now rescale education in months: Let $\text{educ}^* = 12 \cdot \text{education}$

The new regression becomes:

$$\widehat{\text{wage}} = 2000 + \beta_1^* \cdot \text{educ}^* + u$$

We expect: $\beta_1^* = \frac{500}{12} = 41.67$

Interpretation: Each additional month of education increases wage by \$41.67.

But $\mathbf{R^2}$ is still 0.40, t-stat unchanged, p-value unchanged.

Lesson: Report scaling clearly! Use comparable units for audience interpretation.

Functional Forms in Regression

Why use nonlinear (transformed) functional forms?

- **Theoretical motivation:** Many economic relationships are not linear
  - Returns to education (diminishing returns)
  - Demand elasticity (constant vs. variable)
- **Statistical motivation:** Better fit, more stable residuals, easier interpretation
- **Interpretability:** Elasticity (percentage change) often more natural than absolute change

Four main functional forms:

Name	**Model**	**Interpretation of** β
Level-Level	$y = \beta_0 + \beta_1 x$	$\Delta y = \beta_1 \Delta x$ (absolute)
Log-Log	$\ln y = \beta_0 + \beta_1 \ln x$	$\beta_1 =$ elasticity (percent per percent)
Log-Level	$\ln y = \beta_0 + \beta_1 x$	$100\beta_1 =$ percent change in y per unit of x
Level-Log	$y = \beta_0 + \beta_1 \ln x$	$\beta_1 =$ absolute change in y per % of x

Log-Log Model: Elasticity Interpretation (1)

Model: $\ln y = \beta_0 + \beta_1 \ln x + u$

Derivation: Taking derivatives with respect to $\ln x$:

$$\frac{\partial \ln y}{\partial \ln x} = \beta_1$$

Since $\frac{\partial \ln y}{\partial \ln x} = \frac{dy/y}{dx/x}$, this is the **elasticity**:

$$\boxed{\text{Elasticity} = \frac{\%\text{ change in } y}{\%\text{ change in } x} = \beta_1} \tag{2}$$

Log-Log Model: Elasticity Interpretation (2)

Example: Demand model $\ln Q = \beta_0 + \beta_1 \ln P + u$

If $\beta_1 = -0.5$: A 1% increase in price leads to a 0.5% decrease in quantity demanded.

Advantages:

- Constant elasticity across values of x
- Natural scale for many economic variables
- Easy to compare across different units

Log-Level Model: Semi-elasticity Interpretation (1)

Model: $\ln y = \beta_0 + \beta_1 x + ur$

Derivation: Taking derivatives:

$$\frac{\partial \ln y}{\partial x} = \beta_1$$

Since $\frac{\partial \ln y}{\partial x} = \frac{1}{y}\frac{\partial y}{\partial x}$:

$$\boxed{\frac{\%\text{ change in } y}{\text{unit change in } x} \approx \beta_1 \quad \text{(semi-elasticity)}} \tag{3}$$

Interpretation rule: Multiply β_1 by 100 to get percentage change.

Log-Level Model: Semi-elasticity Interpretation (2)

Example: Wage model $\ln(\text{wage}) = \beta_0 + \beta_1 \cdot \text{education} + u$

If $\beta_1 = 0.08$: Each additional year of education increases wage by approximately $100 \times 0.08 = 8\%$.

More precisely (exact formula):

$$\% \text{ change} = 100(\exp(\beta_1) - 1) \approx 100\beta_1 \text{ (for small } \beta_1\text{)}$$

If $\beta_1 = 0.08$: exact percentage change $= 100(\exp(0.08) - 1) = 8.33\%$

Dummy Variables in Linear Models

Model: $y = \beta_0 + \beta_1 D + \beta_2 x + u$, where $D \in \{0, 1\}$

Interpretation:

- When $D = 0$: $E[y|D = 0] = \beta_0 + \beta_2 x$
- When $D = 1$: $E[y|D = 1] = (\beta_0 + \beta_1) + \beta_2 x$
- **Effect:** β_1 is the **level shift** when D changes from 0 to 1

Example: Gender wage gap

$$\ln(\text{wage}) = \beta_0 + \beta_1 \cdot \text{female} + \beta_2 \cdot \text{education} + u$$

If $\beta_1 = -0.10$: Women earn approximately 10% less than men, holding education constant.

Key insight: In a **level model**, β_1 is an absolute difference. In a **log model**, β_1 is a percentage difference.

Dummy Variables When Dependent Variable is Log (1)

Important case: $\ln y = \beta_0 + \beta_1 D + u$, where $D \in \{0, 1\}$

What does β_1 represent?

When $D = 0$: $\ln y = \beta_0 + u \Rightarrow E[\ln y] = \beta_0$
When $D = 1$: $\ln y = \beta_0 + \beta_1 + u \Rightarrow E[\ln y] = \beta_0 + \beta_1$

Taking exponentials:

$$\frac{E[y|D=1]}{E[y|D=0]} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Percentage change formula:

$$\boxed{\%\text{ change in } y = 100 \times (e^{\beta_1} - 1)} \tag{4}$$

Dummy Variables When Dependent Variable is Log (1)

Example: Sales with promotion dummy.

$$\ln(\text{sales}) = 4.5 + 0.25 \cdot \text{promotion} + u$$

If promotion: $\%$ change $= 100 \times (e^{0.25} - 1) = 100 \times 0.2840 = 28.4\%$ increase

Approximation (for small β_1): $100 \times 0.25 = 25\%$ (close enough)

Practical Example: Comparing Functional Forms in Stata

Setup: Wage data with education and experience

Model 1 (Level-Level):

```
regress wage education experience
// Output: wage = 2000 + 500*education + 100*experience
// Interpretation: 1 yr more education --> $500 more wage
```

Model 2 (Log-Level):

```
regress ln_wage education experience
// Output: ln_wage = 2.5 + 0.08*education + 0.03*experience
// Interpretation: 1 yr more education --> 8% more wage
```

Model 3 (Log-Log):

```
regress ln_wage ln_education ln_experience
// Output: ln_wage = 1.0 + 0.5*ln_education + 0.3*ln_experience
// Interpretation: 1% more education --> 0.5% more wage (elasticity)
```

Scaling check: Rescale education in months and compare

```
generate education_months = education * 12
regress wage education_months experience
// Output: education_months coefficient = 500/12 = 41.67
// Same R^2 and t-stat!
```

Block 2: Scaling in Logit/Probit and Average Partial Effects

Core question: How does scaling affect logit/probit models?

- Coefficients in logit/probit DON'T have a natural interpretation
- Scaling a regressor scales the coefficient inversely (just like OLS)
- BUT: Marginal effects change differently
- APE (Average Partial Effect) is the modern standard

In this block:

1. Scaling in logit/probit: what changes?
2. Dummy variables when y is log: exact vs. approximate
3. Definition of Average Partial Effects (APE/AME)
4. Connection to your previous marginal effects session
5. Stata implementation: 'margins' command

Scaling in Logit/Probit Models (1)

Logit model: $P(Y=1|X) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} = \Lambda(X\beta)$

Rescale regressor: $x^* = c \cdot x$. Then:

$$P(Y=1|X) = \Lambda(\beta_0 + \frac{\beta_1}{c} x^*)$$

So: $\beta_1^* = \frac{\beta_1}{c}$ (same scaling as OLS!)

But what about fitted probabilities?

$$P(Y=1|x^*) = \Lambda(\beta_0^* + \beta_1^* x^*) = \Lambda\left(\beta_0 + \frac{\beta_1}{c}(c \cdot x)\right) = \Lambda(\beta_0 + \beta_1 x)$$

Scaling in Logit/Probit Models (2)

Predicted probabilities are UNCHANGED. This is also true for marginal effects:

$$\frac{\partial P}{\partial x} = \lambda(X\beta) \cdot \beta = \text{unchanged}$$

Key insight: Just like OLS:

- Coefficient changes by factor $1/c$
- Standard error changes by factor $1/c$
- **z-statistic and p-value UNCHANGED**
- **Predicted probabilities UNCHANGED**
- **Marginal effects UNCHANGED**

Exact Formula: Dummy Regressor in Log Model (1)

Recall from Block 1: When dependent variable is log and regressor is dummy:

Model: $\ln y = \beta_0 + \beta_1 D + u$ where $D \in \{0, 1\}$

Exact percentage change:

$$\boxed{\%\text{ change} = 100 \times (e^{\beta_1} - 1)} \tag{5}$$

Examples:

- $\beta_1 = 0.10$: $\% = 100(e^{0.10} - 1) = 10.52\%$
- $\beta_1 = 0.50$: $\% = 100(e^{0.50} - 1) = 64.87\%$
- $\beta_1 = -0.10$: $\% = 100(e^{-0.10} - 1) = -9.52\%$

Exact Formula: Dummy Regressor in Log Model (2)

Approximation (linear, valid for small β_1):

$$\% \text{ change} \approx 100\beta_1$$

(Good for $|\beta_1| < 0.10$; use exact formula otherwise)

Why this matters for Stata: The 'margins' command can compute these for you with different levels of regressors.

Average Partial Effects (APE): Concept (1)

Central problem in nonlinear models:

- In OLS: marginal effect $= \beta$ (constant for everyone)
- In logit/probit: marginal effect $= f(X\beta) \cdot \beta$ (varies by person)
- How do we report ONE number?

Solution: Average Partial Effect (APE)

Definition: For a regressor x_k, the APE is:

$$\boxed{\text{APE}_{x_k} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial E[y_i|X_i]}{\partial x_{k,i}}} \tag{6}$$

Average Partial Effects (APE): Concept (2)

In words:

1. Compute the partial effect for each person in your sample
2. Average those effects
3. Report the average

For logit/probit (continuous variable):

$$\text{APE}_{x_k} = \frac{1}{n}\sum_{i=1}^{n} f(X_i\beta)\cdot\beta_k$$

For dummy variable:

$$\text{APE}_D = \frac{1}{n}\sum_{i=1}^{n}[P(y_i = 1|D = 1, X_{-D,i}) - P(y_i = 1|D = 0, X_{-D,i})]$$

APE vs. Coefficient in Logit

Key insight from your previous session:

APE is NOT the same as the coefficient!

Example: Logit model of low birth weight

Variable	**Coefficient**	**APE (at mean)**
Age	-0.024	-0.008 (pp)
Weight (lwt)	-0.007	-0.002 (pp)
Smoker (dummy)	0.417	0.144 (pp)

Interpretation:

- Coefficient -0.024 on age: NOT interpretable on its own
- APE -0.008 on age: One more year of age decreases probability of low birth weight by 0.8 percentage points (on average)
- APE 0.144 on smoking: Being a smoker increases probability by 14.4 percentage points (on average)

Lesson: Always report APE, not coefficients, for interpretation in nonlinear models!

Computing APE in Stata (1)

Command: ‘margins, dydx(*)‘ [0.15cm]

```
* Estimate probit model
probit low age lwt i.smoke

* Compute Average Partial Effects
margins, dydx(*)

* Output shows:
* age:     dy/dx = -0.008  (APE for age)
* lwt:     dy/dx = -0.002  (APE for lwt)
* smoke:   dy/dx =  0.144  (APE for smoking dummy)
```

Computing APE in Stata (2)

Variants:

- ‘margins, dydx(*) atmeans‘ → Marginal effect at the mean (old way, not recommended)
- ‘margins, at(age=(20(5)40))‘ → Predicted probabilities at different ages
- ‘marginsplot‘ → Plot the results

Key difference from your previous session:

- This session: APE defined for ANY model (linear, logit, probit, nonlinear)
- Previous session: Focused on logit/probit marginal effects specifically
- They're the same thing in a nonlinear model context!

Block 3: Goodness-of-Fit and Predictive Ability

Central question: How well does the logit/probit model predict outcomes?

In Block 3 we cover:

1. Review: OLS R^2 and why it's not applicable to logit/probit
2. Classification table: Confusion matrix approach
3. Fraction correctly predicted (accuracy)
4. Sensitivity and specificity
5. Pseudo-R^2: Definition and interpretation
6. Connecting to Wooldridge and Lecture Notes

OLS R^2 is Not Applicable to Logit/Probit (1)

OLS R^2 formula:

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (y_i - \bar{y})^2} = \frac{\text{Explained SS}}{\text{Total SS}}$$

Why this breaks for logit/probit:

- In logit/probit, $\hat{y}_i$ is a PROBABILITY (between 0 and 1), not a binary outcome
- Actual y_i is binary (0 or 1)
- Residuals $\hat{u}_i = y_i - \hat{y}_i$ are not normally distributed
- OLS R^2 becomes hard to interpret

OLS R^2 is Not Applicable to Logit/Probit (2)

Example confusion:

- If $\hat{y}_i = 0.7$ and $y_i = 1$: residual = 0.3
- If $\hat{y}_i = 0.7$ and $y_i = 0$: residual = -0.7
- These are not symmetric or normally distributed!

Solution: Use alternative goodness-of-fit measures based on:

1. **Likelihood function** (pseudo-R^2)
2. **Classification accuracy** (fraction correctly predicted)
3. **Model comparison** (LR tests)

Classification Table (Confusion Matrix)

Method: Convert predicted probabilities to binary predictions using a threshold (usually 0.5)

- If $\hat{P}(y_i = 1|X_i) > 0.5$: predict $\hat{y}_i = 1$
- If $\hat{P}(y_i = 1|X_i) \leq 0.5$: predict $\hat{y}_i = 0$

Confusion Matrix:

	Predicted		
Actual	$\hat{y} = 1$	$\hat{y} = 0$	Total
$y = 1$	n_{11}	n_{10}	$n_{1.}$
$y = 0$	n_{01}	n_{00}	$n_{0.}$
Total	$n_{.1}$	$n_{.0}$	n

Key metrics:

- **Accuracy (fraction correctly predicted):** $\frac{n_{11}+n_{00}}{n}$
- **Sensitivity (true positive rate):** $\frac{n_{11}}{n_{1.}}$ (correctly predicting $y = 1$)
- **Specificity (true negative rate):** $\frac{n_{00}}{n_{0.}}$ (correctly predicting $y = 0$)

Pseudo-R^2 for Logit/Probit (1)

Standard pseudo-R^2 (McFadden's):

$$\text{Pseudo-}R^2 = 1 - \frac{\ell_1}{\ell_0}$$

where:

- ℓ_1 = log-likelihood of the estimated model (with all regressors)
- ℓ_0 = log-likelihood of the constant-only model

Intuition:

- Ranges from 0 to 1 (not directly comparable to OLS R^2)
- Closer to 1 = better fit
- ℓ_0 = baseline (what if we only predict the marginal probability of $y = 1$?)
- ℓ_1 = what we get by adding regressors
- Pseudo-R^2 = relative improvement over baseline

Pseudo-R^2 for Logit/Probit (2)

Alternative interpretation: Related to likelihood ratio test

$$\mathrm{LR} = -2(\ell_0 - \ell_1) = -2\ell_0(1 - (1 - \text{Pseudo-}R^2))$$

Benchmark: Pseudo-R^2 around 0.2 to 0.4 is considered GOOD for logit/probit models (much lower than OLS R^2 for similar data)

References: Wooldridge and Lecture Notes

In Wooldridge (Introductory Econometrics):

- **Chapter 17:** "Limited Dependent Variable Models: Logit and Probit"
  - Section 17.1: Binary response models (logit/probit)
  - Section 17.2c: Goodness-of-fit measures
  - Discusses pseudo-R^2, fraction correctly classified, and comparison with LPM
- **Chapter 6:** "Multiple Regression Analysis: Further Issues"
  - Log functional forms and interpretations
  - Dummy variables in logs
  - Elasticities and semi-elasticities

In Lecture Notes:

- Goodness of fit in logit/probit: classification table approach
- Predictive ability: comparing LPM, logit, probit
- Pseudo-R^2 definition and examples

Practical Example: Predictive Ability in Stata

Logit model of brand choice (from your notes):

```
logit choice price promotion brand_dummy

* Pseudo-R    : reported automatically after logit
* Look for "Pseudo R2 = X.XXXX" in output

* Prediction of probabilities
predict p_hat, pr
* p_hat is Pr(choice=1|X)

* Convert to binary prediction (threshold=0.5)
generate y_pred = (p_hat > 0.5)

* Classification table (manually)
tabulate choice y_pred, matcell(confusion)
* Shows n11, n10, n01, n00

* Fraction correctly predicted
* = (n11 + n00) / n
```

Automatic classification table:

```
estat classification
* Shows confusion matrix and all metrics automatically
```

Summary: What You Should Know

Topic	Key Formula/Concept	Stata Command
Scaling in OLS	$\beta^* = \beta / c$	(just rescale variables)
Log-level model	$\% = 100\beta$ (approx)	‘regress ln_y x‘
Log-log model	$\beta =$ elasticity	‘regress ln_y ln_x‘
Dummy in log y	$\% = 100(e^{\beta} - 1)$	‘regress ln_y d_var‘
APE in nonlinear	$\text{APE} = \frac{1}{n} \sum f'(X_i\beta)\beta$	‘margins, dydx(*)‘
Accuracy	$\frac{n_{11}+n_{00}}{n}$	‘estat classification‘
Pseudo-R^2	$1 - \ell_1 / \ell_0$	(automatic after logit)

Remember:

- ✓ Scaling affects coefficients but not inference or fit
- ✓ Log models give elasticities and semi-elasticities
- ✓ APE is the standard for interpreting nonlinear models
- ✓ Pseudo-R^2 and classification accuracy measure fit

Questions?

Practical assignment:

- Re-estimate your low birth weight model (or any logit/probit)
- Try scaling a regressor (e.g., age in decades instead of years)
- Verify that z-stats and p-values don't change
- Compare pseudo-R^2 with fraction correctly classified