Econ 103: Introduction to Econometrics

Lecture 8 — Estimating Error Variance & Prediction

Ryan Longmuir

UCLA

Summer Session C, 2026

Reading: Hill, Griffiths & Lim (5th ed.), §2.7, 4.1; Stock & Watson (4th ed.), §4.3.

Where we are

Last lecture we showed OLS is unbiased and BLUE, with

$$\mathrm{Var}(b_2 \mid x) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2}.$$

But every variance and standard error hides the one remaining unknown: $\boxed{\sigma^2}$, the error variance.

Today we finish the toolkit. We will

- estimate σ^2 with $\hat{\sigma}^2 = \mathrm{SSE}/(N-2)$, and read off the **standard error of the regression**,
- plug it in to finally compute the standard errors of b_1, b_2,
- and use the fitted line to make a **point prediction** of y at a new x.

After today we have estimates *and* their standard errors — the two ingredients every confidence interval and test needs.

Today's plan

1. Estimating the error variance
2. Standard errors at last
3. Point prediction

Part 1

Estimating the error variance

The last unknown parameter

Under SR2 ($\mathbb{E}(e_i \mid x) = 0$), the error variance is

$$\sigma^2 = \mathrm{Var}(e_i \mid x) = \mathbb{E}(e_i^2 \mid x).$$

A variance is an average of squared deviations, so the natural estimator is the average squared error:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{N}.$$

One fatal problem

The errors $e_i = y_i - \beta_1 - \beta_2 x_i$ are unobservable — they use the unknown β's. We can never compute this.

The fix is the move we have made all along: replace the unobservable error by its observable stand-in, the residual.

From errors to residuals, with a correction

Swap each error e_i for its residual $\hat{e}_i = y_i - b_1 - b_2 x_i$:

$$\frac{\sum \hat{e}_i^2}{N} \quad \text{is a biased estimator of } \sigma^2.$$

Why biased? OLS *chose* b_1, b_2 to make $\sum \hat{e}_i^2$ as small as possible, so residuals are a touch too small on average. The remedy is a **degrees-of-freedom** correction:

$$\boxed{\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{e}_i^2}{N-2} = \frac{\mathrm{SSE}}{N-2}} \qquad \text{and then} \qquad \mathbb{E}(\hat{\sigma}^2 \mid x) = \sigma^2.$$

- Subtract 2 because we used the data to estimate 2 parameters (β_1, β_2) before forming residuals.
- $N - 2$ is the **degrees of freedom**. (For the sample variance of one variable you divide by $N - 1$ — same logic, one parameter.)

The standard error of the regression

$\hat{\sigma}^2$ is in squared units of y. Its square root is back in the units of y and has a name:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} \qquad = \qquad \textbf{standard error of the regression (SER)}.$$

How to read the SER

$\hat{\sigma}$ measures the typical vertical distance of a data point from the fitted line — the size of a typical residual, in the units of the dependent variable. Small SER $\Rightarrow$ points hug the line; large SER $\Rightarrow$ a loose scatter.

Software labels it various ways: "S.E. of regression" (EViews), "Residual standard error" (R), "Root MSE" (Stata). Same number.

Food data: $\hat{\sigma}^2$ and the SER

Sum the 40 squared residuals (HGL Example 2.5):

$$\text{SSE} = \sum \hat{e}_i^2 = 304{,}505.2.$$

Divide by $N - 2 = 38$:

$$\hat{\sigma}^2 = \frac{304{,}505.2}{38} = 8013.29,$$

$$\hat{\sigma} = \sqrt{8013.29} = 89.52.$$

A *typical* household sits about $89.52 above or

below the fitted line — a big spread, reminding us income alone leaves much unexplained.

x	y	$\hat{y}$	$\hat{e} = y - \hat{y}$
3.69	115.22	121.09	−5.87
4.39	135.98	128.24	7.74
4.75	119.34	131.91	−12.57
6.03	114.96	144.98	−30.02
12.47	187.05	210.73	−23.68

Residuals for the first five households, using $\hat{y} = 83.42 + 10.21x$.

Part 2

Standard errors at last

Plugging $\hat{\sigma}^2$ into the variances

In Lecture 7 the estimator variances contained the unknown σ^2. Replace it with $\hat{\sigma}^2$ to get *estimated* variances:

$$\widehat{\mathrm{Var}}(b_2 \mid x) = \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}, \qquad \widehat{\mathrm{Var}}(b_1 \mid x) = \hat{\sigma}^2 \left[\frac{\sum x_i^2}{N \sum (x_i - \bar{x})^2} \right],$$

$$\widehat{\mathrm{Cov}}(b_1, b_2 \mid x) = \hat{\sigma}^2 \left[\frac{-\bar{x}}{\sum (x_i - \bar{x})^2} \right].$$

The **standard errors** are the square roots of the diagonal:

$$\mathrm{se}(b_1) = \sqrt{\widehat{\mathrm{Var}}(b_1 \mid x)}, \qquad \mathrm{se}(b_2) = \sqrt{\widehat{\mathrm{Var}}(b_2 \mid x)}.$$

Food data: the standard errors

Software arranges the estimated variances and covariance in a matrix — the **variance–covariance matrix**:

$$\begin{bmatrix} \widehat{\mathrm{Var}}(b_1) & \widehat{\mathrm{Cov}}(b_1, b_2) \\ \widehat{\mathrm{Cov}}(b_1, b_2) & \widehat{\mathrm{Var}}(b_2) \end{bmatrix} = \begin{bmatrix} 1884.44 & -85.90 \\ -85.90 & 4.3818 \end{bmatrix}$$

Square-rooting the diagonal:

$$\mathrm{se}(b_1) = \sqrt{1884.44} = 43.41,$$

$$\mathrm{se}(b_2) = \sqrt{4.3818} = 2.09.$$

Does it pass a sanity check?

Across the 10 hypothetical samples in Lecture 7, the estimates b_2 had standard deviation 1.58. Our standard error from *one* sample, 2.09, is in the same ballpark — it really is estimating the sample-to-sample spread of b_2.

A standard error is an estimate of the standard deviation of an estimator across repeated samples.

Reporting a fitted regression

The compact convention puts the standard error in parentheses beneath each estimate (HGL Example 4.3):

$$\widehat{\text{FOOD\_EXP}} = \underset{(43.41)}{83.42} + \underset{(2.09)}{10.21}\,\text{INCOME}, \qquad \hat{\sigma} = 89.52.$$

Why standard errors and not just the point estimates? Because they let any reader immediately

- gauge precision (a small se means a sharply estimated coefficient),
- build a quick interval — roughly $b_k \pm 2\,\mathrm{se}(b_k)$ when $N - 2 > 30$ (Lecture 9),
- and test hypotheses, e.g. $t = (10.21 - 8)/2.09$ for $H_0 : \beta_2 = 8$ (Lecture 10).

Estimates without standard errors are nearly useless.

Part 3

Point prediction

Predicting y at a new x

A second use of the fitted line: forecast y for a household *not* in the sample, with income x_0. By SR1 it obeys the same model,

$$y_0 = \beta_1 + \beta_2 x_0 + e_0, \qquad \mathbb{E}(e_0) = 0.$$

Split it into the systematic part and the new noise:

$$y_0 = \underbrace{\beta_1 + \beta_2 x_0}_{\text{estimate this}} + \underbrace{e_0}_{\text{best guess 0}}.$$

Estimate the systematic part by $b_1 + b_2 x_0$ and predict the noise by its mean, 0. The **least squares predictor** is the point on the fitted line:

$$\boxed{\hat{y}_0 = b_1 + b_2 x_0}$$

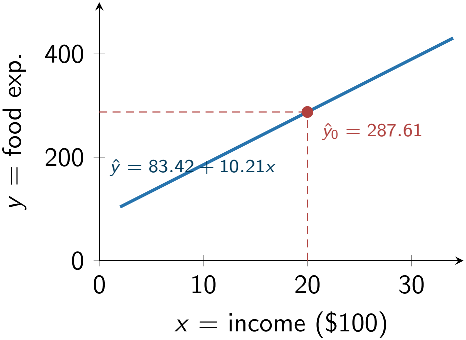
Food data: a point prediction

Predict weekly food spending for a household earning \$2,000/week, i.e. $x_0 = 20$:

$$\hat{y}_0 = 83.42 + 10.21(20) = 287.61.$$

Our best single guess is \$287.61.

Same arithmetic as "plug x into the line" — but now interpreted as a forecast for a specific out-of-sample household.



How good is the prediction?

$\hat{y}_0$ is built from the random b_1, b_2, so it is itself random. Judge it by the **forecast error**

$$f = y_0 - \hat{y}_0 = (\beta_1 + \beta_2 x_0 + e_0) - (b_1 + b_2 x_0).$$

Its conditional mean is zero:

$$\mathbb{E}(f \mid x) = \beta_1 + \beta_2 x_0 + 0 - \left[\beta_1 + \beta_2 x_0\right] = 0.$$

So $\hat{y}_0$ is an unbiased predictor — in fact the **best linear unbiased predictor** (BLUP), inheriting that title from b_1, b_2 being BLUE.

But unbiased $\neq$ pin-point. The forecast misses for *two* reasons:

- we estimated the line ($b_1, b_2 \neq \beta_1, \beta_2$), and
- the new household has its own fresh shock e_0 we cannot anticipate.

Two predictions, two questions

Be careful what you are forecasting:

The conditional mean $\mathbb{E}(y_0 \mid x_0)$

"What do households at income x_0 spend *on average*?" Uncertainty comes only from estimating the line.

The outcome y_0

"What will *this* household spend?" Adds the household's own shock e_0, so it is harder — a wider band.

Both have the same point estimate $\hat{y}_0 = b_1 + b_2 x_0$, but *different* margins of error. Prediction is also less reliable the farther x_0 is from $\bar{x}$ — we know least where we have seen least.

Looking ahead

Quantifying these margins — the **prediction interval** $\hat{y}_0 \pm t_c \, \mathrm{se}(f)$ — needs the t-distribution and the sum-of-squares decomposition. That is Lecture 11, once we have inference (Lectures 9–10) in hand.

Recap

Estimating σ^2

- errors unobservable $\Rightarrow$ use residuals
- $\hat{\sigma}^2 = \dfrac{\text{SSE}}{N-2}$ (unbiased; df $= N - 2$)
- SER $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$: typical miss from the line
- food: $\hat{\sigma}^2 = 8013.3$, $\hat{\sigma} = 89.52$

Standard errors

- $\text{se}(b_2) = \sqrt{\hat{\sigma}^2 / \sum(x_i - \bar{x})^2}$
- food: $\text{se}(b_1) = 43.41$, $\text{se}(b_2) = 2.09$

Point prediction

- $\hat{y}_0 = b_1 + b_2 x_0$ (point on the line)
- food, $x_0 = 20$: $\hat{y}_0 = 287.61$
- unbiased (BLUP); forecast error $f = y_0 - \hat{y}_0$, $\mathbb{E}(f \mid x) = 0$
- less reliable as x_0 leaves $\bar{x}$; mean vs. outcome differ

Next time (Lecture 9): confidence intervals

We have estimates and standard errors. Next: turn them into a range of plausible values for β_2, using the ***t*-distribution** — our first formal act of statistical inference.

Questions?