Linear Models Lecture 12: MLE

Robert Gulotty

University of Chicago

April 27, 2025

Testing "Restrictions"

- Consider the following model of incumbent support:

$$\text{incumbent support} = \beta_1 + \beta_2\text{war dead} + \beta_3\text{economic growth} + \beta_4\text{inflation} + e$$

- An alternative (non-testable) theory says that voters care about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_2\text{war dead} + \beta_3(\text{economic growth-inflation}) + \beta_4\text{inflation} + e$$

- A testable theory could be that voters care about *only* about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_3(\text{economic growth-inflation}) + e$$

- This suggests $\beta_2 = 0$ and $\beta_3 + \beta_4 = 0$

Testing "Restrictions"

- Consider the following model of incumbent support:

$$\text{incumbent support} = \beta_1 + \beta_2\text{war dead} + \beta_3\text{economic growth} + \beta_4\text{inflation} + e$$

- An alternative (non-testable) theory says that voters care about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_2\text{war dead} + \beta_3(\text{economic growth-inflation}) + \beta_4\text{inflation} + e$$

- A testable theory could be that voters care about *only* about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_3(\text{economic growth-inflation}) + e$$

- This suggests $\beta_2 = 0$ and $\beta_3 + \beta_4 = 0$

Testing "Restrictions"

- Consider the following model of incumbent support:

$$\text{incumbent support} = \beta_1 + \beta_2\text{war dead} + \beta_3\text{economic growth} + \beta_4\text{inflation} + e$$

- An alternative (non-testable) theory says that voters care about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_2\text{war dead} + \beta_3(\text{economic growth-inflation}) + \beta_4\text{inflation} + e$$

- A testable theory could be that voters care about *only* about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_3(\text{economic growth-inflation}) + e$$

- This suggests $\beta_2 = 0$ and $\beta_3 + \beta_4 = 0$

Testing "Restrictions"

- Consider the following model of incumbent support:

$$\text{incumbent support} = \beta_1 + \beta_2\text{war dead} + \beta_3\text{economic growth} + \beta_4\text{inflation} + e$$

- An alternative (non-testable) theory says that voters care about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_2\text{war dead} + \beta_3(\text{economic growth-inflation}) + \beta_4\text{inflation} + e$$

- A testable theory could be that voters care about *only* about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_3(\text{economic growth-inflation}) + e$$

- This suggests $\beta_2 = 0$ and $\beta_3 + \beta_4 = 0$

Testing "Restrictions"

- Consider the following model of incumbent support:

$$\text{incumbent support} = \beta_1 + \beta_2\text{war dead} + \beta_3\text{economic growth} + \beta_4\text{inflation} + e$$

- An alternative (non-testable) theory says that voters care about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_2\text{war dead} + \beta_3(\text{economic growth-inflation}) + \beta_4\text{inflation} + e$$

- A testable theory could be that voters care about *only* about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_3(\text{economic growth-inflation}) + e$$

- This suggests $\beta_2 = 0$ and $\beta_3 + \beta_4 = 0$

Testing "Restrictions"

- Consider the following model of incumbent support:

$$\text{incumbent support} = \beta_1 + \beta_2 \text{war dead} + \beta_3 \text{economic growth} + \beta_4 \text{inflation} + e$$

- An alternative (non-testable) theory says that voters care about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_2 \text{war dead} + \beta_3 (\text{economic growth-inflation}) + \beta_4 \text{inflation} + e$$

- A testable theory could be that voters care about *only* about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_3 (\text{economic growth-inflation}) + e$$

- This suggests $\beta_2 = 0$ and $\beta_3 + \beta_4 = 0$

Testing "Restrictions"

- Consider the following model of incumbent support:

$$\text{incumbent support} = \beta_1 + \beta_2 \text{war dead} + \beta_3 \text{economic growth} + \beta_4 \text{inflation} + e$$

- An alternative (non-testable) theory says that voters care about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_2 \text{war dead} + \beta_3 (\text{economic growth-inflation}) + \beta_4 \text{inflation} + e$$

- A testable theory could be that voters care about *only* about real growth (net of inflation).

$$\text{incumbent support} = \beta_1 + \beta_3 (\text{economic growth-inflation}) + e$$

- This suggests $\beta_2 = 0$ and $\beta_3 + \beta_4 = 0$

Testing "Restrictions"

- We can specify joint hypotheses by specifying them as a matrix equation:

$$\underbrace{\boldsymbol{R}}_{q\times K}\beta = \underbrace{\boldsymbol{r}}_{q\times 1}$$

- The null hypothesis $\mathbb{H}_0$: $\beta_3 = 0$ is presented as:

$$[0\ 0\ 1\ 0\ldots 0]\beta = [0]$$

- The null hypothesis $\mathbb{H}_0$: $\beta_2 = \beta_3$ and $\beta_4 + \beta_5 = 1$ is:

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}\beta = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Testing "Restrictions"

- We can specify joint hypotheses by specifying them as a matrix equation:

$$\underbrace{\boldsymbol{R}}_{q\times K}\boldsymbol{\beta} = \underbrace{\boldsymbol{r}}_{q\times 1}$$

- The null hypothesis $\mathbb{H}_0$: $\beta_3 = 0$ is presented as:

$$[0\ 0\ 1\ 0\ldots 0]\boldsymbol{\beta} = [0]$$

- The null hypothesis $\mathbb{H}_0$: $\beta_2 = \beta_3$ and $\beta_4 + \beta_5 = 1$ is:

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}\boldsymbol{\beta} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Testing "Restrictions"

- We can specify joint hypotheses by specifying them as a matrix equation:

$$\underbrace{\boldsymbol{R}}_{q \times K} \boldsymbol{\beta} = \underbrace{\boldsymbol{r}}_{q \times 1}$$

- The null hypothesis $\mathbb{H}_0$: $\beta_3 = 0$ is presented as:

$$[0\ 0\ 1\ 0 \ldots 0]\boldsymbol{\beta} = [0]$$

- The null hypothesis $\mathbb{H}_0$: $\beta_2 = \beta_3$ and $\beta_4 + \beta_5 = 1$ is:

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Testing "Restrictions"

- We can specify joint hypotheses by specifying them as a matrix equation:

$$\underbrace{\boldsymbol{R}}_{q \times K} \boldsymbol{\beta} = \underbrace{\boldsymbol{r}}_{q \times 1}$$

- The null hypothesis $\mathbb{H}_0$: $\beta_3 = 0$ is presented as:

$$[0\ 0\ 1\ 0 \ldots 0]\boldsymbol{\beta} = [0]$$

- The null hypothesis $\mathbb{H}_0$: $\beta_2 = \beta_3$ and $\beta_4 + \beta_5 = 1$ is:

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Testing "Restrictions"

- We can specify joint hypotheses by specifying them as a matrix equation:

$$\underbrace{\boldsymbol{R}}_{q\times K}\boldsymbol{\beta} = \underbrace{\boldsymbol{r}}_{q\times 1}$$

- The null hypothesis $\mathbb{H}_0$: $\beta_3 = 0$ is presented as:

$$[0\ 0\ 1\ 0\ldots 0]\boldsymbol{\beta} = [0]$$

- The null hypothesis $\mathbb{H}_0$: $\beta_2 = \beta_3$ and $\beta_4 + \beta_5 = 1$ is:

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}\boldsymbol{\beta} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Testing "Restrictions"

- We can specify joint hypotheses by specifying them as a matrix equation:

$$\underbrace{\boldsymbol{R}}_{q\times K}\boldsymbol{\beta} = \underbrace{\boldsymbol{r}}_{q\times 1}$$

- The null hypothesis $\mathbb{H}_0$: $\beta_3 = 0$ is presented as:

$$[0\ 0\ 1\ 0 \ldots 0]\boldsymbol{\beta} = [0]$$

- The null hypothesis $\mathbb{H}_0$: $\beta_2 = \beta_3$ and $\beta_4 + \beta_5 = 1$ is:

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Testing "Restrictions"

- We can specify joint hypotheses by specifying them as a matrix equation:

$$\underbrace{\boldsymbol{R}}_{q\times K}\boldsymbol{\beta} = \underbrace{\boldsymbol{r}}_{q\times 1}$$

- The null hypothesis $\mathbb{H}_0$: $\beta_3 = 0$ is presented as:

$$[0\ 0\ 1\ 0\ldots 0]\boldsymbol{\beta} = [0]$$

- The null hypothesis $\mathbb{H}_0$: $\beta_2 = \beta_3$ and $\beta_4 + \beta_5 = 1$ is:

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}\boldsymbol{\beta} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Test Statistic for Restrictions in Normal Regression

$$F = (\boldsymbol{R}\hat{\beta} - \boldsymbol{r})'(\boldsymbol{R}(Var(\hat{\beta}))\boldsymbol{R}')^{-1}(\boldsymbol{R}\hat{\beta} - \boldsymbol{r})/q \sim F(q,\ N-K)$$

or equivalently

$$F = \frac{(\boldsymbol{e}_r'\boldsymbol{e}_r - \boldsymbol{e}'\boldsymbol{e})/q}{\boldsymbol{e}'\boldsymbol{e}/(N-K)} \sim F(q,\ N-K)$$

Where $\boldsymbol{e}$ is the residual from the regression, $\boldsymbol{e}_r$ is the residual from the regression with restrictions.

Test Statistic for Restrictions in Normal Regression

$$F = (\boldsymbol{R}\hat{\beta} - \boldsymbol{r})'(\boldsymbol{R}(Var(\hat{\beta}))\boldsymbol{R}')^{-1}(\boldsymbol{R}\hat{\beta} - \boldsymbol{r})/q \sim F(q,\ N-K)$$

or equivalently

$$F = \frac{(\boldsymbol{e}_r'\boldsymbol{e}_r - \boldsymbol{e}'\boldsymbol{e})/q}{\boldsymbol{e}'\boldsymbol{e}/(N-K} \sim F(q,\ N-K)$$

Where $\boldsymbol{e}$ is the residual from the regression, $\boldsymbol{e}_r$ is the residual from the regression with restrictions.

Test Statistic for Restrictions in Normal Regression

$$F = (\boldsymbol{R}\hat{\beta} - \boldsymbol{r})'(\boldsymbol{R}(Var(\hat{\beta}))\boldsymbol{R}')^{-1}(\boldsymbol{R}\hat{\beta} - \boldsymbol{r})/q \sim F(q,\ N-K)$$

or equivalently

$$F = \frac{(\boldsymbol{e}_r'\boldsymbol{e}_r - \boldsymbol{e}'\boldsymbol{e})/q}{\boldsymbol{e}'\boldsymbol{e}/(N-K} \sim F(q,\ N-K)$$

Where $\boldsymbol{e}$ is the residual from the regression, $\boldsymbol{e}_r$ is the residual from the regression with restrictions.

Test Statistic for Restrictions in Normal Regression

$$F = (\boldsymbol{R}\hat{\beta} - \boldsymbol{r})'(\boldsymbol{R}(Var(\hat{\beta}))\boldsymbol{R}')^{-1}(\boldsymbol{R}\hat{\beta} - \boldsymbol{r})/q \sim F(q,\ N-K)$$

or equivalently

$$F = \frac{(\boldsymbol{e}_r'\boldsymbol{e}_r - \boldsymbol{e}'\boldsymbol{e})/q}{\boldsymbol{e}'\boldsymbol{e}/(N-K} \sim F(q,\ N-K)$$

Where $\boldsymbol{e}$ is the residual from the regression, $\boldsymbol{e}_r$ is the residual from the regression with restrictions.

```
library(car)
set.seed(10)
x1 <- rnorm(100)
x2 <- rnorm(100)
x3 <- rnorm(100)
y <- 5*x1 + 0.5*x2 - 0.5*x3 + rnorm(100)
m1 <- lm(y~x1+x2+x3)
linearHypothesis(m1,c("x2 = 0.5","x3 = -0.5"),test="F")
```

Prediction versus forecasts

Prediction: The best point estimate for $\bar{y}_0$ given $\boldsymbol{x}_0$ is $\hat{y}_0 = \boldsymbol{x}_0'\hat{\beta}$.

$$\bar{e}_0 = \bar{y}_0 - \hat{y}_0 = x_0'\beta - x_0'\hat{\beta} = -x_0'(\beta - \hat{\beta})$$

$$\begin{aligned} Var(\bar{e}_0) &= Var(x_0'\hat{\beta}) = x_0' Var(\hat{\beta}) x_0 \\ &= \sigma_e^2 (x_0'(\boldsymbol{X}'\boldsymbol{X})^{-1} x_0) \end{aligned}$$

Forecast: The best point estimate for a future value y_{n+1} given $\boldsymbol{x}$ is also $\boldsymbol{x}'\hat{\beta}$.

$$\begin{aligned} \hat{e}_{n+1} &= y_{n+1} - \boldsymbol{x}'\hat{\beta} = (\boldsymbol{x}'\beta + e) - \boldsymbol{x}'\hat{\beta} \\ &= e - \boldsymbol{x}'(\beta - \hat{\beta}) \end{aligned}$$

$$Var(\hat{e}_{n+1}) = Var(e - \boldsymbol{x}'(\beta - \hat{\beta})) = \sigma_e^2 + \boldsymbol{x}' Var(\hat{\beta}) \boldsymbol{x}$$

Prediction versus forecasts

Prediction: The best point estimate for $\bar{y}_0$ given $\boldsymbol{x}_0$ is $\hat{y}_0 = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}}$.

$$\bar{e}_0 = \bar{y}_0 - \hat{y}_0 = \boldsymbol{x}_0'\boldsymbol{\beta} - \boldsymbol{x}_0'\hat{\boldsymbol{\beta}} = -\boldsymbol{x}_0'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

$$\begin{aligned} Var(\bar{e}_0) &= Var(x_0'\hat{\beta}) = x_0' Var(\hat{\beta}) x_0 \\ &= \sigma_e^2 (x_0'(\boldsymbol{X}'\boldsymbol{X})^{-1} x_0) \end{aligned}$$

Forecast: The best point estimate for a future value y_{n+1} given $\boldsymbol{x}$ is also $\boldsymbol{x}'\hat{\beta}$.

$$\begin{aligned} \hat{e}_{n+1} &= y_{n+1} - \boldsymbol{x}'\hat{\beta} = (\boldsymbol{x}'\beta + e) - \boldsymbol{x}'\hat{\beta} \\ &= e - \boldsymbol{x}'(\beta - \hat{\beta}) \end{aligned}$$

$$Var(\hat{e}_{n+1}) = Var(e - \boldsymbol{x}'(\beta - \hat{\beta})) = \sigma_e^2 + \boldsymbol{x}' Var(\hat{\beta}) \boldsymbol{x}$$

Prediction versus forecasts

Prediction: The best point estimate for $\bar{y}_0$ given $\boldsymbol{x}_0$ is $\hat{y}_0 = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}}$.

$$\bar{e}_0 = \bar{y}_0 - \hat{y}_0 = \boldsymbol{x}_0'\boldsymbol{\beta} - \boldsymbol{x}_0'\hat{\boldsymbol{\beta}} = -\boldsymbol{x}_0'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

$$Var(\bar{e}_0) = Var(\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}) = \boldsymbol{x}_0' Var(\hat{\boldsymbol{\beta}})\boldsymbol{x}_0$$
$$= \sigma_e^2(x_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}x_0)$$

Forecast: The best point estimate for a future value y_{n+1} given $\boldsymbol{x}$ is also $\boldsymbol{x}'\hat{\beta}$.

$$\hat{e}_{n+1} = y_{n+1} - \boldsymbol{x}'\hat{\beta} = (\boldsymbol{x}'\beta + e) - \boldsymbol{x}'\hat{\beta}$$
$$= e - \boldsymbol{x}'(\beta - \hat{\beta})$$
$$Var(\hat{e}_{n+1}) = Var(e - \boldsymbol{x}'(\beta - \hat{\beta})) = \sigma_e^2 + \boldsymbol{x}' Var(\hat{\beta})\boldsymbol{x}$$

Prediction versus forecasts

Prediction: The best point estimate for $\bar{y}_0$ given $\boldsymbol{x}_0$ is $\hat{y}_0 = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}}$.

$$\begin{aligned} \bar{e}_0 &= \bar{y}_0 - \hat{y}_0 = \boldsymbol{x}_0'\boldsymbol{\beta} - \boldsymbol{x}_0'\hat{\boldsymbol{\beta}} = -\boldsymbol{x}_0'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\ Var(\bar{e}_0) &= Var(\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}) = \boldsymbol{x}_0' Var(\hat{\boldsymbol{\beta}})\boldsymbol{x}_0 \\ &= \sigma_e^2(\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0) \end{aligned}$$

Forecast: The best point estimate for a future value y_{n+1} given $\boldsymbol{x}$ is also $\boldsymbol{x}'\hat{\beta}$.

$$\begin{aligned} \hat{e}_{n+1} &= y_{n+1} - \boldsymbol{x}'\hat{\beta} = (\boldsymbol{x}'\beta + e) - \boldsymbol{x}'\hat{\beta} \\ &= e - \boldsymbol{x}'(\beta - \hat{\beta}) \\ Var(\hat{e}_{n+1}) &= Var(e - \boldsymbol{x}'(\beta - \hat{\beta})) = \sigma_e^2 + \boldsymbol{x}' Var(\hat{\beta})\boldsymbol{x} \end{aligned}$$

Prediction versus forecasts

Prediction: The best point estimate for $\bar{y}_0$ given $\boldsymbol{x}_0$ is $\hat{y}_0 = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}}$.

$$\begin{aligned}\bar{e}_0 &= \bar{y}_0 - \hat{y}_0 = \boldsymbol{x}_0'\boldsymbol{\beta} - \boldsymbol{x}_0'\hat{\boldsymbol{\beta}} = -\boldsymbol{x}_0'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\ Var(\bar{e}_0) &= Var(\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}) = \boldsymbol{x}_0' Var(\hat{\boldsymbol{\beta}})\boldsymbol{x}_0 \\ &= \sigma_e^2(\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0)\end{aligned}$$

Forecast: The best point estimate for a future value y_{n+1} given $\boldsymbol{x}$ is also $\boldsymbol{x}'\hat{\boldsymbol{\beta}}$.

$$\begin{aligned}\hat{e}_{n+1} &= y_{n+1} - \boldsymbol{x}'\hat{\boldsymbol{\beta}} = (\boldsymbol{x}'\boldsymbol{\beta} + e) - \boldsymbol{x}'\hat{\boldsymbol{\beta}} \\ &= e - \boldsymbol{x}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\end{aligned}$$

$$Var(\hat{e}_{n+1}) = Var(e - \boldsymbol{x}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})) = \sigma_e^2 + \boldsymbol{x}' Var(\hat{\boldsymbol{\beta}})\boldsymbol{x}$$

Prediction versus forecasts

Prediction: The best point estimate for $\bar{y}_0$ given $\boldsymbol{x}_0$ is $\hat{y}_0 = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}}$.

$$\begin{aligned}
\bar{e}_0 &= \bar{y}_0 - \hat{y}_0 = \boldsymbol{x}_0'\boldsymbol{\beta} - \boldsymbol{x}_0'\hat{\boldsymbol{\beta}} = -\boldsymbol{x}_0'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\
Var(\bar{e}_0) &= Var(\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}) = \boldsymbol{x}_0' Var(\hat{\boldsymbol{\beta}})\boldsymbol{x}_0 \\
&= \sigma_e^2(\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0)
\end{aligned}$$

Forecast: The best point estimate for a future value y_{n+1} given $\boldsymbol{x}$ is also $\boldsymbol{x}'\hat{\boldsymbol{\beta}}$.

$$\begin{aligned}
\hat{e}_{n+1} &= y_{n+1} - \boldsymbol{x}'\hat{\boldsymbol{\beta}} = (\boldsymbol{x}'\boldsymbol{\beta} + e) - \boldsymbol{x}'\hat{\boldsymbol{\beta}} \\
&= e - \boldsymbol{x}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\
Var(\hat{e}_{n+1}) &= Var(e - \boldsymbol{x}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})) = \sigma_e^2 + \boldsymbol{x}' Var(\hat{\boldsymbol{\beta}})\boldsymbol{x}
\end{aligned}$$

Prediction versus forecasts

Prediction: The best point estimate for $\bar{y}_0$ given $\boldsymbol{x}_0$ is $\hat{y}_0 = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}}$.

$$\begin{aligned}
\bar{e}_0 &= \bar{y}_0 - \hat{y}_0 = \boldsymbol{x}_0'\boldsymbol{\beta} - \boldsymbol{x}_0'\hat{\boldsymbol{\beta}} = -\boldsymbol{x}_0'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\
Var(\bar{e}_0) &= Var(\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}) = \boldsymbol{x}_0' Var(\hat{\boldsymbol{\beta}})\boldsymbol{x}_0 \\
&= \sigma_e^2(\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0)
\end{aligned}$$

Forecast: The best point estimate for a future value y_{n+1} given $\boldsymbol{x}$ is also $\boldsymbol{x}'\hat{\boldsymbol{\beta}}$.

$$\begin{aligned}
\hat{e}_{n+1} &= y_{n+1} - \boldsymbol{x}'\hat{\boldsymbol{\beta}} = (\boldsymbol{x}'\boldsymbol{\beta} + e) - \boldsymbol{x}'\hat{\boldsymbol{\beta}} \\
&= e - \boldsymbol{x}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\
Var(\hat{e}_{n+1}) &= Var(e - \boldsymbol{x}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})) = \sigma_e^2 + \boldsymbol{x}' Var(\hat{\boldsymbol{\beta}})\boldsymbol{x}
\end{aligned}$$

Predictions and Forecasts in R

- You can get the confidence interval for the prediction of the mean of y in R (for each value of x specified in a data frame called dtax) with
  `predict(mod, interval= "confidence", newdata=dtax)`
- You can get the forecast for a value of y for some points $\boldsymbol{x}$ with:
  `predict(mod, interval= "prediction", newdata=dtax)`

Likelihood of Probit Regression

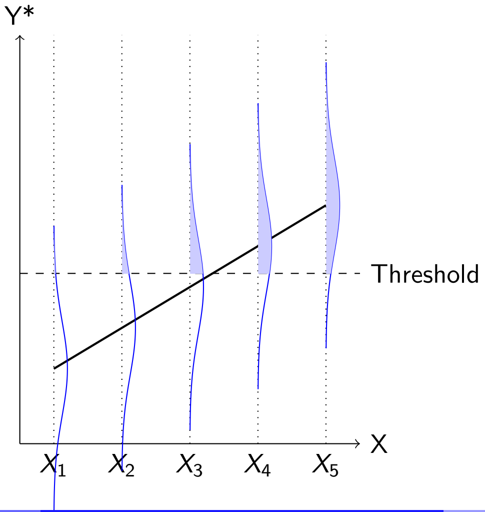
- Suppose we only observe binary manifestations of a *latent* Y^*:

$$Y = \begin{Bmatrix} 1 & Y^* > 0 \\ 0 & \text{otherwise} \end{Bmatrix} = \begin{Bmatrix} 1 & \boldsymbol{x}'\beta + e > 0 \\ 0 & \text{otherwise} \end{Bmatrix}$$

- If $e \sim N(0, 1)$, then $P(Y = 1|X) = \Phi(\boldsymbol{x}'\beta)$.
- The likelihood of a single observation $(y_i,\ x_i)$ is then:

$$\mathcal{L}(\beta; y_i,\ x_i) = \Phi(\boldsymbol{x}'\beta)^{y_i}[1 - \Phi(\boldsymbol{x}'\beta)]^{(1-y_i)}$$

Visualization of Probit Regression



Score Function in Probit Regression

- The score function for a single observation (y_i, x_i) is:

$$\frac{\partial \log \mathcal{L}(\beta; y_i, x_i)}{\partial \beta} = (y_i - \Phi(x_i'\beta))\frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)(1-\Phi(x_i'\beta))}x_i$$

- In GLMs (including probit), the score simplifies to:

$$\text{score} = (\text{residual}) \times (\text{covariate})$$

- The model says $P(Y=1|X) = g^{-1}(X'\beta)$, where g^{-1} is a CDF (link inverse).
- the derivative of Φ is ϕ and chain rule applies).
- score $= (y_i - \Phi(X_i'\beta)) \times \left(\frac{\phi(X_i'\beta)}{\Phi(X_i'\beta)(1-\Phi(X_i'\beta))}\right) X_i$
- Take the residual, weight it by the local slope, and multiply by the covariates.

Score Function in Probit Regression

- The score function for a single observation (y_i, x_i) is:

$$\frac{\partial \log \mathcal{L}(\beta; y_i, x_i)}{\partial \beta} = (y_i - \Phi(x_i'\beta)) \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)(1 - \Phi(x_i'\beta))} x_i$$

- In GLMs (including probit), the score simplifies to:

$$\text{score} = (\text{residual}) \times (\text{covariate})$$

- The model says $P(Y = 1|X) = g^{-1}(X'\beta)$, where g^{-1} is a CDF (link inverse).
- the derivative of Φ is ϕ and chain rule applies).
- score $= (y_i - \Phi(X_i'\beta)) \times \left(\frac{\phi(X_i'\beta)}{\Phi(X_i'\beta)(1-\Phi(X_i'\beta))}\right) X_i$
- Take the residual, weight it by the local slope, and multiply by the covariates.

```
probitmodel <- glm( y ~ x,
        family= binomial(link = "probit") ,
        data = df)
summary(probitmodel)
predict(probitmodel)
```

The probit slope estimates can be divided by 2.5 to make them comparable to the linear probability model estimates.

Variance and Standard Errors in Probit Regression

- The asymptotic variance of $\hat{\beta}$ is based on the Fisher Information Matrix:

$$\mathcal{I}(\beta) = \mathbb{E}\left[\left(\frac{\partial \log \mathcal{L}(\beta; y, x)}{\partial \beta}\right)\left(\frac{\partial \log \mathcal{L}(\beta; y, x)}{\partial \beta}\right)'\right]$$

- In practice, we estimate $\mathcal{I}(\hat{\beta})$ with the observed information:

$$\widehat{\mathcal{I}}(\hat{\beta}) = \sum_{i=1}^{n}\left(\frac{\phi(x_i'\hat{\beta})^2}{\Phi(x_i'\hat{\beta})(1-\Phi(x_i'\hat{\beta}))}x_i x_i'\right)$$

- Then:

$$\widehat{\text{Var}}(\hat{\beta}) = \left(\widehat{\mathcal{I}}(\hat{\beta})\right)^{-1}$$

- The **standard errors** are the square roots of the diagonal elements of $\widehat{\text{Var}}(\hat{\beta})$.

Variance and Standard Errors in Probit Regression

- The asymptotic variance of $\hat{\beta}$ is based on the Fisher Information Matrix:

$$\mathcal{I}(\beta) = \mathbb{E}\left[\left(\frac{\partial \log \mathcal{L}(\beta; y, x)}{\partial \beta}\right)\left(\frac{\partial \log \mathcal{L}(\beta; y, x)}{\partial \beta}\right)'\right]$$

- In practice, we estimate $\mathcal{I}(\hat{\beta})$ with the observed information:

$$\widehat{\mathcal{I}}(\hat{\beta}) = \sum_{i=1}^{n} \left(\frac{\phi(x_i'\hat{\beta})^2}{\Phi(x_i'\hat{\beta})(1 - \Phi(x_i'\hat{\beta}))} x_i x_i'\right)$$

- Then:

$$\widehat{\text{Var}}(\hat{\beta}) = \left(\widehat{\mathcal{I}}(\hat{\beta})\right)^{-1}$$

- The **standard errors** are the square roots of the diagonal elements of $\widehat{\text{Var}}(\hat{\beta})$.

Variance and Standard Errors in Probit Regression

- The asymptotic variance of $\hat{\beta}$ is based on the Fisher Information Matrix:

$$\mathcal{I}(\beta) = \mathbb{E}\left[\left(\frac{\partial \log \mathcal{L}(\beta; y, x)}{\partial \beta}\right)\left(\frac{\partial \log \mathcal{L}(\beta; y, x)}{\partial \beta}\right)'\right]$$

- In practice, we estimate $\mathcal{I}(\hat{\beta})$ with the observed information:

$$\widehat{\mathcal{I}}(\hat{\beta}) = \sum_{i=1}^{n}\left(\frac{\phi(x_i'\hat{\beta})^2}{\Phi(x_i'\hat{\beta})(1-\Phi(x_i'\hat{\beta}))}x_i x_i'\right)$$

- Then:

$$\widehat{\text{Var}}(\hat{\beta}) = \left(\widehat{\mathcal{I}}(\hat{\beta})\right)^{-1}$$

- The **standard errors** are the square roots of the diagonal elements of $\widehat{\text{Var}}(\hat{\beta})$.

Interpreting Probit Coefficients

- In probit regression, β_j does not directly measure the marginal effect on $P(Y = 1|X)$.
- The marginal effect of x_j is:

$$\frac{\partial P(Y = 1|X)}{\partial x_j} = \phi(X'\beta)\beta_j$$

  where $\phi(\cdot)$ is the standard normal PDF.
- **Interpretation:**
  - β_j indicates the *direction* (positive or negative) of the effect.
  - The *magnitude* of the marginal effect depends on X through $\phi(X'\beta)$.
  - Effects are largest when $X'\beta$ is near 0 (where ϕ is largest).
- Marginal effects are often evaluated:
  - **At the mean of** X (Marginal Effect at the Mean, MEM), or
  - **Averaged across the sample** (Average Marginal Effect, AME).

Interpreting Probit Coefficients

- In probit regression, β_j does not directly measure the marginal effect on $P(Y = 1|X)$.
- The marginal effect of x_j is:

$$\frac{\partial P(Y = 1|X)}{\partial x_j} = \phi(X'\beta)\beta_j$$

  where $\phi(\cdot)$ is the standard normal PDF.
- **Interpretation:**
  - β_j indicates the *direction* (positive or negative) of the effect.
  - The *magnitude* of the marginal effect depends on X through $\phi(X'\beta)$.
  - Effects are largest when $X'\beta$ is near 0 (where ϕ is largest).
- Marginal effects are often evaluated:
  - **At the mean of** X (Marginal Effect at the Mean, MEM), or
  - **Averaged across the sample** (Average Marginal Effect, AME).

Interpreting Probit Coefficients

- In probit regression, β_j does not directly measure the marginal effect on $P(Y = 1|X)$.
- The marginal effect of x_j is:

$$\frac{\partial P(Y = 1|X)}{\partial x_j} = \phi(X'\beta)\beta_j$$

  where $\phi(\cdot)$ is the standard normal PDF.
- **Interpretation:**
  - β_j indicates the *direction* (positive or negative) of the effect.
  - The *magnitude* of the marginal effect depends on X through $\phi(X'\beta)$.
  - Effects are largest when $X'\beta$ is near 0 (where ϕ is largest).
- Marginal effects are often evaluated:
  - **At the mean of** X (Marginal Effect at the Mean, MEM), or
  - **Averaged across the sample** (Average Marginal Effect, AME).

Interpreting Probit Coefficients

- In probit regression, β_j does not directly measure the marginal effect on $P(Y = 1|X)$.
- The marginal effect of x_j is:

$$\frac{\partial P(Y = 1|X)}{\partial x_j} = \phi(X'\beta)\beta_j$$

  where $\phi(\cdot)$ is the standard normal PDF.
- **Interpretation:**
  - β_j indicates the *direction* (positive or negative) of the effect.
  - The *magnitude* of the marginal effect depends on X through $\phi(X'\beta)$.
  - Effects are largest when $X'\beta$ is near 0 (where ϕ is largest).
- Marginal effects are often evaluated:
  - **At the mean of** X (Marginal Effect at the Mean, MEM), or
  - **Averaged across the sample** (Average Marginal Effect, AME).

Warnings about nonlinear models

- In linear models, heteroskedasticity is a nuisance, but doesn't affect the estimates and can be addressed (partially) with robust standard errors.
- In nonlinear models, heteroskedasticity can produce inconsistent estimates.
- Including fixed effects is also more complicated, as they don't have a de-meaning interpretation.