Unit 2. The single-equation linear model

Block D. How sure are we?

Instructor: Henry Redondo

Dpto. de Fundamentos del Análisis Económico
Universidad de Alicante

Econometrics II · Academic Year 2026/2027

Outline of Block D

1. D. How sure are we?
    - Testing hypotheses
    - The LM test

Inference

- We will first study exact inference under the additional assumption that the error term follows a normal distribution.
- Then we will remove the normality assumption and study asymptotic inference.

Exact inference under normality

- Consider the linear model

$$y = \mathbf{x}'\boldsymbol{\beta} + u$$

satisfying $OLS2$, and add the following assumption on the distribution of u conditional on $\mathbf{x}$:

$$u \mid \mathbf{x} \sim N(0, \sigma^2) \qquad OLS4$$

- $OLS4$ implies

$$E(u \mid \mathbf{x}) = 0 \qquad OLS1'$$
$$E(u^2 \mid \mathbf{x}) = \sigma^2 \qquad OLS3'$$

so all previous results under $OLS1'$, $OLS2$ and $OLS3'$ still hold.

Exact inference under normality

- Since $y = \mathbf{x}'\boldsymbol{\beta} + u$, *OLS*4 is equivalent to

$$y \mid \mathbf{x} \sim N(\mathbf{x}'\boldsymbol{\beta}, \sigma^2)$$

- Since $\{(\mathbf{x}_i', y_i)', \; i = 1, \ldots, N\}$ are *iid*, *OLS*4 holds for every observation:

$$u_i \mid \mathbf{x}_i \sim N(0, \sigma^2) \quad \Longleftrightarrow \quad y_i \mid \mathbf{x}_i \sim N(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^2), \qquad i = 1, \ldots, N$$

Exact inference under normality

- By independence across i, $u_i \mid \mathbf{X}$ has the same distribution as $u_i \mid \mathbf{x}_i$, and the u_i are independent conditional on $\mathbf{X}$. Hence

$$\mathbf{u} \mid \mathbf{X} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N) \quad \Longleftrightarrow \quad \mathbf{y} \mid \mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_N)$$

where

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_N' \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \qquad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

- Conditional on $\mathbf{X}$, $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is a linear function of $\mathbf{y}$, so $\widehat{\boldsymbol{\beta}} \mid \mathbf{X}$ is also normal.
- Since $E(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) = \boldsymbol{\beta}$ and $\mathrm{Var}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ (shown earlier under *OLS*1′, *OLS*2, *OLS*3′, all implied here), we have

$$\widehat{\boldsymbol{\beta}} \mid \mathbf{X} \sim N\big(\boldsymbol{\beta},\, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\big)$$

The t test for a single parameter

- We consider three tests, depending on the alternative hypothesis:

$$a) \quad \begin{matrix} H_0 : \beta_j = \beta_j^0 \\ H_1 : \beta_j \neq \beta_j^0 \end{matrix} \qquad b) \quad \begin{matrix} H_0 : \beta_j = \beta_j^0 \\ H_1 : \beta_j > \beta_j^0 \end{matrix} \qquad c) \quad \begin{matrix} H_0 : \beta_j = \beta_j^0 \\ H_1 : \beta_j < \beta_j^0 \end{matrix}$$

- Now, using the distribution of $\widehat{\boldsymbol{\beta}}$

$$\widehat{\boldsymbol{\beta}} \,|\, \mathbf{X} \sim N\big(\boldsymbol{\beta},\, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\big)$$

- For a single coefficient, with $a_{jj} = \big[(\mathbf{X}'\mathbf{X})^{-1}\big]_{jj}$,

$$\widehat{\beta}_j \,|\, \mathbf{X} \sim N\big(\beta_j,\, \sigma^2 a_{jj}\big)$$

Standardizing $\widehat{\beta}_j$

Subtracting the mean and dividing by the standard deviation,

$$\frac{\widehat{\beta}_j - \beta_j}{\sqrt{\sigma^2 a_{jj}}} \,\Big|\, \mathbf{X} \sim N(0,1).$$

This would be enough to test $H_0 : \beta_j = \beta_j^0$, except that σ^2 is unknown, so the ratio cannot be computed.

The natural fix is to replace σ^2 by its estimator

$$\widehat{\sigma}^2 = \frac{\widehat{\mathbf{u}}'\widehat{\mathbf{u}}}{N-k}.$$

But $\widehat{\sigma}^2$ is itself random, so the ratio is no longer $N(0,1)$. To find its distribution we need two things: the distribution of $\widehat{\sigma}^2$, and its relation to $\widehat{\beta}_j$.

The variance estimator as a quadratic form

The residuals are $\widehat{\mathbf{u}} = \mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{u}$, with

$$\mathbf{M} = \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \qquad \mathbf{M}\mathbf{X} = \mathbf{0}.$$

$\mathbf{M}$ is symmetric, idempotent ($\mathbf{M}\mathbf{M} = \mathbf{M}$) and has rank $\operatorname{tr}(\mathbf{M}) = N - k$. Therefore

$$(N-k)\widehat{\sigma}^2 = \widehat{\mathbf{u}}'\widehat{\mathbf{u}} = \mathbf{u}'\mathbf{M}'\mathbf{M}\mathbf{u} = \mathbf{u}'\mathbf{M}\mathbf{u},$$

and, writing $\mathbf{z} = \mathbf{u}/\sigma$, so that $\mathbf{z} \mid \mathbf{X} \sim N(\mathbf{0}, \mathbf{I}_N)$,

$$\frac{(N-k)\widehat{\sigma}^2}{\sigma^2} = \mathbf{z}'\mathbf{M}\mathbf{z}.$$

Its distribution follows in three steps.

Step 1. Diagonalize M

Because $\mathbf{M}$ is symmetric, it has a spectral decomposition

$$\mathbf{M} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}', \qquad \mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}_N, \qquad \mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_N).$$

Since $\mathbf{M}$ is symmetric, it can be written as $\mathbf{M} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$, with $\mathbf{Q}$ orthogonal and $\mathbf{\Lambda}$ diagonal. Since $\mathbf{M}$ is idempotent, $\mathbf{\Lambda}^2 = \mathbf{\Lambda}$, so every eigenvalue is 0 or 1. The rank of $\mathbf{M}$ equals the number of nonzero eigenvalues, so exactly $N - k$ of them equal 1.

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{I}_{N-k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_k \end{pmatrix}.$$

Step 2. Rotate the normal vector

Define $\mathbf{w} = \mathbf{Q}'\mathbf{z}$. It is a linear function of a normal vector, so it is normal, with

$$E(\mathbf{w} \mid \mathbf{X}) = \mathbf{Q}'E(\mathbf{z} \mid \mathbf{X}) = \mathbf{0}, \qquad \operatorname{Var}(\mathbf{w} \mid \mathbf{X}) = \mathbf{Q}'\mathbf{I}_N\mathbf{Q} = \mathbf{I}_N.$$

Hence $\mathbf{w} \mid \mathbf{X} \sim N(\mathbf{0}, \mathbf{I}_N)$: $w_1, \ldots, w_N$ are independent $N(0, 1)$.

Step 3. Evaluate the quadratic form

Substituting the decomposition of $\mathbf{M}$,

$$\mathbf{z}'\mathbf{M}\mathbf{z} = (\mathbf{Q}'\mathbf{z})'\mathbf{\Lambda}(\mathbf{Q}'\mathbf{z}) = \mathbf{w}'\mathbf{\Lambda}\mathbf{w} = \sum_{j=1}^{N-k} w_j^2,$$

a sum of $N-k$ squares of independent $N(0,1)$ variables. Therefore

$$\boxed{\left.\frac{(N-k)\widehat{\sigma}^2}{\sigma^2}\,\right|\,\mathbf{X} \sim \chi^2_{N-k}}$$

$\widehat{\beta}$ and $\widehat{\sigma}^2$ are independent

Both depend on $\mathbf{u}$ through linear maps:

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}, \qquad \widehat{\mathbf{u}} = \mathbf{M}\mathbf{u}.$$

So they are jointly normal given $\mathbf{X}$, with covariance

$$\operatorname{Cov}\left(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{u}} \mid \mathbf{X}\right) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{M} = \mathbf{0},$$

because $\mathbf{X}'\mathbf{M} = (\mathbf{M}\mathbf{X})' = \mathbf{0}$. For jointly normal vectors, zero covariance means independence.

Since $\widehat{\sigma}^2$ is a function of $\widehat{\mathbf{u}}$ only, $\widehat{\beta}_j$ and $\widehat{\sigma}^2$ are independent given $\mathbf{X}$.

The t statistic

Divide the standardized coefficient by the square root of the χ^2 variable over its degrees of freedom:

$$\frac{\dfrac{\widehat{\beta}_j - \beta_j}{\sqrt{\sigma^2 a_{jj}}}}{\sqrt{\dfrac{(N-k)\widehat{\sigma}^2}{\sigma^2(N-k)}}} = \frac{\widehat{\beta}_j - \beta_j}{\sqrt{\widehat{\sigma}^2 a_{jj}}} = \frac{\widehat{\beta}_j - \beta_j}{SE(\widehat{\beta}_j)}.$$

The numerator is $N(0,1)$, the term inside the root is $\chi^2_{N-k}/(N-k)$, and the two are independent. By the definition of the t distribution,

$$\boxed{\frac{\widehat{\beta}_j - \beta_j}{SE(\widehat{\beta}_j)} \,\Big|\, \mathbf{X} \sim t_{N-k}}$$

The unknown σ^2 cancels, so the statistic can be computed once β_j is fixed by H_0. The distribution does not depend on $\mathbf{X}$, so it also holds unconditionally.

The t test decision rule

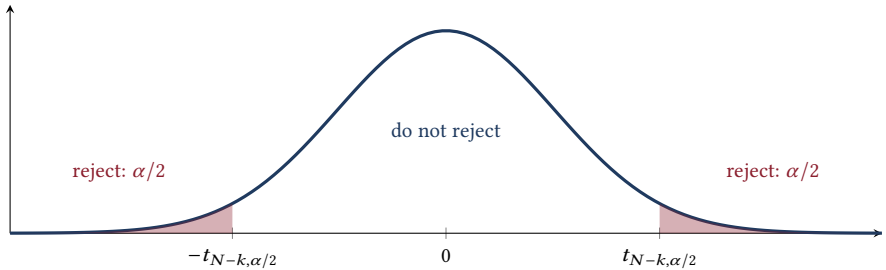
- The test statistic then is

$$t = \frac{\widehat{\beta}_j - \beta_j^0}{SE(\widehat{\beta}_j)} \sim t_{N-k} \quad \text{under } H_0$$

- If the significance level is α, the rejection regions are

$$a)\ \{|t| > t_{N-k,\alpha/2}\}$$
$$b)\ \{t > t_{N-k,\alpha}\}$$
$$c)\ \{t < -t_{N-k,\alpha}\}$$

Rejection regions at a glance



Two-sided test (a). Reject if $|t| > t_{N-k,\alpha/2}$: extreme values in *either* tail are evidence against H_0.

One-sided tests (b), (c). All of α sits in one tail: $t > t_{N-k,\alpha}$ or $t < -t_{N-k,\alpha}$.

The t test for a linear combination

- We can consider three types of tests depending on the alternative hypothesis

$$a) \quad \begin{array}{l} H_0 : \mathbf{r}'\boldsymbol{\beta} = r_0 \\ H_1 : \mathbf{r}'\boldsymbol{\beta} \neq r_0 \end{array} \qquad b) \quad \begin{array}{l} H_0 : \mathbf{r}'\boldsymbol{\beta} = r_0 \\ H_1 : \mathbf{r}'\boldsymbol{\beta} > r_0 \end{array} \qquad c) \quad \begin{array}{l} H_0 : \mathbf{r}'\boldsymbol{\beta} = r_0 \\ H_1 : \mathbf{r}'\boldsymbol{\beta} < r_0 \end{array}$$

  where $\mathbf{r}$ is a $k \times 1$ vector and r_0 is a scalar.

- Since $\widehat{\boldsymbol{\beta}} \mid \mathbf{X} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, we have that

$$\mathbf{r}'\widehat{\boldsymbol{\beta}} \mid \mathbf{X} \sim N(\mathbf{r}'\boldsymbol{\beta}, \sigma^2\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r})$$

- and we also know that

$$\frac{(N-k)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-k}$$

The t test for a linear combination (cont.)

- and since, conditional on $\mathbf{X}$, $\widehat{\sigma}^2$ and $\widehat{\boldsymbol{\beta}}$ are independent, we have that

$$\frac{\frac{\mathbf{r}'\widehat{\boldsymbol{\beta}}-\mathbf{r}'\boldsymbol{\beta}}{\sqrt{\sigma^2\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}}}}{\sqrt{\frac{(N-k)\widehat{\sigma}^2}{\sigma^2(N-k)}}} = \frac{\mathbf{r}'\widehat{\boldsymbol{\beta}}-\mathbf{r}'\boldsymbol{\beta}}{\sqrt{\widehat{\sigma}^2\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}}} = \frac{\mathbf{r}'\widehat{\boldsymbol{\beta}}-\mathbf{r}'\boldsymbol{\beta}}{\sqrt{\mathbf{r}'\widehat{var(\widehat{\boldsymbol{\beta}})}\mathbf{r}}} \sim t_{N-k}$$

- The test statistic then is

$$t = \frac{\mathbf{r}'\widehat{\boldsymbol{\beta}}-r_0}{\sqrt{\mathbf{r}'\widehat{var(\widehat{\boldsymbol{\beta}})}\mathbf{r}}} \sim t_{N-k} \quad \text{under } H_0$$

- If the significance level is α, the rejection regions are

$$\begin{aligned} &a)\ \{|t| > t_{N-k,\alpha/2}\} \\ &b)\ \{t > t_{N-k,\alpha}\} \\ &c)\ \{t < -t_{N-k,\alpha}\} \end{aligned}$$

The F test for multiple restrictions

- Let's now consider that we want to test a set of J linear restrictions

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$
$$H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$$

  where $\mathbf{R}$ is a $J \times k$ matrix and $\mathbf{r}$ is a $J \times 1$ vector.
- Since $\widehat{\boldsymbol{\beta}} \mid \mathbf{X} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, we have

$$\mathbf{R}\widehat{\boldsymbol{\beta}} \mid \mathbf{X} \sim N(\mathbf{R}\boldsymbol{\beta}, \sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')$$

  and

$$(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})'(\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}) \sim \chi^2_J$$

- We also know that

$$\frac{(N-k)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-k}$$

The F test for multiple restrictions (cont.)

- Moreover, since conditional on $\mathbf{X}$, $\widehat{\sigma}^2$ and $\widehat{\boldsymbol{\beta}}$ are independent, we have that

$$\begin{aligned}
& \frac{(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})'(\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})/J}{\frac{(N-k)\widehat{\sigma}^2}{\sigma^2(N-k)}} \\
= \; & \frac{(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})/J}{\widehat{\sigma}^2} \\
= \; & (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})'(\widehat{\sigma}^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})/J \\
= \; & (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})'(\mathbf{R}\widehat{var(\widehat{\boldsymbol{\beta}})}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})/J \sim F_{J,N-k}
\end{aligned}$$

- The test statistic then is

$$F = (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}\widehat{var(\widehat{\boldsymbol{\beta}})}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})/J \sim F_{J,N-k} \quad \text{under } H_0$$

- If the significance level is α, the rejection region is $\{F > F_{J,N-k,\alpha}\}$.

Wald tests

- When the distribution of the errors conditional on the regressors is unknown we cannot do exact inference. However, using the asymptotic approximation to the distribution of the *OLS* estimator, we can still do inference when the sample size is large.
- We know that under *OLS*1 and *OLS*2

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow_d N(0, \mathbf{B}) \tag{47}$$

where

$$\mathbf{B} = E(\mathbf{x}\mathbf{x}')^{-1} E(u^2 \mathbf{x}\mathbf{x}') E(\mathbf{x}\mathbf{x}')^{-1}$$

- Furthermore if *OLS*3 is satisfied the expression for $\mathbf{B}$ simplifies to

$$\mathbf{B} = \sigma^2 E(\mathbf{x}\mathbf{x}')^{-1}$$

Estimating the asymptotic variance for Wald tests

- We have shown above that

$$\widehat{Nvar(\widehat{\boldsymbol{\beta}})} \rightarrow_p \mathbf{B} \tag{48}$$

where

$$\widehat{var(\widehat{\boldsymbol{\beta}})} = \left(\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \sum_{i=1}^{N} \widehat{u}_i^2 \mathbf{x}_i \mathbf{x}_i' \left(\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i'\right)^{-1}$$

- When *OLS*3 is satisfied we can also estimate $\mathbf{B}$ consistently by $\widehat{Nvar(\widehat{\boldsymbol{\beta}})}$, i.e. under *OLS*3

$$\widehat{Nvar(\widehat{\boldsymbol{\beta}})} \rightarrow_p \mathbf{B} \tag{49}$$

where

$$\widehat{var(\widehat{\boldsymbol{\beta}})} = \widehat{\sigma}^2 \left(\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i'\right)^{-1} = \widehat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$$

Wald t test for a single parameter

- We consider as before three types of tests depending on the alternative hypothesis

$$a) \quad \begin{matrix} H_0 : \beta_j = \beta_j^0 \\ H_1 : \beta_j \neq \beta_j^0 \end{matrix} \qquad b) \quad \begin{matrix} H_0 : \beta_j = \beta_j^0 \\ H_1 : \beta_j > \beta_j^0 \end{matrix} \qquad c) \quad \begin{matrix} H_0 : \beta_j = \beta_j^0 \\ H_1 : \beta_j < \beta_j^0 \end{matrix}$$

- Using (47), we have that

$$\sqrt{N}(\widehat{\beta}_j - \beta_j) \rightarrow_d N(0, b_{jj})$$

where b_{jj} is the (j, j) element of $\mathbf{B}$.

- Using (48) or (49), we have that

$$N d_{jj} \rightarrow_p b_{jj}$$

where d_{jj} is the (j, j) element of $\widehat{var(\widehat{\boldsymbol{\beta}})}$ or $\widetilde{var(\widehat{\boldsymbol{\beta}})}$ depending on whether we assume that *OLS*3 is satisfied or not.

Wald t test for a single parameter (cont.)

- Then, the Wald test statistic is

$$t = \frac{\sqrt{N}(\widehat{\beta}_j - \beta_j^0)}{\sqrt{N\widehat{d}_{jj}}} = \frac{(\widehat{\beta}_j - \beta_j^0)}{\sqrt{\widehat{d}_{jj}}} = \frac{\widehat{\beta}_j - \beta_j^0}{SE(\widehat{\beta}_j)} \rightarrow_d N(0,1) \quad \text{under } H_0$$

- Notice that, if $\widehat{var(\widehat{\beta})}$ is used, the test statistic coincides with the exact test statistic under normality.
- If the significance level is α, the rejection regions are

$$a) \{|t| > z_{\alpha/2}\}$$
$$b) \{t > z_{\alpha}\}$$
$$c) \{t < -z_{\alpha}\}$$

Analysis D2 · Is a terrace worth anything?

`lm(f_log)`, robust s.e. $(N - k = 976)$

$$\widehat{\beta}_{\text{terraza}} = 0.1041, \quad \text{s.e.} = 0.0189, \quad t = 5.50, \quad p < 10^{-7}$$

95% CI: $\widehat{\beta}_{\text{terraza}} \pm t_{0.975,\,976} \cdot \text{s.e.} = [0.0669,\ 0.1413] \implies$ flats with a terrace sell for 6.9% to 15.2% more.

Analysis D2 · Is a terrace worth anything?

`lm(f_log)`, robust s.e. $(N - k = 976)$

$$\widehat{\beta}_{\text{terraza}} = 0.1041, \quad \text{s.e.} = 0.0189, \quad t = 5.50, \quad p < 10^{-7}$$

95% CI: $\widehat{\beta}_{\text{terraza}} \pm t_{0.975,\,976} \cdot \text{s.e.} = [0.0669,\ 0.1413] \implies$ flats with a terrace sell for 6.9% to 15.2% more.

- $H_0 : \beta_{\text{terraza}} = 0$ is rejected at any conventional level.
- The interval is *narrow*: the estimate is precise.

Analysis D2 · Is a terrace worth anything?

`lm(f_log)`, robust s.e. $(N - k = 976)$

$$\widehat{\beta}_{\text{terraza}} = 0.1041, \quad \text{s.e.} = 0.0189, \quad t = 5.50, \quad p < 10^{-7}$$

95% CI: $\widehat{\beta}_{\text{terraza}} \pm t_{0.975,\,976} \cdot \text{s.e.} = [0.0669,\ 0.1413] \implies$ flats with a terrace sell for 6.9% to 15.2% more.

- $H_0 : \beta_{\text{terraza}} = 0$ is rejected at any conventional level.
- The interval is *narrow*: the estimate is precise.

From Block C: holding quality fixed, the premium falls to 0.033 (about 3.4%), outside this interval. A confidence interval measures sampling noise only; it says nothing about omitted-variable bias. A precise estimate can be precisely wrong.

Wald t test for a linear combination

- We can consider three types of tests depending on the alternative hypothesis

$$a) \quad \begin{array}{l} H_0 : \mathbf{r}'\boldsymbol{\beta} = r_0 \\ H_1 : \mathbf{r}'\boldsymbol{\beta} \neq r_0 \end{array} \qquad b) \quad \begin{array}{l} H_0 : \mathbf{r}'\boldsymbol{\beta} = r_0 \\ H_1 : \mathbf{r}'\boldsymbol{\beta} > r_0 \end{array} \qquad c) \quad \begin{array}{l} H_0 : \mathbf{r}'\boldsymbol{\beta} = r_0 \\ H_1 : \mathbf{r}'\boldsymbol{\beta} < r_0 \end{array}$$

  where $\mathbf{r}$ is a $k \times 1$ vector and r_0 is a scalar.

- Using (47), we have that

$$\sqrt{N}(\mathbf{r}'\widehat{\boldsymbol{\beta}} - \mathbf{r}'\boldsymbol{\beta}) \rightarrow_d N(0, \mathbf{r}'\mathbf{B}\mathbf{r})$$

- and using (48), we have that

$$N\mathbf{r}'\widehat{var(\widehat{\boldsymbol{\beta}})}\mathbf{r} \rightarrow_p \mathbf{r}'\mathbf{B}\mathbf{r}$$

Wald t test for a linear combination (cont.)

- The Wald test statistic is

$$t = \frac{\sqrt{N}(\mathbf{r}'\widehat{\boldsymbol{\beta}} - r_0)}{\sqrt{N\mathbf{r}'\widetilde{var(\widehat{\boldsymbol{\beta}})}\mathbf{r}}} = \frac{\mathbf{r}'\widehat{\boldsymbol{\beta}} - r_0}{\sqrt{\mathbf{r}'\widetilde{var(\widehat{\boldsymbol{\beta}})}\mathbf{r}}} \rightarrow_d N(0,1) \quad \text{under } H_0$$

- If $OLS3$ is satisfied $\widetilde{var(\widehat{\boldsymbol{\beta}})}$ can be replaced by $\widehat{var(\widehat{\boldsymbol{\beta}})}$.
- Notice that, if $\widehat{var(\widehat{\boldsymbol{\beta}})}$ is used, the test statistic coincides with the exact test statistic under normality.
- If the significance level is α, the rejection regions are

$$\begin{aligned} &a)\ \{|t| > z_{\alpha/2}\} \\ &b)\ \{t > z_{\alpha}\} \\ &c)\ \{t < -z_{\alpha}\} \end{aligned}$$

Analysis D4 · The appraisers' rule

When Spanish appraisers value a flat, covered terraces typically count at **half** of the interior area. Is that what buyers pay?

In the levels model: $H_0 : \beta_{m2_terraza} = 0.5\,\beta_{m2}$, i.e. $\theta = \beta_{m2_terraza} - 0.5\,\beta_{m2} = 0$.

```
lincom(m, c(m2_terraza = 1, m2 = -0.5))
```

$\widehat{\beta}_{m2}$ = 1,615.76, $\widehat{\beta}_{m2_terraza}$ = 1,475.43 (ratio 0.913)

$$\widehat{\theta} = 667.55, \qquad \text{robust s.e.} = 431.92, \qquad t = 1.55\ (p = 0.12)$$

With *classical* s.e. (271.40): $t = 2.46$, $p = 0.014$.

Analysis D4 · The appraisers' rule

When Spanish appraisers value a flat, covered terraces typically count at **half** of the interior area. Is that what buyers pay?

In the levels model: $H_0 : \beta_{m2_terraza} = 0.5\,\beta_{m2}$, i.e. $\theta = \beta_{m2_terraza} - 0.5\,\beta_{m2} = 0$.

```
lincom(m, c(m2_terraza = 1, m2 = -0.5))
```

$\widehat{\beta}_{m2} = 1{,}615.76$, $\widehat{\beta}_{m2_terraza} = 1{,}475.43$ (ratio 0.913)

$$\widehat{\theta} = 667.55, \qquad \text{robust s.e.} = 431.92, \qquad t = 1.55\ (p = 0.12)$$

With *classical* s.e. (271.40): $t = 2.46$, $p = 0.014$.

- With robust standard errors we **cannot reject** the 50% rule. With classical ones we **would reject** it. The verdict hinges on the standard errors, and Analysis D1 showed that the levels model is strongly heteroskedastic, so the robust verdict is the valid one.
- Not rejecting is not confirming: the point estimate says a terrace m^2 is worth about 90% of an interior one. The data cannot tell 50% from 90% apart.
- Note the variance: $\mathrm{Var}(\widehat{\beta}_2 - 0.5\widehat{\beta}_1)$ needs the **covariance** of the two estimates. `lincom` takes care of it.

Wald test for multiple restrictions

- Let's now consider that we want to test a set of J linear restrictions

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$
$$H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$$

  where $\mathbf{R}$ is a $J \times k$ matrix and $\mathbf{r}$ is a $J \times 1$ vector.

- Using (47), we have that

$$\sqrt{N}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}) \rightarrow_d N(0, \mathbf{RBR}')$$

- and using (48), we have that

$$N\mathbf{R}\widehat{var(\widehat{\boldsymbol{\beta}})}\mathbf{R}' \rightarrow_p \mathbf{RBR}'$$

Wald test for multiple restrictions (cont.)

- The Wald test statistic is

$$\begin{aligned} W &= \sqrt{N}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'(N\mathbf{R}\widehat{var(\widehat{\boldsymbol{\beta}})}\mathbf{R}')^{-1}\sqrt{N}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}) \\ &= (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}\widehat{var(\widehat{\boldsymbol{\beta}})}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}) \rightarrow_d \chi^2_J \quad \text{under } H_0 \end{aligned}$$

- If $OLS3$ is satisfied $\widehat{var(\widehat{\boldsymbol{\beta}})}$ can be replaced by $\widehat{var(\widehat{\boldsymbol{\beta}})}$.
- Notice that if $\widehat{var(\widehat{\boldsymbol{\beta}})}$ is used the statistic above is equal to JF, where F is the exact test statistic under normality.
- If the significance level is α, the rejection region is $\{W > \chi^2_{J,\alpha}\}$.

The *LM* test: idea, and step 1, the restricted OLS estimator

- We test the same hypothesis as the Wald test, $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ (J linear restrictions). The Wald test only estimates the **unrestricted** model and asks whether $\mathbf{R}\widehat{\boldsymbol{\beta}}$ is far from $\mathbf{r}$. The *LM* test only estimates the **restricted** model and asks how much the fit would improve without the restrictions.
- The construction has four steps:
  1. the restricted *OLS* estimator $\widetilde{\boldsymbol{\beta}}$;
  2. its residuals, and how much larger their sum of squares is;
  3. from W to the *LM* statistic;
  4. computing LM as NR_u^2 from an auxiliary regression.
- **Step 1.** The **restricted** *OLS* **estimator** $\widetilde{\boldsymbol{\beta}}$ solves

$$\min_{\mathbf{b}} (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) \qquad s.t.\ \mathbf{R}\mathbf{b} = \mathbf{r}$$

  with $\mathbf{X}$ and $\mathbf{y}$ as in Block A.

Step 1. The Lagrangian and the first-order conditions

- The Lagrangian of this problem is

$$L(\mathbf{b}) = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) + 2\boldsymbol{\lambda}'(\mathbf{R}\mathbf{b} - \mathbf{r})$$

- The partial derivatives are

$$\begin{aligned} \frac{\partial L(\mathbf{b})}{\partial \mathbf{b}} &= -2\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b}) + 2\mathbf{R}'\boldsymbol{\lambda} \\ \frac{\partial L(\mathbf{b})}{\partial \boldsymbol{\lambda}} &= \mathbf{R}\mathbf{b} - \mathbf{r} \end{aligned}$$

- Then, the *FOC* are

$$\begin{aligned} \mathbf{X}'(\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) - \mathbf{R}'\widetilde{\boldsymbol{\lambda}} &= 0 \\ \mathbf{R}\widetilde{\boldsymbol{\beta}} - \mathbf{r} &= 0 \end{aligned}$$

Step 1. Solving for $\widetilde{\beta}$

- Multiplying the first equation by $(\mathbf{X}'\mathbf{X})^{-1}$

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\widetilde{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\widetilde{\boldsymbol{\lambda}} = 0$$

- and simplifying we have that

$$\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\widetilde{\boldsymbol{\lambda}} = 0 \tag{50}$$

- Multiplying by $\mathbf{R}$ and using the fact that $\widetilde{\boldsymbol{\beta}}$ is the restricted estimator and therefore it satisfies the restrictions (i.e. $\mathbf{R}\widetilde{\boldsymbol{\beta}} = \mathbf{r}$), we have that

$$\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r} - \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\widetilde{\boldsymbol{\lambda}} = 0$$

- Then

$$\widetilde{\boldsymbol{\lambda}} = (\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})$$

- and substituting in (50) we have that

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}) \tag{51}$$

Step 2. Restricted residuals and the increase in *SSR*

- The residuals of the restricted model are $\widetilde{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}$. Adding and subtracting $\mathbf{X}\widehat{\boldsymbol{\beta}}$ and using (51) we have

$$\begin{aligned} \widetilde{\mathbf{u}} &= \mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}) \\ &= \widehat{\mathbf{u}} + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}) \end{aligned} \tag{52}$$

- Then, since $\mathbf{X}'\widehat{\mathbf{u}} = \mathbf{0}$ (algebraic property 1), the cross terms vanish and

$$\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}} = \widehat{\mathbf{u}}'\widehat{\mathbf{u}} + (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})$$

- Imposing the restrictions can only raise the sum of squared residuals. The question is whether it raises it by much.

Step 3. From W to LM

- If $OLS1$-$OLS3$ are satisfied, the Wald statistic for $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ with the usual variance estimator is

$$\begin{aligned} W &= (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}\widehat{var(\widehat{\boldsymbol{\beta}})}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}) \\ &= \frac{(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})}{\widehat{\sigma}^2} \end{aligned}$$

- and, by step 2, its numerator is the increase in SSR:

$$W = \frac{\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}} - \widehat{\mathbf{u}}'\widehat{\mathbf{u}}}{\widehat{\sigma}^2} \rightarrow_d \chi^2_J \quad \text{under } H_0$$

- The LM test statistic is obtained by replacing the estimator of σ^2 in the denominator by $\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}}/N$ which is a consistent estimator of σ^2 under the null

$$LM = \frac{\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}} - \widehat{\mathbf{u}}'\widehat{\mathbf{u}}}{\frac{\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}}}{N}} \rightarrow_d \chi^2_J \quad \text{under } H_0 \tag{53}$$

- Notice that this LM test statistic is only valid if assumption $OLS3$ is satisfied.

Step 4. Computing LM as NR_u^2

- The LM test statistic can also be written as

$$LM = NR_u^2$$

  where R_u^2 is the $R-$squared from the auxiliary regression of $\widetilde{\mathbf{u}}$ on $\mathbf{x}$.
- *Proof*
  - Using the definition of $R-$squared

$$R_u^2 = 1 - \frac{\widehat{\widetilde{\mathbf{u}}}'\widehat{\widetilde{\mathbf{u}}}}{\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}}}$$

    where $\widehat{\widetilde{\mathbf{u}}}$ is the vector of residuals from the auxiliary regression of $\widetilde{\mathbf{u}}$ on x.
  - Then

$$NR_u^2 = N\left(1 - \frac{\widehat{\widetilde{\mathbf{u}}}'\widehat{\widetilde{\mathbf{u}}}}{\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}}}\right) = \frac{\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}} - \widehat{\widetilde{\mathbf{u}}}'\widehat{\widetilde{\mathbf{u}}}}{\frac{\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}}}{N}}$$

  - and to prove that NR_u^2 coincides with the LM test statistic in (53), we are going to show that $\widehat{\widetilde{\mathbf{u}}} = \widehat{\mathbf{u}}$.

Step 4. Proof that $LM = NR_u^2$, continued

- *Proof (cont.)*
  - The vector of residuals from this auxiliary regression is $\widetilde{\widehat{\mathbf{u}}} = \widetilde{\mathbf{u}} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widetilde{\mathbf{u}}$. Then, using the algebraic property 3 of the unrestricted *OLS* residuals, we have that

$$\widetilde{\widehat{\mathbf{u}}} = \widetilde{\mathbf{u}} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widetilde{\mathbf{u}} = \mathbf{M}\widetilde{\mathbf{u}} = \mathbf{M}(\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) = \mathbf{M}\mathbf{y} - \mathbf{M}\mathbf{X}\widetilde{\boldsymbol{\beta}} = \widehat{\mathbf{u}}$$

  - and

$$NR_u^2 = \frac{\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}} - \widetilde{\widehat{\mathbf{u}}}'\widetilde{\widehat{\mathbf{u}}}}{\frac{\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}}}{N}} = \frac{\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}} - \widehat{\mathbf{u}}'\widehat{\mathbf{u}}}{\frac{\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}}}{N}} = LM$$

Robust *LM* tests

- When *OLS*3 is not satisfied the *LM* test above is no longer valid. However, for some linear restrictions, it is possible to obtain a heteroskedasticity-robust *LM* statistic.
- In particular, in section 4.2.4 of Wooldridge book you can find a heteroskedasticity-robust *LM* statistic for testing whether a subset of the variables of the model are jointly significant.
- Notice that the test of multiple restrictions that most often arises in practice is precisely whether a subset of the explanatory variables of the model are jointly significant.

Analysis D6 · The LM test for the zones

Recipe. H_0: the four zone coefficients are zero ($J = 4$).

1. Estimate the **restricted** model by imposing H_0: drop the zone dummies and run OLS. This is $\widetilde{\boldsymbol{\beta}}$ of step 1. Keep the residuals $\widetilde{u}$.
2. Regress $\widetilde{u}$ on *all* regressors of the unrestricted model (step 4).
3. $LM = N \cdot R^2_{\widetilde{u}} \;\sim\; \chi^2_4$ under H_0.

In the data

R^2 of the auxiliary regression = 0.4172 $\Rightarrow$ $LM = 1{,}000 \times 0.4172 = 417.2$, $p \approx 10^{-89}$.

Analysis D6 · The LM test for the zones

Recipe. H_0: the four zone coefficients are zero ($J = 4$).

1. Estimate the **restricted** model by imposing H_0: drop the zone dummies and run OLS. This is $\widetilde{\boldsymbol{\beta}}$ of step 1. Keep the residuals $\widetilde{u}$.
2. Regress $\widetilde{u}$ on *all* regressors of the unrestricted model (step 4).
3. $LM = N \cdot R^2_{\widetilde{u}} \;\sim\; \chi^2_4$ under H_0.

In the data
R^2 of the auxiliary regression $= 0.4172 \;\Rightarrow\; LM = 1{,}000 \times 0.4172 = 417.2,\; p \approx 10^{-89}$.

The same conclusion as the F/Wald test, from the other side: LM only estimates the *restricted* model. Here $LM = 417$ against $4F = 699$. Under H_0 they are asymptotically equivalent; far from H_0 they differ, and $LM \leq W$.

Which test when?

Statistic	Assumptions	Distribution under H_0
t (exact)	$OLS2$, $OLS4$	t_{N-k}
F (exact)	$OLS2$, $OLS4$	$F_{J,N-k}$
Wald t, usual SE	$OLS1$–$OLS3$	$N(0,1)$ asymptotically
Wald t, robust SE	$OLS1$, $OLS2$	$N(0,1)$ asymptotically
Wald W	$OLS1$, $OLS2$ (robust) or $OLS1$–$OLS3$ (usual)	χ^2_J asymptotically
$LM = NR_u^2$	$OLS1$–$OLS3$	χ^2_J asymptotically

Useful links between the statistics.

- With the usual variance estimator: $W = J \cdot F$, and the Wald t equals the exact t.
- When $J = 1$: $W = t^2$ and the two-sided p-values coincide.
- In large samples $t_{N-k} \approx N(0,1)$ and $J \cdot F_{J,N-k} \approx \chi^2_J$.

Exact or asymptotic?

Exercise. In Analysis D4 with classical standard errors ($N - k = 976$, $J = 1$), would the exact F test and the Wald test lead to different conclusions?

Exact or asymptotic?

Exercise. In Analysis D4 with classical standard errors ($N - k = 976$, $J = 1$), would the exact F test and the Wald test lead to different conclusions?

Answer. No. With the usual variance estimator, $F = W/J = t^2 = 2.46^2 \approx 6.05$, and with 976 degrees of freedom the $F_{1,976}$ and χ^2_1 distributions are practically identical: both p-values are about 0.014.