EC203: Applied Econometrics
Non-normality in the errors: consequences, tests and solutions

Dr. Tom Martin

University of Warwick

Illustrative reading:

- Wooldridge: Chapter 5

Non-normality in the errors

Given the population model:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + \epsilon_i$$

Suppose all of CLRM assumptions hold except for the normality assumption. That is, the assumption $\epsilon | X_1, ..., X_k \sim N(0, \sigma^2)$ does not hold.

What are the consequences for our OLS estimators?

1. OLS estimators are still unbiased.
2. The usual OLS the variance formula is still correct and estimators still have minimum variance.
3. However, we no longer know the shape of the OLS sampling distribution, so we cannot carry out any inference.

Non-normality: consequences

If all CLRM assumptions hold then it is possible to show:

$$b_j \sim N(\beta_j, Var(b_j))$$

Where,

$$Var(b_j) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2(1 - R_j^2)}$$

Replacing σ^2 with its estimator S^2 and normalising we get the result:

$$\frac{b_j - \beta_j}{se(b)} \sim t_{n-k-1}$$

That is, the standardised coefficient has an EXACT t-distribution with $n - k - 1$ degrees of freedom. Further, the F-statistic only has an exact F-distribution if all CLRM assumptions hold.

Non-normality: consequences

Crucially, the normality of the OLS estimators relies on the errors being normally distributed. To give the intuition of why this is the case:

1. Recall, in the SLR case b can be represented in the following fashion:[1]

$$b = \beta + \frac{\sum_{i=1}^{n}(X_i - \bar{X})\epsilon_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

2. Thus, after conditioning on our sample values of X, b is a linear combination of the errors, which under the normality assumption are normally distributed random variables.

3. Any linear combination of normally distributed random variables is itself a normally distributed random variable.

[1] Slide 5 in SLR III lecture.

Non-normality: consequences

Thus, if the errors are not normally distributed the estimators are no longer normally distributed, meaning:

$$b_j \sim ?(\beta_j, Var(b_j))$$

So the major consequence is that t (F) statistic no longer have t (F) distributions.

Given the shape of the sampling distribution is no longer known and we cannot carry out our usual hypothesis test, obtain p-values or construct confidence intervals.

Non-normality: check

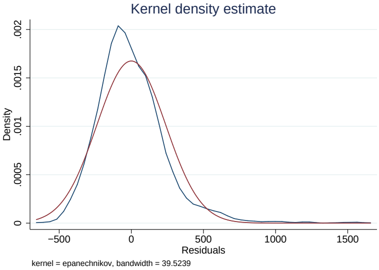
Empirically checking the plot of your residuals. Simply run the regression, save the residuals and plot them.

For example, using our wage.dta type the following commands:

```
reg wage educ age IQ
predict resid, residuals
kdensity resid, normal
```

Non-normality: tests

Test 1:



- There seems to me a mass of errors just below zero and around 500.
- This is not a formal test, but it is an empirically important check.

Non-normality of errors: tests

Test 2: more formally, recall a normally distributed random variable is symmetric about the mean, constant variance, zero skewness (3rd moment) and kurtosis (4th moment). sktest in Stata tests the skewness and kurtosis of a random variable.

For example, in Stata with wage.dta type the following commands:

- reg wage educ age IQ
- predict resid, residuals
- sktest resid

In this case we get a p-value for skewness of 0.000 and one for kurtosis 0.000 which implies our residuals are highly non-normal, as the previous graphic implied. (Recall, our null is normality so the larger the p-value the better: the smaller the p-value the more evidence against the null.)

Non-normality: tests

A note of caution: an otherwise normal distribution is sometimes distorted by outliers in the data.

- It is always important check the distributions of your dependent, independent variables and residuals to check for such outliers.
- Tabulate any discrete variables. Summarise and plot histograms of the continuous variables.
- This is not a non-normality problem it is an outlier problem.

Non-normality: tests

Always check the robustness of your results to possible outliers.
Either by:

- including a dummy variable for any possible outliers in your regression: $out = 1$ if outlier, $out = 0$ otherwise.
- or running the analysis with and without the outliers to see if your results are robust (results do not change drastically) to their presence.

Non-normality of errors: solutions

Solution 1: One common occurrence is that Y has a skewed distribution, such as wages. It is often then the case that taking the log of Y gives a normal distribution. A random variable which has this property is said to have a log-normal distribution.

Solution 2: Rely on asymptotic results and large sample inference. In short, we do nothing. We rely on asymptotic normality instead of normality. That is, we say that b has an asymptotic normal distribution $b \overset{a}{\sim} N(.)$, instead of the usual $b{\sim}N(.)$. We will continue using the latter notation.

Asymptotics: note I

Recall from our discussion of the mean estimator:

- If the population was normal, $X_i \sim N(\mu, \sigma^2)$, the sampling distribution of the mean estimator was normal: $\bar{X} \sim N(\mu, \sigma^2/n)$.
- If the population was non-normal we had to rely on the central limit theorem: the distribution of the mean estimator approached normality as the sample size grew: $\bar{X} \overset{a}{\sim} N(\mu, \sigma^2/n)$.

Asymptotics: note I

The equivalent holds in the OLS estimator case:

- If the errors are normal: the sampling distribution of the OLS estimators are normal. Further, t-statistics have an exact t-distribution: $t = \frac{b_j - \beta_j}{se(b_j)} \sim t_{n-k-1}$.
- If the errors are non-normal: the distribution of the OLS estimator approaches normality as the sample size grows.[2] Further, in large samples the t-statistic converges to the normal, such that $z = \frac{b_j - \beta_j}{se(b_j)} \overset{a}{\sim} Z$.

[2] Are a rule of thumb will we say a 'large sample' is 30 observations.

Asymptotics: note I

The main point is, if the normality assumption fails, the t (F) statistics don't have exact t (F) distributions. The larger the sample, the closer the t (F) statistics will be to having a t (F) distribution. An alternative way of thinking about this is, if the sample size is not very large then the t-distribution (F) can be a very poor approximation to the distribution of the t-statistic (F) if the error is not normally distributed.

Asymptotics: note II

It is common to see results in applied academic papers to refer to t-statistics only.[3] This is because t-tests are a more conservative test, since the t-distribution has fatter tails than the Z-distribution. When the number of observations gets large (30, 50, 100 ish) there is very little difference at all between t-test and z-tests.

[3] Stata also uses t-tests as default.

Asymptotics: note III

It is worth mentioning that a lot of estimator properties and test-statistics rely on asymptotic results: what happens to their distributions as the sample size increases to infinity. The above result is one example.

Typically, the mathematics of such results are more involved so we omit the discussion from this course. Instead we focus on finite sample properties such as unbiasedness.[4] Further, we will continue to use t and F statistics.

[4] Finite sample properties are results that hold in all sample sizes. For an introductory discussion on some of these topics including consistency (the large sample equivalent of unbiasedness), asymptotic normality and Wald, Likelihood Ratio and LM statistics, see Wooldridge Chapter 5. Note: under the CLRM assumptions the OLS estimators are consistent and still have minimum variance.