Econometrics 2

Lecture 21: Misspecification in the Probit Model, Indirect Inference & Asymptotic Normality

AUEB | Spring Semester 2026
Taught by: Prof. S. Arvanitis
Notes digitized by: T. Kourtalis

1 Introduction to Misspecified Probit Models

In this lecture, we examine the behavior of the Probit Maximum Likelihood Estimator (MLE) under model misspecification. Specifically, we consider the case where the true distribution of the error terms deviates from the standard normal distribution assumed by the Probit model.

Let the true latent error terms be distributed according to a Student-t distribution with ν degrees of freedom:

$$\varepsilon_{(i)} \sim t_\nu, \quad \forall i = 1, \ldots, n \tag{1}$$

where ν is assumed to be "sufficiently large" (such that $\nu > 2$, ensuring that the variance of the distribution is finite).

2 Normal Approximation and the Pseudo-True Value

To analyze this framework, we rely on the normal approximation of the Student-t distribution for large degrees of freedom:

$$t_\nu \simeq N\left(0, \frac{\nu}{\nu - 2}\right) \tag{2}$$

Using this approximation, the true variance of the error term is $\sigma^2 = \frac{\nu}{\nu-2}$. Under this setup, the probability limit of the misspecified Probit MLE—known as the **pseudo-true value** and denoted by β^*—is well-defined.

It can be shown that the Probit estimator $\hat{\beta}_n$ converges in probability to this pseudo-true value:

$$\hat{\beta}_n \xrightarrow{p} \beta^* = \sqrt{\frac{\nu - 2}{\nu}}\beta_0 \tag{3}$$

where β_0 is the true underlying parameter vector.

3 Analytical Feasibility vs. General Settings

The tractability of this misspecification analysis highlights two important structural properties:

- Due to the normality approximation of the Student-t error distribution, deriving the analytical expression of the pseudo-true value is mathematically feasible.
- Furthermore, the relation between the pseudo-true value β^* and the true parameter β_0 is a simple **linear function** scaled by the standard deviation of the true error distribution.

More generally, however, establishing that a pseudo-true value exists and is unique is a highly demanding task. Deriving its closed-form analytical representation in more complex non-linear misspecifications is typically not possible.

4 Construction of a Consistent Correction Estimator

If the degrees of freedom parameter ν is known to the researcher (an assumption that is generally difficult to satisfy in empirical practice), we can exploit the linear relationship between β^* and β_0 to construct a consistent estimator for the true parameters.

By inverting the scaling function, we can adjust the misspecified Probit MLE $\hat{\beta}_n$ directly. We define the corrected estimator $\widetilde{\beta}_n$ as:

$$\widetilde{\beta}_n = \sqrt{\frac{\nu}{\nu-2}}\,\hat{\beta}_n \tag{4}$$

Applying the Continuous Mapping Theorem (CMT), the asymptotic limit of this corrected estimator is:

$$\widetilde{\beta}_n = \sqrt{\frac{\nu}{\nu-2}}\,\hat{\beta}_n \xrightarrow{p} \sqrt{\frac{\nu}{\nu-2}}\sqrt{\frac{\nu-2}{\nu}}\beta_0 = \beta_0 \tag{5}$$

Thus, by correcting for the scale discrepancy introduced by the Student-t errors, $\widetilde{\beta}_n$ recovers consistency for the true parameter vector β_0.

5 The Methodology of Indirect Inference

The correction technique developed in the previous section is a specific application of a broader econometrics literature.

Consequently, we have shown that an estimator obtained within a misspecified model framework is, on its own, inconsistent. However, if we analytically know the function that characterizes this inconsistency (the **binding function** mapping $\beta_0 \mapsto \beta^*$), or if we can approximate it numerically, we can recover a consistent estimator for the true parameter β_0 by evaluating the inverse of this function (or its approximation) at the initial inconsistent estimator.

This core principle governs the methodology of **Indirect Inference**. Under this two-stage framework:

1. **First Stage:** We employ a misspecified model (often selected for its computational ease or analytical tractability) to obtain an auxiliary, albeit inconsistent, estimator.
2. **Second Stage:** We resolve this inconsistency by using the binding function (or its numerical approximation) to map the auxiliary estimator back to a consistent estimator of the true parameter of interest β_0.

6 Transition to MLE Asymptotic Normality

Having examined the properties of the estimator under misspecification, we now proceed with the remaining asymptotic theory of the Maximum Likelihood Estimator. For the derivations that follow, we return to the standard framework where the statistical model is assumed to be **well-specified**.

For simplicity, we will continue to assume that the sample elements satisfy **(conditional) independence**, so that the average log-likelihood function takes the form:

$$l_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \ln p(z_{(i)} \mid X_{(i)}, \beta) \tag{6}$$

We note that the results presented below generalize also to cases where the aforementioned independence assumption does not hold (e.g., under dependence).

7 The Score Vector

To apply our general asymptotic theory, we assume that the log-likelihood function $l_n(\beta)$ is **twice continuously differentiable** with respect to β for all $\beta \in \Theta$.

Under this regularity condition, the **Score vector** ($s_n(\beta)$) is well-defined as the gradient of the log-likelihood function:

$$\underbrace{s_n(\beta)}_{p\times 1} := \frac{\partial l_n(\beta)}{\partial \beta} = \frac{\partial}{\partial \beta} \left[\frac{1}{n} \sum_{i=1}^{n} \ln p(z_{(i)} \mid X_{(i)}, \beta) \right] = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta} \tag{7}$$

where p represents the dimension of the parameter vector β.

7.1 Example: The Probit Model

Within the framework of the Probit model, the sample average log-likelihood function is given by:

$$l_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \left[Y_{(i)} \ln \Phi(X_{(i)}\beta) + (1 - Y_{(i)}) \ln(1 - \Phi(X_{(i)}\beta)) \right] \tag{8}$$

Differentiating this function with respect to β yields the Score vector:

$$\frac{\partial l_n(\beta)}{\partial \beta} = \frac{1}{n} \sum_{i=1}^{n} \frac{\overbrace{\left(Y_{(i)} - \Phi(X_{(i)}\beta)\right)}^{1\times 1} \phi(X_{(i)}\beta)}{\Phi(X_{(i)}\beta)\left(1 - \Phi(X_{(i)}\beta)\right)} \underbrace{X'_{(i)}}_{p\times 1} \quad \blacksquare \tag{9}$$

where $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ represents the probability density function (PDF) of the standard normal distribution $N(0,1)$.

8 Properties of the Score Vector

Let us now derive some of the fundamental asymptotic properties of the Score vector.

First, we observe that under the assumption of (conditional) independence, the joint probability density function of the sample is represented by the product of individual densities:

$$\prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta) \tag{10}$$

assuming that the true parameter vector matches β. Since this product is a valid joint probability density function over the support of the sample space $\mathcal{W}_n$, its multiple integral must integrate to 1 for any parameter vector $\beta \in \Theta$:

$$\int_{\mathcal{W}_n} \cdots \int \prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta) \, \mathrm{d}z_{(1)} \mathrm{d}z_{(2)} \ldots \mathrm{d}z_{(n)} = 1, \quad \forall \beta \in \Theta \implies \tag{11}$$

Differentiating both sides of this identity with respect to β yields:

$$\frac{\partial}{\partial \beta} \left[\int_{\mathcal{W}_n} \cdots \int \prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta) \, \mathrm{d}z_{(1)} \ldots \mathrm{d}z_{(n)} \right] = \frac{\partial 1}{\partial \beta} = \mathbf{0}_{p \times 1}, \quad \forall \beta \in \Theta \tag{12}$$

Assuming that the regularity conditions allowing the interchange of differentiation and integration are satisfied (for instance, because the requirements of the **Dominated Convergence Theorem (DCT)** hold in this model), we can bring the derivative inside the integral sign to obtain:

$$\int_{\mathcal{W}_n} \cdots \int \frac{\partial}{\partial \beta} \left[\prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta) \right] \mathrm{d}z_{(1)} \ldots \mathrm{d}z_{(n)} = \mathbf{0}_{p \times 1}, \quad \forall \beta \in \Theta \tag{*}$$

8.1 A Useful Mathematical Identity

Let $f(\beta)$ be a positive and differentiable function with respect to the parameter vector β. Applying the multivariate chain rule to the natural logarithm of $f(\beta)$, we establish the following property:

$$\frac{\partial \ln f(\beta)}{\partial \beta} = \frac{1}{f(\beta)} \frac{\partial f(\beta)}{\partial \beta} \iff \frac{\partial f(\beta)}{\partial \beta} = f(\beta) \frac{\partial \ln f(\beta)}{\partial \beta} \tag{13}$$

In the next step, we will apply this identity by substituting $f(\beta) = \prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta)$.

8.2 Expectation of the Score Vector

Therefore, defining $f(\beta) = \prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta)$ and applying the derivative identity (13), we can substitute the integrand of (*):

$$(*) \iff \int_{\mathcal{W}_n} \cdots \int \left(\frac{\partial \ln \prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta} \right) \prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta) \, \mathrm{d}z_{(1)} \dots \mathrm{d}z_{(n)} = \mathbf{0}_{p\times 1}, \quad \forall \beta \in \Theta \tag{14}$$

$$\iff \int_{\mathcal{W}_n} \cdots \int \left(\frac{\partial}{\partial \beta} \sum_{i=1}^{n} \ln p(z_{(i)} \mid X_{(i)}, \beta) \right) \prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta) \, \mathrm{d}z_{(1)} \dots \mathrm{d}z_{(n)} = \mathbf{0}_{p\times 1}, \quad \forall \beta \in \Theta \tag{15}$$

$$\iff \int_{\mathcal{W}_n} \cdots \int \left(\sum_{i=1}^{n} \frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta} \right) \prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta) \, \mathrm{d}z_{(1)} \dots \mathrm{d}z_{(n)} = \mathbf{0}_{p\times 1}, \quad \forall \beta \in \Theta \tag{16}$$

Multiplying both sides of the last equivalence by $\frac{1}{n}$, we obtain:

$$\int_{\mathcal{W}_n} \cdots \int \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta} \right) \prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta) \, \mathrm{d}z_{(1)} \dots \mathrm{d}z_{(n)} = \mathbf{0}_{p\times 1}, \quad \forall \beta \in \Theta \tag{17}$$

Recognizing the bracketed term as the sample average Score vector $s_n(\beta)$, and the remaining term as the joint probability density function of the sample, this multiple integration is exactly the definition of the expected value. Thus:

$$\mathbb{E}_\beta \left(s_n(\beta) \right) = \mathbf{0}_{p\times 1}, \quad \forall \beta \in \Theta \tag{18}$$

where the expectation $\mathbb{E}_\beta$ is taken with respect to the joint distribution of the sample corresponding to the parameter vector β.

Evaluating this property at the true parameter vector $\beta = \beta_0$, we arrive at a landmark result in maximum likelihood theory: the expectation of the Score vector at the true parameter is zero:

$$\boxed{\mathbb{E}_{\beta_0} \left(s_n(\beta_0) \right) = \mathbf{0}_{p\times 1}} \tag{19}$$

8.3 Verification for the Probit Example

To verify this result, we evaluate the Probit Score vector (9) at the true parameter vector β_0:

$$s_n(\beta_0) = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_{(i)} - \Phi(X_{(i)}\beta_0)}{\Phi(X_{(i)}\beta_0) \left(1 - \Phi(X_{(i)}\beta_0)\right)} \phi(X_{(i)}\beta_0) X'_{(i)} \tag{20}$$

Taking the expectation on both sides yields:

$$\mathbb{E} \left(s_n(\beta_0) \right) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left(\frac{Y_{(i)} - \Phi(X_{(i)}\beta_0)}{\Phi(X_{(i)}\beta_0) \left(1 - \Phi(X_{(i)}\beta_0)\right)} \phi(X_{(i)}\beta_0) X'_{(i)} \right) \tag{21}$$

By the **Law of Iterated Expectations (L.I.E.)**, we can evaluate the expectation of each term in the sum by conditioning on $X_{(i)}$:

$$\mathbb{E}\left[\mathbb{E}\left[\frac{Y_{(i)} - \Phi(X_{(i)}\beta_0)}{\Phi(X_{(i)}\beta_0)\left(1 - \Phi(X_{(i)}\beta_0)\right)}\phi(X_{(i)}\beta_0)X'_{(i)}\,\middle|\, X_{(i)}\right]\right] \tag{22}$$

Since all terms in the fraction except $Y_{(i)}$ are functions of $X_{(i)}$ (and thus behave as constants under the conditional expectation), (22) simplifies to:

$$= \mathbb{E}\left[\frac{\mathbb{E}(Y_{(i)} \mid X_{(i)}) - \Phi(X_{(i)}\beta_0)}{\Phi(X_{(i)}\beta_0)\left(1 - \Phi(X_{(i)}\beta_0)\right)}\phi(X_{(i)}\beta_0)X'_{(i)}\right] \tag{23}$$

Under correct specification, the true conditional distribution is Bernoulli:

$$Y_{(i)} \mid X_{(i)} \sim \text{Ber}\left(\Phi(X_{(i)}\beta_0)\right) \implies \mathbb{E}(Y_{(i)} \mid X_{(i)}) = \Phi(X_{(i)}\beta_0) \tag{24}$$

Substituting this expectation back into the numerator, we obtain:

$$= \mathbb{E}\left(0 \cdot \phi(X_{(i)}\beta_0)X'_{(i)}\right) = \mathbf{0}_{p\times 1} \quad \blacksquare \tag{25}$$

9 The Hessian Matrix

Next, we examine the **Hessian matrix** of the sample average log-likelihood function $l_n(\beta)$, denoted by $\mathcal{H}_n(\beta)$. This is a symmetric $p \times p$ matrix containing the second-order partial derivatives with respect to the parameter vector:

$$\mathcal{H}_n(\beta) := \frac{\partial^2 l_n(\beta)}{\partial\beta\partial\beta'} \tag{26}$$

$$= \frac{\partial s_n(\beta)}{\partial\beta'} \tag{27}$$

$$= \frac{\partial}{\partial\beta'}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial\beta}\right) \tag{28}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial\beta\partial\beta'} \tag{29}$$

Exercise

Derive the Hessian matrix $\mathcal{H}_n(\beta)$ in the context of the Probit model by differentiating the Score vector $s_n(\beta)$ with respect to β'.