# Econometrics 2

## Lecture 22: The Information Matrix Equality

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

## 1 Introduction and Differentiating the Score Identity

Recall from Lecture 21 that under correct model specification and (conditional) independence, the expected value of the sample average Score vector is zero. This fundamental property is represented by the following integral identity:

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

To derive further asymptotic properties of the maximum likelihood estimator, we differentiate both sides of this identity with respect to the transposed parameter vector $\beta'$:

$$\frac{\partial A(\beta)}{\partial \beta'} = \frac{\partial \mathbf{0}_{p \times 1}}{\partial \beta'} \quad \forall \beta \in \Theta \iff \frac{\partial A(\beta)}{\partial \beta'} = \mathbf{0}_{p \times p}, \quad \forall \beta \in \Theta \tag{2}$$

where the differentiation of the $p \times 1$ zero vector with respect to the $1 \times p$ row vector $\beta'$ yields a $p \times p$ null matrix.

## 2 Interchanging Differentiation and Integration

Assuming once again that the regularity conditions permitting the interchange of differentiation and integration are satisfied (for example, under conditions that justify the application of the Leibniz rule or the Dominated Convergence Theorem), we can bring the partial derivative operator inside the integral sign.

This yields the following expression:

$$\int_{\mathcal{W}_n} \cdots \int \frac{\partial}{\partial \beta'} \left[ \left( \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta} \right) \prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta) \right] \mathrm{d}z_{(1)} \ldots \mathrm{d}z_{(n)} = \mathbf{0}_{p \times p}, \quad \forall \beta \in \Theta \tag{3}$$

# 3 Application of the Product Rule

To evaluate the derivative inside the bracket in (3), we apply the multivariate product rule of differentiation:

$$\frac{\partial}{\partial \beta'}\left[F(\beta)G(\beta)\right] = \left(\frac{\partial F(\beta)}{\partial \beta'}\right) G(\beta) + F(\beta)\left(\frac{\partial G(\beta)}{\partial \beta'}\right) \tag{4}$$

We analyze each component of the product rule individually:

1. **Differentiating the first factor (the average Score term):** Differentiating the average Score vector with respect to $\beta'$ yields the sample average Hessian matrix of the log-likelihood function:

$$\frac{\partial}{\partial \beta'}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta}\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta \partial \beta'} \tag{5}$$

2. **Differentiating the second factor (the joint density term):** Using the logarithmic derivative identity, the derivative of the product of individual densities with respect to $\beta'$ is:

$$\frac{\partial}{\partial \beta'}\left[\prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta)\right] = \left(\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta'}\right)\prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta) \tag{6}$$

Substituting these two components back into the integral equation (3), we obtain the sum of two distinct multiple integrals:

$$\begin{aligned}
&\int_{\mathcal{W}_n}\cdots\int\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta \partial \beta'}\right)\prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta)\,\mathrm{d}z_{(1)}\ldots\mathrm{d}z_{(n)} \\
&+\int_{\mathcal{W}_n}\cdots\int\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta}\right)\left(\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta'}\right)\prod_{i=1}^{n} p(z_{(i)} \mid X_{(i)}, \beta)\,\mathrm{d}z_{(1)}\ldots\mathrm{d}z_{(n)} \\
&= \mathbf{0}_{p\times p}, \quad \forall \beta \in \Theta
\end{aligned} \tag{7}$$

This resulting equation provides the mathematical basis for deriving the asymptotic variance of the MLE, a relationship known as the **Information Matrix Equality**.

# 4 Expectation Formulation and the True Parameter Case

Recognizing that the multiple integrals in (7) weighted by the joint density product define the expectation operator $\mathbb{E}_\beta$, we can express the identity compactly as:

$$\mathbb{E}_\beta\left[\mathcal{H}_n(\beta)\right] + \mathbb{E}_\beta\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta)}{\partial \beta'}\right)\right] = \mathbf{0}_{p\times p}, \quad \forall \beta \in \Theta \tag{8}$$

where we split the factor $\frac{1}{n}$ into $\frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}}$ and grouped each factor with its respective sum.

Evaluating this identity at the true parameter vector $\beta = \beta_0$, we write:

$$\mathbb{E}_{\beta_0}\left[\mathcal{H}_n(\beta_0)\right] + \underbrace{\mathbb{E}_{\beta_0}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta'}\right)\right]}_{B} = \mathbf{0}_{p\times p} \tag{9}$$

# 5 Decomposition of the Term B

To evaluate the expectation term $B$ under correct specification, we first re-introduce the factor $\frac{1}{n} = \frac{1}{\sqrt{n}}\frac{1}{\sqrt{n}}$ and express the product of the two single sums as a double sum:

$$B = \frac{1}{n}\mathbb{E}_{\beta_0}\left[\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta}\sum_{j=1}^{n}\frac{\partial \ln p(z_{(j)} \mid X_{(j)}, \beta_0)}{\partial \beta'}\right] \tag{10}$$

We can decompose this double sum into two distinct expected value components: the diagonal terms (where the observation indices are equal, $i = j$) and the cross-product terms (where the indices are unequal, $i \neq j$):

$$\begin{aligned} B = &\underbrace{\frac{1}{n}\mathbb{E}_{\beta_0}\left[\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta'}\right]}_{\Delta} \\ &+ \underbrace{\frac{1}{n}\mathbb{E}_{\beta_0}\left[\sum_{\substack{i,j=1 \\ i\neq j}}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta}\frac{\partial \ln p(z_{(j)} \mid X_{(j)}, \beta_0)}{\partial \beta'}\right]}_{\Gamma} \end{aligned} \tag{11}$$

# 6 Independence and the Cross-Product Term $\Gamma$

We assume that the observations $(z_{(i)}, X_{(i)})$ are independent with respect to the index $i$. Under this independence assumption, the joint expectation of the product of two distinct observations ($i \neq j$) factorizes into the product of their individual marginal expectations:

$$\Gamma = \frac{1}{n}\sum_{\substack{i,j=1 \\ i\neq j}}^{n}\mathbb{E}_{\beta_0}\left(\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta}\right)\mathbb{E}_{\beta_0}\left(\frac{\partial \ln p(z_{(j)} \mid X_{(j)}, \beta_0)}{\partial \beta'}\right) \tag{12}$$

From the fundamental properties of the Score vector derived in Lecture 21, the expectation of the individual observation Score is zero ($\mathbf{0}_{p\times 1}$ and $\mathbf{0}_{1\times p}$, respectively). Thus, the cross-product term completely vanishes:

$$\Gamma = \frac{1}{n}\sum_{\substack{i,j=1 \\ i\neq j}}^{n}\mathbf{0}_{p\times 1}\,\mathbf{0}_{1\times p} = \mathbf{0}_{p\times p} \tag{13}$$

# 7 The Diagonal Term $\Delta$ and the Fisher Information Matrix

Since the observations are identically distributed, the expected value of any individual term in the sum $\Delta$ is identical to that of the first observation ($i = 1$). This allows us to simplify the diagonal term $\Delta$ as follows:

$$\Delta = \mathbb{E}_{\beta_0}\left[\frac{\partial \ln p(z_{(1)} \mid X_{(1)}, \beta_0)}{\partial \beta} \frac{\partial \ln p(z_{(1)} \mid X_{(1)}, \beta_0)}{\partial \beta'}\right] := \mathcal{I}(\beta_0) \tag{14}$$

where $\mathcal{I}(\beta_0)$ represents the **Fisher Information Matrix** of a single observation.

Additionally, applying the identical distribution property to the expectation of the sample average Hessian matrix $\mathcal{H}_n(\beta_0)$ yields:

$$\begin{aligned}\mathbb{E}_{\beta_0}\left[\mathcal{H}_n(\beta_0)\right] &= \mathbb{E}_{\beta_0}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta \partial \beta'}\right) \\ &= \mathbb{E}_{\beta_0}\left[\frac{\partial^2 \ln p(z_{(1)} \mid X_{(1)}, \beta_0)}{\partial \beta \partial \beta'}\right] = -\mathcal{I}(\beta_0)\end{aligned} \tag{15}$$

# 8 The Fisher Information Matrix Equality

Substituting the simplified components $\Gamma = \mathbf{0}_{p\times p}$ and $\Delta = \mathcal{I}(\beta_0)$ back into the true parameter identity (9), we establish the **Fisher Information Matrix Equality** (within the context of our assumed independence framework):

$$\boxed{\mathbb{E}_{\beta_0}\left[\frac{\partial \ln p(z_{(1)} \mid X_{(1)}, \beta_0)}{\partial \beta} \frac{\partial \ln p(z_{(1)} \mid X_{(1)}, \beta_0)}{\partial \beta'}\right] = \mathcal{I}(\beta_0) = -\mathbb{E}_{\beta_0}\left[\mathcal{H}_n(\beta_0)\right]} \tag{16}$$

# 9 Asymptotic Interpretation of the Normalized Score Variance

Let us now interpret the left-hand side of this identity. We scale the sample average Score vector $s_n(\beta_0)$ by $\sqrt{n}$:

$$\sqrt{n}\, s_n(\beta_0) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta} \tag{17}$$

From our previous results, we know that the expectation of this normalized Score vector is zero:

$$\mathbb{E}_{\beta_0}\left(\sqrt{n}\, s_n(\beta_0)\right) = \sqrt{n}\, \mathbb{E}_{\beta_0}\left(s_n(\beta_0)\right) = \sqrt{n}\, \mathbf{0}_{p\times 1} = \mathbf{0}_{p\times 1} \tag{18}$$

Consequently, the variance-covariance matrix of the normalized Score vector is simply the expected value of its outer product:

$$\begin{aligned}\mathrm{Var}_{\beta_0}\left[\sqrt{n}\, s_n(\beta_0)\right] &= \mathbb{E}_{\beta_0}\left(n\, s_n(\beta_0) s_n'(\beta_0)\right) \\ &= n\,\mathbb{E}_{\beta_0}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta'}\right)\right] \\ &= \frac{1}{n}\mathbb{E}_{\beta_0}\left[\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta}\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta'}\right] \end{aligned} \tag{19}$$

Under the assumption of independence, this algebraic expression matches exactly the term $B$ defined in (9). Since $B = \Delta + \Gamma = \mathcal{I}(\beta_0) + \mathbf{0}_{p\times p} = \mathcal{I}(\beta_0)$, we obtain:

$$\boxed{\mathrm{Var}_{\beta_0}\left[\sqrt{n}\, s_n(\beta_0)\right] = \mathcal{I}(\beta_0)} \tag{20}$$

This confirms that the Fisher Information Matrix $\mathcal{I}(\beta_0)$ serves as the exact variance-covariance matrix of the normalized Score vector under the assumption of independence.

# 10 Asymptotic Normality of the MLE

We now investigate how these results impact the asymptotic properties of the Maximum Likelihood Estimator. We apply our general theory under the assumption that the true parameter vector lies in the interior of the parameter space:

$$\beta_0 \in \mathrm{Int}\,\Theta \tag{21}$$

Due to the independence assumption, the classical multivariate Central Limit Theorem (CLT) is directly applicable to the normalized Score vector:

$$\sqrt{n}\, s_n(\beta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \ln p(z_{(i)} \mid X_{(i)}, \beta_0)}{\partial \beta} \xrightarrow{d} N\left(\mathbf{0}_{p\times 1}, \mathcal{I}(\beta_0)\right) \tag{22}$$

which is highly compatible with our variance derivation.

Additionally, under standard regularity conditions, the weak Law of Large Numbers (WLLN) applies to the sample average Hessian evaluated at any consistent mean value $\beta_n^*$ (lying on the segment joining the estimator $\hat{\beta}_n$ and the true parameter $\beta_0$):

$$\mathcal{H}_n(\beta_n^*) \xrightarrow{p} \mathbb{E}_{\beta_0}\left[\mathcal{H}_1(\beta_0)\right] = -\mathcal{I}(\beta_0) \tag{23}$$

# 11 Asymptotic Efficiency and the Cramer-Rao Bound

Combining the results of the CLT (22) and the WLLN (23) via a standard Taylor expansion of the first-order conditions, our general asymptotic theory yields:

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) &\xrightarrow{d} N\left(\mathbf{0}_{p\times 1}, \left(-\mathcal{I}(\beta_0)\right)^{-1}\mathcal{I}(\beta_0)\left(-\mathcal{I}(\beta_0)\right)^{-1}\right) \\ &= N\left(\mathbf{0}_{p\times 1}, \mathcal{I}^{-1}(\beta_0)\right)\end{aligned} \tag{24}$$

This represents a crucial simplification: under correct specification and independence, the asymptotic variance-covariance matrix of the MLE collapses from the standard "sandwich" formula to the inverse of the Fisher Information Matrix, $\mathcal{I}^{-1}(\beta_0)$.

This result is of paramount importance due to the asymptotic version of the **Cramer-Rao Lemma**: the smallest asymptotic variance that a consistent, $\sqrt{n}$-consistent, and asymptotically normal estimator of $\beta_0$ can achieve is $\mathcal{I}^{-1}(\beta_0)$.

The term "smallest" is interpreted in the matrix sense: if $V$ is the asymptotic variance-covariance matrix of any other regular consistent estimator, then the matrix difference is positive semi-definite:

$$V - \mathcal{I}^{-1}(\beta_0) \succeq \mathbf{0}_{p\times p} \tag{25}$$

The matrix $\mathcal{I}^{-1}(\beta_0)$ is defined as the **Cramer-Rao Bound**, and any estimator that achieves this lower bound is called **asymptotically efficient**. Under the regularity conditions outlined, the Maximum Likelihood Estimator (MLE) achieves this bound and is therefore asymptotically efficient. *Warning:* This efficiency property does not always hold if the regularity conditions are violated.

# 12 Specialization to the Probit Model

Let us examine how these properties apply to the well-specified Probit model. To ensure complete consistency with our framework, we assume that the rows of the design matrix $X_n$ are mutually independent (ensuring that the observations are independent over $i$).

In this setting, the variance-covariance matrix of the normalized Score vector is:

$$\mathrm{Var}_{\beta_0}\left[\sqrt{n}\, s_n(\beta_0)\right] = \mathrm{Var}_{\beta_0}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\left(Y_{(i)} - \Phi(X_{(i)}\beta_0)\right)\phi(X_{(i)}\beta_0)}{\Phi(X_{(i)}\beta_0)(1-\Phi(X_{(i)}\beta_0))}X'_{(i)}\right) \tag{26}$$

Since the observations are independent and identically distributed, the variance of the sum is $n$ times the variance of a single observation. Given that the factor $1/\sqrt{n}$ squared is $1/n$, the $n$ terms cancel out. Additionally, because the expectation of each individual Score term is zero, the variance equals the expectation of its outer product. For the first observation ($i = 1$), this is given by:

$$= \mathbb{E}_{\beta_0}\left[\frac{\left(Y_{(1)} - \Phi(X_{(1)}\beta_0)\right)^2\phi^2(X_{(1)}\beta_0)}{\Phi^2(X_{(1)}\beta_0)(1-\Phi(X_{(1)}\beta_0))^2}X'_{(1)}X_{(1)}\right] \tag{27}$$

Applying the **Law of Iterated Expectations (L.I.E.)** by conditioning on $X_{(1)}$, we can rewrite this expectation as:

$$= \mathbb{E}_{\beta_0}\left[\mathbb{E}_{\beta_0}\left[\frac{\left(Y_{(1)} - \Phi(X_{(1)}\beta_0)\right)^2\phi^2(X_{(1)}\beta_0)}{\Phi^2(X_{(1)}\beta_0)(1-\Phi(X_{(1)}\beta_0))^2}X'_{(1)}X_{(1)}\,\middle|\, X_{(1)}\right]\right] \tag{28}$$

Since the terms depending solely on $X_{(1)}$ are measurable with respect to the conditioning set, the conditional expectation operator $\mathbb{E}(\cdot \mid X_{(1)})$ treats them as constants. We can therefore

pull them out:

$$= \mathbb{E}_{\beta_0}\left[\underbrace{\mathbb{E}_{\beta_0}\left[\left(Y_{(1)} - \Phi(X_{(1)}\beta_0)\right)^2 \middle| X_{(1)}\right]}_{\Delta_Y} \frac{\phi^2(X_{(1)}\beta_0)}{\Phi^2(X_{(1)}\beta_0)(1-\Phi(X_{(1)}\beta_0))^2} X'_{(1)}X_{(1)}\right] \tag{29}$$

Recall that under correct specification, the conditional distribution of $Y_{(1)}$ given $X_{(1)}$ is Bernoulli:

$$Y_{(1)} \mid X_{(1)} \sim \text{Ber}\left(\Phi(X_{(1)}\beta_0)\right) \tag{30}$$

The conditional variance of a Bernoulli random variable with parameter $p = \Phi(X_{(1)}\beta_0)$ is $p(1-p)$. Thus:

$$\Delta_Y = \text{Var}_{\beta_0}\left(Y_{(1)} \mid X_{(1)}\right) = \Phi(X_{(1)}\beta_0)\left(1 - \Phi(X_{(1)}\beta_0)\right) \tag{31}$$

Substituting $\Delta_Y$ back into our expression, the term $\Phi(X_{(1)}\beta_0)\left(1-\Phi(X_{(1)}\beta_0)\right)$ cancels out one power in the denominator. This yields the final simplified expression for the variance of the normalized Probit Score vector:

$$\text{Var}_{\beta_0}\left[\sqrt{n}\, s_n(\beta_0)\right] = \mathbb{E}_{\beta_0}\left[\frac{\phi^2(X_{(1)}\beta_0)}{\Phi(X_{(1)}\beta_0)(1-\Phi(X_{(1)}\beta_0))} X'_{(1)}X_{(1)}\right] := \mathcal{I}(\beta_0) \tag{32}$$

Consequently, the Information Matrix Equality holds in this model.

### Exercise

Find the Hessian matrix $\mathcal{H}_n(\beta)$ of the Probit model and show that:

$$\mathbb{E}_{\beta_0}\left[\mathcal{H}_n(\beta_0)\right] = -\mathcal{I}(\beta_0) \tag{33}$$

*Hint:* Apply the Law of Iterated Expectations (L.I.E.) and utilize the fact that:

$$\mathbb{E}_{\beta_0}\left[Y_{(1)} - \Phi(X_{(1)}\beta_0) \middle| X_{(1)}\right] = 0 \quad \blacksquare \tag{34}$$

# 13 Conclusion

Consequently, under the relevant regularity conditions (correct specification, homogeneity, independence, etc.), the Maximum Likelihood Estimator (MLE) for the Probit model satisfies our asymptotic theory and achieves the asymptotic Cramer-Rao bound, ensuring its efficiency.