CHARTA LEARNING · LESSON NOTES

Regression Diagnostics

FRM Part I

1 A regression that looks fine, but is not

- A high R^2 and big t-statistics can still mislead
- Errors may be wilder in some periods than others
- Regressors may be near copies of each other
- One wild point or one missing variable bends the line

IN THE LESSON

- Picture an analyst who studies how a fund's monthly return responds to the market and two other drivers.
- She runs a regression.
- The R squared is high, the signs look sensible, and the t statistics clear the usual bar.
- She writes her report.
- Months later, the numbers fall apart.
- What went wrong?
- Perhaps the errors were far larger in turbulent months, so her standard errors were too small.
- Perhaps two of her drivers were nearly copies of each other.
- Perhaps one wild month dragged the whole line.
- Perhaps she left out a driver that mattered.
- Fitting the line is the easy part.
- Checking it is the craft, and that is today's subject.

2 The map of problems

- 1 Unequal error spread: heteroskedasticity
- 2 Tangled regressors: multicollinearity
- 3 Wrong variable list: omitted or irrelevant
- 4 Odd points: outliers and leverage
- 5 Tools: residual plots, model selection, BLUE

IN THE LESSON

- Here is the map.
- First, errors whose spread is not constant.
- Second, regressors that move together.
- Third, getting the list of variables wrong, by leaving out one that matters or adding one that does not.
- Fourth, odd points that pull on the fitted line.
- Along the way we need tools to choose between models and pictures of the residuals.
- At the end, the Gauss Markov theorem says exactly when ordinary least squares is the best linear unbiased estimator.
- Keep one idea in mind throughout.
- For every problem, ask two questions: does it bias the coefficients, or does it only damage the standard errors?
- The answers differ, and the course returns to that difference again and again.

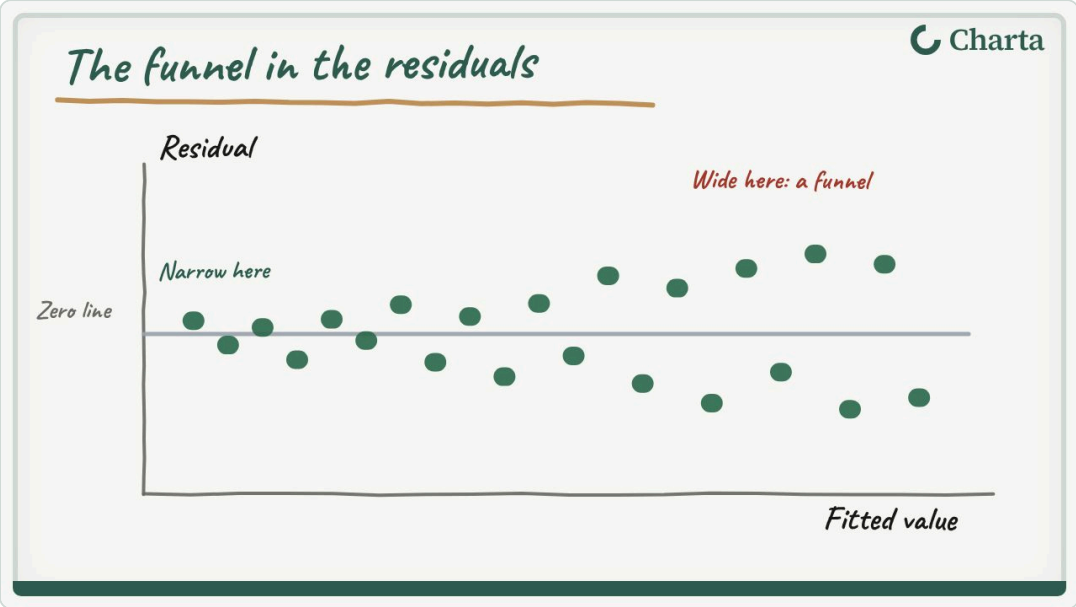
3 Heteroskedasticity in plain words

- Constant variance: same error spread everywhere
- Heteroskedastic: spread depends on regressors or fit
- OLS coefficients stay unbiased
- But the usual standard errors are wrong
- So t-tests and confidence intervals mislead

IN THE LESSON

- Start with the word.
- Homoskedasticity means the errors have the same variance whatever the values of the regressors.
- Heteroskedasticity means the variance changes.
- A natural case is the gap between predicted and actual loss, which tends to be small for small firms and large for big firms, because bigger firms simply have more going on.
- What does it break?
- Not the coefficients.
- Ordinary least squares still gives unbiased estimates of the slopes.
- What breaks is the formula for the standard errors, which assumes constant spread.
- When that assumption fails, the standard errors are wrong, so the t tests and confidence intervals built from them mislead.

4 The funnel in the residuals



IN THE LESSON

- The first check is by eye.
- Plot the residuals against the fitted values, and then against each regressor.
- If the errors have constant variance, the points form an even horizontal band around zero, with the same thickness from left to right.
- Here is what trouble looks like.
- On the left the points hug the zero line.
- As the fitted value grows, they fan out into a funnel.
- The spread changes with the fitted value, and that is heteroskedasticity.
- Plotting against each regressor in turn tells you which variable is driving the spread.
- A picture is quick, but it is a judgment call, so we also want a formal test.

5 Testing: Breusch-Pagan and White

Statistic = n × R^2_{aux}
Compare with χ^2 (q df)

Test statistic	The test statistic, n × R^2_{aux}
Sample size	Number of observations
Auxiliary R squared	R^2 of e^2 on the regressors
q = degrees of freedom	Auxiliary regressors, no intercept

IN THE LESSON

- Two formal tests are standard.
- In the Breusch Pagan test, you first run your regression and keep the residuals.
- Then you run an auxiliary regression of the squared residuals on the original regressors.
- White's test does the same but also adds the squares of the regressors and their cross products, so it can catch more shapes of variance.
- In both tests the null hypothesis is constant variance.
- The test statistic is the sample size times the R squared of the auxiliary regression.
- It is compared with a chi square distribution whose degrees of freedom, called q, equal the number of auxiliary regressors, leaving out the intercept.
- A large statistic rejects homoskedasticity.

6 Example 1: a Breusch-Pagan test

Illustration: n = 200, two regressors, auxiliary R^2 = 0.06, 5% level

Test statistic	200 × 0.06	**12.0**
Critical value	χ^2 with 2 df, 5%	**5.99**
Decision	12.0 > 5.99	**Reject**
The slip	0.06 vs 5.99, no n	**Never rejects**

IN THE LESSON

- Let's run the test with made up numbers.
- We have two hundred observations and two regressors.
- We regress the squared residuals on those two regressors and get an R squared of point zero six.
- The statistic is the sample size times that R squared: two hundred times point zero six, which is twelve.
- Now the critical value.
- The chi square distribution with two degrees of freedom, one for each auxiliary regressor, at the five per cent level, is five point nine nine.
- Twelve is greater than five point nine nine, so we reject constant variance.
- We have heteroskedasticity.
- Now see the common slip.
- If you compared the R squared itself, point zero six, with five point nine nine, you would never reject anything.
- Always multiply by the sample size first.

7 Three ways to work with it

- 1 Robust (White) standard errors: same coefficients
- 2 Transform: logs, or scale a loss by firm size
- 3 Weighted least squares: weight by 1 ÷ error variance
- Robust fixes inference only; WLS changes coefficients

IN THE LESSON

- Once you find heteroskedasticity there are three standard responses.
- The first keeps your ordinary least squares coefficients and replaces the standard errors with heteroskedasticity robust ones, often called White standard errors.
- They are valid whatever the form of the variance, so they are an easy, safe default.
- The second is to transform the data so that the variance becomes constant.
- Taking logs often does this, and so does scaling a loss by firm size when the spread grows with size.
- The third is weighted least squares, a form of generalized least squares.
- It changes the coefficients as well as the standard errors, and it is more efficient if your model of the variance is right.
- Remember the difference: robust standard errors change inference only.

8 Example 2: robust standard errors

Illustration: coefficient 0.90, usual SE 0.20, robust SE 0.30

Coefficient (unchanged)	robust SEs keep it	**0.90**
t with usual SE	0.90 ÷ 0.20	**4.50**
t with robust SE	0.90 ÷ 0.30	**3.00**
95% interval, robust	0.90 ± 1.96 × 0.30	**0.31 to 1.49**

IN THE LESSON

- Let's see the effect of robust standard errors.
- Suppose a coefficient estimate is point nine.
- The usual standard error is point two, which gives a t statistic of four point five.
- Now switch to robust standard errors.
- The coefficient does not move, it is still point nine, but the robust standard error is point three.
- The t statistic falls to three.
- A ninety five per cent interval uses one point nine six standard errors, so it runs from point nine minus point five eight eight to point nine plus point five eight eight, about point three one to one point four nine.
- The conclusion survived, but the evidence is weaker than it first looked.

9 Example 3: weighted least squares

Illustration: error variance = $\sigma^2 x^2$, with $\sigma^2 = 1$

Divide every term by x	y/x = a(1/x) + b + e/x	**new model**
Variance of e/x	$\sigma^2 x^2 \div x^2$	**σ^2**
Weight at x = 2	1 ÷ 4	**0.25**
Weight at x = 10	1 ÷ 100	**0.01**
Noisy point counts	0.25 ÷ 0.01	**25 times less**

IN THE LESSON

- Now weighted least squares.
- Suppose the error variance is sigma squared times the variable squared, so errors are bigger when the variable is bigger.
- Divide every term of the model by the variable, including the intercept.
- The model becomes the outcome over the variable equals the intercept times one over the variable, plus the slope, plus the error over the variable.
- The new error has variance sigma squared times the variable squared, divided by the variable squared, which is just sigma squared, constant.
- Dividing is the same as weighting each observation by one over its error variance.
- Setting sigma squared to one, an observation at two gets weight point two five, and one at ten gets point zero one.
- The noisy observation counts twenty five times less.

10 Multicollinearity and perfect collinearity

- Multicollinearity: high but not perfect correlation
- OLS still works; coefficients unbiased but imprecise
- Big SEs, low t, unstable estimates; R^2 and F still good
- Perfect collinearity: an exact linear combination
- Dummy trap: intercept plus a dummy for every category

IN THE LESSON

- Next, regressors that move together.
- Multicollinearity means two or more regressors are highly correlated, but not perfectly.
- Think of two interest rates that rise and fall almost together.
- The regression can still be computed, and the coefficients stay unbiased.
- The damage is to precision.
- Each coefficient has a large standard error, so the t statistics are small, and the estimates jump around from one sample to the next.
- Strangely, the R squared and the joint F test can still look strong.
- Perfect collinearity is different: one regressor is an exact linear combination of the others.
- The classic case is the dummy variable trap, an intercept plus a dummy for every category.
- Then no unique solution exists.

11 The variance inflation factor

$VIF_j = 1 \div (1 - R_j^2)$
SE is inflated by √VIF

Variance inflation factor	Factor multiplying the variance
R squared of regressor j	R^2 of regressor j on the others
Standard error	Inflated by the square root
Rule of thumb	Flag a VIF above 10

IN THE LESSON

- The variance inflation factor measures the damage for one regressor.
- Take regressor j and regress it on all the other regressors.
- Call the R squared of that regression R j squared.
- If the others explain it well, R j squared is large.
- The variance inflation factor equals one divided by one minus R j squared.
- If R j squared is zero, nothing is inflated.
- As it approaches one, the factor explodes.
- The variance of the coefficient is multiplied by the factor, so the standard error is inflated by the square root of the factor.
- A common rule of thumb flags a factor above ten.
- Perfect collinearity is a specification error, while this is a precision problem.

12 Example 4: how bad is a VIF of ten?

Illustration: $R_j^2 = 0.90$; coefficient 0.60, SE 0.10 if uncorrelated

VIF	1 ÷ (1 − 0.90)	**10**
SE inflation	√10	**3.16 times**
New standard error	0.10 × 3.16	**0.316**
New t (was 6.0)	0.60 ÷ 0.316	**1.90**

IN THE LESSON

- Suppose regressor j has an R squared of point nine when regressed on the others.
- The variance inflation factor is one divided by one minus point nine, one over point one, which is ten.
- The standard error is inflated by the square root of ten, three point one six.
- Now see the effect.
- The coefficient is point six, and with uncorrelated regressors its standard error would be point one, giving a t statistic of six.
- With the inflation, the standard error becomes point one times three point one six, about point three one six.
- The t statistic falls to point six divided by point three one six, about one point nine.
- That is below the usual critical value, so a strong effect looks insignificant.
- The coefficient has not been biased; it has only become hard to pin down.

13 A quick VIF reference

Standard error inflation = √VIF, $VIF = 1 \div (1 - R_j^2)$

R^2 of regressor j	VIF and SE inflation
0.50	VIF 2, SE × 1.41
0.75	VIF 4, SE × 2.00
0.90	VIF 10, SE × 3.16
0.99	VIF 100, SE × 10

IN THE LESSON

- Here is a quick reference.
- At an R squared of point five, the factor is two, and standard errors are inflated by about one point four one.
- At point seven five, the factor is four and standard errors double.
- At point nine, the factor is ten and standard errors are about three point one six times larger.
- At point nine nine, the factor is one hundred and standard errors are ten times larger.
- Notice how suddenly it explodes as the R squared nears one.
- Common responses are to collect more data, or to combine regressors that carry the same information.
- But dropping a variable has its own danger, which is next.

14 Two ways to get the variable list wrong

- Omit a variable that matters: bias
- Keep a variable that does not: extra variance
- Omission: biased and inconsistent
- Inclusion: unbiased and consistent, less precise
- Bias is usually the worse problem

IN THE LESSON

- Choosing which variables to include can go wrong in two opposite ways, and the damage differs.
- Suppose you leave out a variable that truly affects the outcome.
- If it is correlated with a regressor you kept, the kept coefficient absorbs part of its effect, so it is biased and inconsistent.
- Now suppose you keep a variable that has no true effect.
- The estimates stay unbiased and consistent, because its true coefficient is zero.
- But the estimates become less precise, with larger variances, especially if the useless variable is correlated with the others.
- So omission costs bias and inclusion costs precision.
- Bias is usually the worse problem.

15 The omitted variable bias formula

$E(\text{slope}) = \beta_1 + \beta_2 \times \delta$
$\text{Bias} = \beta_2 \times \delta$

Expected slope	Average slope you would estimate
Own coefficient	True effect of included variable
Omitted coefficient	True effect of omitted variable
Delta	Slope of omitted on included

IN THE LESSON

- There is a formula for the bias.
- Suppose the true model has two variables, but you run a regression with only the first.
- The expected value of your slope equals the first variable's own coefficient, plus the omitted variable's true coefficient times delta.
- Delta is the slope you would get by regressing the omitted variable on the included one.
- So the bias is the omitted effect times that relationship.
- Bias needs two ingredients: the omitted variable must matter, and it must be related to the included variable.
- If it is uncorrelated with the included ones, the slopes stay unbiased, though the error variance is larger.

16 Example 5: bias from an omitted variable

Illustration: true model has both variables; you omit the second

True own effect, β_1	included variable	**0.50**
True omitted effect, β_2	left out	**0.80**
Delta	omitted on included	**0.50**
Expected slope	0.50 + 0.80 × 0.50	**0.90**
Bias	0.80 × 0.50	**+0.40**

IN THE LESSON

- An illustration.
- The true effect of the included variable is point five.
- The omitted variable has a true effect of point eight.
- Regressing the omitted variable on the included one gives a slope, delta, of point five.
- The expected estimate of your slope is point five plus point eight times point five, which is point nine.
- The truth is point five, but on average you would see point nine.
- The bias is point eight times point five, which is plus point four.
- It does not shrink with a bigger sample, so the estimate is inconsistent too.
- If delta were zero, the bias would vanish.
- If the omitted effect were negative, the bias would push the estimate down.

17 Example 6: a useless variable costs precision

Illustration: add a useless regressor correlated 0.60 with a kept one

Squared correlation	0.60^2	**0.36**
One minus that	$1 - 0.36$	**0.64**
Variance multiplier	$1 \div 0.64$	**1.56**
Standard error multiplier	$\sqrt{1.5625}$	**1.25**

IN THE LESSON

- Now the other error.
- Suppose you add a useless variable that happens to be correlated with an included regressor, with a correlation of point six.
- Nothing is biased: the true coefficient of the useless variable is zero, and the included slope stays centered on the truth.
- But the precision falls.
- In the simple case of two regressors, the variance of the included coefficient is multiplied by one divided by one minus the squared correlation.
- The squared correlation is point three six, so one minus it is point six four, and one over point six four is one point five six.
- The standard error is multiplied by the square root of that, one point two five.
- A quarter wider, for nothing.

18 Omission against inclusion

Omission costs bias; inclusion costs precision

Case	What happens
Omit, correlated with kept	Biased and inconsistent
Omit, uncorrelated	Unbiased; error variance up
Include an irrelevant one	Unbiased; variance rises
Irrelevant and correlated	Variance rises even more

IN THE LESSON

- Let's put the cases side by side.
- Omit a relevant variable that is correlated with an included one: the estimates are biased and inconsistent.
- Omit a relevant variable that is uncorrelated with the included ones: the slopes are unbiased, but the error variance is bigger.
- Include an irrelevant variable: the estimates remain unbiased and consistent, with larger variance.
- And if that irrelevant variable is correlated with the other regressors, the variance rises even more.
- The practical lesson is that the costs are asymmetric.
- Extra variance shows up honestly, as a wider interval and a smaller t statistic, but bias hides inside a confident but wrong number.
- That does not mean throwing in every variable, as the next section on model selection shows.

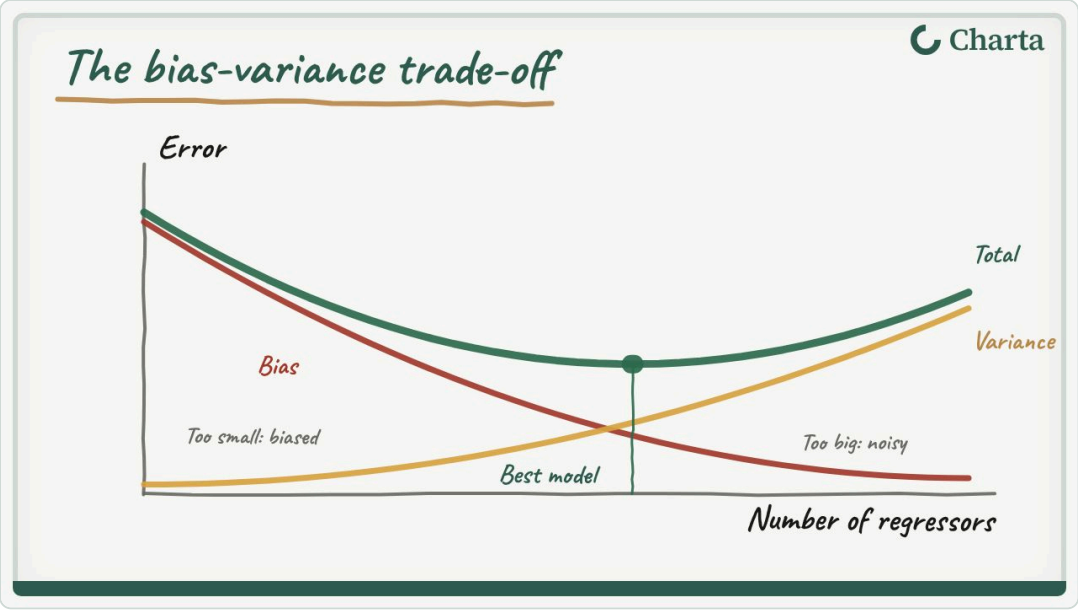
19 Model selection: bias against variance

- Too few regressors: biased, relevant effects omitted
- Too many: low bias, high variance, fits noise
- In-sample R^2 never falls when you add a regressor
- So R^2 cannot choose the model
- Two tools: general-to-specific and cross-validation

IN THE LESSON

- Choosing a model is a balance between bias and variance.
- A model with too few regressors is biased, because it omits relevant effects.
- A model with too many has low bias but high variance: it fits the noise in the sample, and it forecasts poorly out of sample.
- The in sample R squared never falls when a regressor is added, even a useless one, so it always favors the biggest model.
- It cannot choose.
- We look at two procedures.
- General to specific starts big and prunes.
- Cross validation tests how well each model predicts data it has not seen.

20 The bias-variance trade-off



IN THE LESSON

- Here is the trade off in a picture.
- Along the bottom is model size, the number of regressors.
- Up the side is error.
- The red curve is the bias, the error from leaving things out.
- It starts high with a tiny model and falls as relevant variables come in.
- The gold curve is the variance, the error from estimating too many parameters from limited data.
- It starts low and climbs as the model grows.
- The green curve adds them together, and it is shaped like a U.
- The best model sits near the bottom of the U, where the two are balanced.
- A model that is too small is on the left, biased.
- A model that is too big is on the right, noisy, and overfitted.
- Model selection is the search for that bottom.

21 Example 7: general to specific

Illustration: keep regressors with |t| above about 2 (5% level)

Start: five regressors	smallest |t| = 0.3	**drop it**
Re-estimate: four left	smallest |t| = 0.8	**drop it**
Re-estimate: three left	smallest |t| = 1.5	**drop it**
Re-estimate: two left	smallest |t| = 2.4	**stop**

IN THE LESSON

- General to specific starts with a large model that contains every candidate regressor.
- Look at the t statistics.
- Drop the regressor with the smallest absolute t statistic, which is the one with the largest p value.
- Then re estimate, because the other t statistics change when a variable leaves.
- Repeat until every remaining regressor is significant.
- Here we begin with five regressors.
- The weakest has a t statistic of point three, so it goes.
- With four left, the weakest is point eight, so it goes too.
- With three left, the weakest is one point five, still short of about two, so it goes.
- With two left, the weakest is two point four, so we stop.
- Starting big guards against omitted variable bias, but a crowded start can leave the variance high.

22 Example 8: five-fold cross-validation

Illustration: five blocks; mean squared error on each held-out block

Procedure	fit on four, test on one	**5 rounds**
In-sample R squared	B higher: more terms	**B looks better**
Model A, 3 regressors	(4.2+3.8+4.6+4.0+4.4)÷5	**4.2**
Model B, 8 regressors	(5.0+5.4+4.6+5.2+4.8)÷5	**5.0**
Winner	lowest CV error	**Model A**

IN THE LESSON

- Cross validation is usually done in several folds.
- It splits the data into blocks.
- Fit on all blocks but one, measure the squared error on the held out block, and repeat for each of the blocks.
- Then average.
- Here we use five blocks, and compare a small model, A, with three regressors against a large model, B, with eight.
- In sample, model B always has the higher R squared, because it has more terms.
- But look at the held out errors.
- Model A has errors of four point two, three point eight, four point six, four, and four point four.
- Their average is four point two.
- Model B has five, five point four, four point six, five point two, and four point eight, averaging five.
- The lowest cross validated error wins, so model A wins.
- It directly penalizes overfitting.

23 Which residual plot for which problem

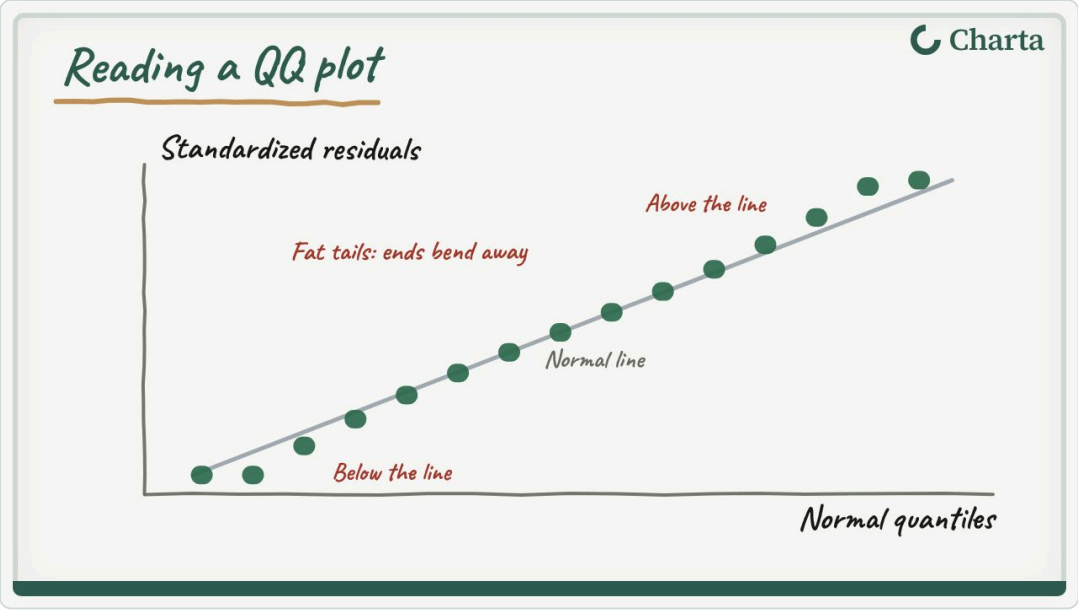
Match plot to assumption; standardize residuals to spot outliers

Plot	What it reveals
Residuals vs fitted	Curve (non-linear), funnel
Residuals vs one regressor	Which variable is to blame
Time order, or ACF plot	Runs of same-sign values
Histogram	Shape, but depends on bins
QQ plot	Tails, with no bin choice

IN THE LESSON

- Residual plots reveal violated assumptions that summary statistics hide, and each plot suits a different problem.
- A scatter of residuals against fitted values shows non linearity, as a curve or U shape, and heteroskedasticity, as a funnel.
- In a multiple regression it summarizes all the regressors in one picture.
- Residuals against each individual regressor locate which variable is the cause.
- Residuals plotted in time order, or their autocorrelation function, show serial correlation as runs of same sign values.
- A histogram gives a quick look at the shape, but it depends on the bin width you choose.
- A normal quantile plot avoids that choice and exposes the tails.
- Standardizing the residuals first makes outliers easier to spot.
- The rule is simple: match the plot to the assumption you are checking.

24 Reading a QQ plot



IN THE LESSON

- A QQ plot lines up the standardized residuals against the quantiles you would expect if they were normal.
- If the residuals really are normal, the points fall on the straight line.
- Here the middle of the data hugs the line, so the bulk looks normal.
- But at both ends, the points bend away from it.
- At the left end they sit below the line, and at the right end above.
- That means the extreme residuals are larger in size than a normal distribution would produce.
- In other words, fat tails.
- It beats a histogram for tail behavior, because there are no bins to choose.

25 Outliers: studentized residual, Cook's D

Studentized = $e \div (s \times \sqrt{(1 - h)})$
Cook's D = $e^2 \div (k\ s^2) \times h \div (1 - h)^2$

Residual	Residual of the observation
Leverage	Leverage, between 1÷n and 1
Residual standard error	Standard error of the regression
Parameters	Parameters incl. intercept
Cook's distance	Above about 1: influential

IN THE LESSON

- An outlier is an observation far from the pattern of the rest, either in the outcome, where it has a large residual, or in the regressors, where it has high leverage.
- Leverage is the diagonal element of the hat matrix.
- It lies between one over the sample size and one, and the leverages add up to the number of estimated parameters, including the intercept, so the average is that number divided by the sample size.
- A studentized residual puts the residuals on a common scale: the residual divided by the residual standard error times the square root of one minus the leverage.
- Cook's distance combines size and leverage: the squared residual over the number of parameters times the residual variance, multiplied by leverage over one minus leverage, squared.
- A value above about one flags a point whose removal would change the fit materially.

26 Example 9: how much leverage is normal?

Illustration: n = 20 observations, k = 2 parameters (intercept and slope)

Sum of the leverages	equals k	**2**
Average leverage	k ÷ n = 2 ÷ 20	**0.10**
Smallest possible	1 ÷ n = 1 ÷ 20	**0.05**
Largest possible	never above one	**1.00**
A leverage of 0.60	0.60 ÷ 0.10	**6 × average**

IN THE LESSON

- Start with leverage itself.
- Suppose we have twenty observations and two estimated parameters, an intercept and a slope.
- The leverages add up to the number of parameters, which is two.
- So the average leverage is two divided by twenty, which is point one.
- The smallest a leverage can be is one over twenty, point zero five, and the largest is one.
- A point with leverage of point six is six times the average, and it pulls strongly on the fitted line.
- High leverage alone is not a fault.
- It simply means the point is far from the center of the regressors.
- Whether it matters depends on Cook's distance, which combines leverage with the residual.

27 Example 10: two suspicious points

Illustration: s = 2, k = 2. A: residual 6, leverage 0.10. B: residual 2, leverage 0.60

A: studentized residual	6 ÷ (2 × √0.9)	**3.16**
A: Cook's distance	4.5 × 0.1 ÷ 0.81	**0.56**
B: studentized residual	2 ÷ (2 × √0.4)	**1.58**
B: Cook's distance	0.5 × 0.6 ÷ 0.16	**1.88**

IN THE LESSON

- Now two suspicious points, with a residual standard error of two.
- Point A has a large residual of six, and average leverage, point one.
- Its studentized residual is six divided by two times the square root of point nine, which is three point one six.
- Its Cook's distance is six squared over two times two squared, which is four point five, multiplied by point one over point nine squared, point one two three.
- The result is point five six, below one.
- Point B has a small residual of only two, but a leverage of point six.
- Its studentized residual is only one point five eight, modest next to the first point.
- Yet its Cook's distance is point five times point six over point four squared, which is one point eight eight, above one.
- The high leverage point with the small residual is the more misleading.

28 When is OLS the best linear unbiased estimator?

- 1 Linear in the parameters (logs and squares are fine)
- 2 Exogeneity: E(error | X) = 0
- 3 No exact linear combination among regressors
- 4 Constant error variance (homoskedasticity)
- 5 Uncorrelated errors (no serial correlation)

IN THE LESSON

- The Gauss Markov theorem says ordinary least squares is the best linear unbiased estimator when five conditions hold.
- One, the model is linear in the parameters, though it may still use logs or squares of the variables.
- Two, the errors have zero mean conditional on the regressors.
- This is called exogeneity.
- It means no regressor is correlated with the error, and it is a different condition from having independent observations.
- Three, no regressor is an exact linear combination of the others, so the coefficients are identified.
- Four, the errors have constant variance.
- Five, the errors are uncorrelated with one another, with no serial correlation.
- Best means the smallest variance among all linear unbiased estimators.
- Know all five conditions and what each one protects.

29 Break a condition, see what fails

Gauss-Markov: which failures cost unbiasedness, and which cost efficiency

Condition broken	Consequence
Exogeneity	Biased: unbiasedness lost
Exact collinearity	No unique solution
Constant variance	Unbiased, not best; bad SEs
No serial correlation	Unbiased, not best; bad SEs
Normality of errors	Not needed for BLUE

IN THE LESSON

- Now connect the five conditions to everything we have met.
- If exogeneity fails, as with an omitted variable correlated with an included one, OLS is biased, so it loses even unbiasedness.
- If a regressor is an exact combination of the others, there is no unique solution at all.
- If the errors are heteroskedastic or autocorrelated, OLS stays unbiased but is no longer best, and the usual standard errors are wrong, which is why robust errors or weighted least squares help.
- Notice what is missing from the list: normality of the errors.
- It is not required for the best linear unbiased result.
- It is needed only for exact small sample t and F tests.

30 Four mistakes to avoid

- Testing R^2 instead of $n \times R^2$ against chi-square
- Calling collinearity bias: it inflates SEs, no bias
- Omitting a relevant variable to cut variance: keep it
- Deleting outliers freely: check, remove only errors

IN THE LESSON

- Mistake one: comparing the R squared of the auxiliary regression with the chi square critical value.
- The statistic is the sample size times that R squared.
- Mistake two: treating multicollinearity as a bias.
- It inflates standard errors and lowers t statistics, but the coefficients stay unbiased.
- Mistake three: leaving out a relevant variable to save variance.
- Omission biases the estimates and more data cannot fix that, while inclusion only costs precision.
- Mistake four: deleting an outlier just because it is awkward.
- First investigate it, because a high leverage point may be telling you something real about the model.
- Remove only data errors.

31 Recap: the main ideas

- Hetero: funnel, n × R^2 test; robust SE, logs or WLS
- Collinearity: SEs × √VIF, no bias; perfect: no fit
- Omit relevant: bias $\beta_2 \times \delta$. Add useless: more variance
- Select by general-to-specific or CV, never in-sample R^2
- Match plot to problem; Cook's D > 1; five GM terms

IN THE LESSON

- Let's pull it together.
- Heteroskedasticity leaves the coefficients unbiased but spoils the standard errors.
- Look for a funnel, then test with the sample size times the auxiliary R squared against a chi square.
- Respond with robust standard errors, a transformation, or weighted least squares.
- Multicollinearity inflates standard errors by the square root of the variance inflation factor, without bias, while perfect collinearity has no solution.
- Omitting a relevant variable biases the estimate by its coefficient times delta, and an irrelevant variable only adds variance.
- Choose models with general to specific or cross validation, never in sample R squared.
- Match each residual plot to its assumption, check leverage and Cook's distance, and know the five Gauss Markov conditions.

32 Try three quick questions

1. n=100, aux R^2=0.09, 3 regressors, critical 7.81. Decision?
   Answer: 100 × 0.09 = 9.0 > 7.81, so reject constant variance.
2. Regressor's R^2 on the others is 0.80. VIF and SE inflation?
   Answer: VIF = 1 ÷ 0.20 = 5. SE inflated by √5 = 2.24.
3. Included slope 0.4, omitted effect 0.5, delta 0.6. Expected slope?
   Answer: 0.4 + 0.5 × 0.6 = 0.7, biased up by 0.3.

IN THE LESSON

- Now try three quick questions yourself.
- Pause after each one.
- Question one.
- A Breusch Pagan test has one hundred observations, an auxiliary R squared of point zero nine, and three auxiliary regressors, with a critical value of seven point eight one.
- What is the decision?
- One hundred times point zero nine is nine, which is greater than seven point eight one, so reject constant variance.
- Question two.
- A regressor has an R squared of point eight on the other regressors.
- What are the variance inflation factor and the standard error inflation?
- One over point two is five, and the square root of five is about two point two four.
- Question three.
- The included slope is point four, the omitted effect is point five, and delta is point six.
- What slope do we expect?
- Point four plus point three, which is point seven.