Heteroskedasticity and Serial Correlation

Christopher Taber

Department of Economics
University of Wisconsin-Madison

April 22, 2011

Outline

In this set of lecture notes we will learn about heteroskedasticity and serial correlation.

They are closely related problems so I will deal with them together

Lets go back to think about the classic regression model.

I think Wooldridge makes this point best in Chapter 10 which is on Time Series

Details of time series is not important but the difference is

- So far we have only thought about random sampling. That is we have assumed that there is a population and we have a random sample of that population.
- In time series we think of each observation coming as a point in time. For example we could have data where an observation is a year (and we have had examples of that kind of data such as when we looked at forecasting or the fish market).

It is not particularly important, but we will index an observation by t (for time) rather than i (for individual).

This notation could refer to either time series or individual data and we do not need to distinguish at these points.

Here are the assumptions:

Assumption TS.1 (Linear in parameters)

The stochastic process

$$\{(x_{t1}, x_{t2}, ..., x_{tk}, y_t) : t = 1, 2, ..., T\}$$

follows the linear model

$$y_t = \beta_0 + \beta_1 x_{t1} + ... + \beta_k x_{tk} + u_t$$

where $\{u_t : t = 1, 2, ..., T\}$ is the sequence of errors or disturbances. Here T is the number of observations (time periods).

Assumption TS.2 (No Perfect Collinearity)

In the sample (and therefore in the underlying time series process), no independent variable is constant nor a perfect linear combination of the others

Assumption TS.3 (Zero Conditional Mean)

For each t, the expected value of the error u_t, given the explanatory variables for all time periods is zero. Mathematically,

$$E(u_t \mid X) = 0, t = 1, 2, ..., T$$

We use these to get unbiasedness (and also consistency)

Theorem 10.1 (Unbiasedness of OLS)

Under Assumption TS.1,TS.2, and TS.3, the OLS estimators are unbiased conditional on X,and therefore unconditionally as well: $E\left(\widehat{\beta}_j\right) = \beta_j, j = 0, 1, ..., k$

We then used the following assumptions to get asymptotic normality and to do inference.

Assumption TS.4 (Homoskedasticity)

Conditional on X the variance of u_t is the same for all t:

$$Var(u_t \mid X) = var(u_t) = \sigma^2$$

for $t = 1, 2, ..., T$

Assumption TS.5 (No Serial Correlation)

Conditional on X, the errors in two different time periods are uncorrelated:

$$cov(u_t, u_s \mid X) = 0$$

for all $t \neq s$.

From these two we get two additional results

Theorem 10.2 (OLS Sampling Variances)

Under the time series Gauss-Markov Assumptions TS.1 through TS.5, the variance of $\widehat{\beta}_j$, conditional on X, is

$$var\left(\hat{\beta}_j \mid X\right) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)}$$

where SST_j is the total some of squares of x_{tj} and R_j^2 is the R-squared from the regression of x_j on the other independent variables.

Theorem 10.4 (Gauss-Markov Theorem)

Under Assumptions TS.1 through TS.5, the OLS estimators are the best linear unbiased estimators conditional on X.

The key thing that I want you to understand is how the different assumptions are important for how we construct the standard errors

To see where it comes from think about the variance of the estimate of the slope coefficient in a simple regression model

$$var\left(\hat{\beta}_1 \mid X\right)$$

Recalling that

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})\, u_t}{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

so that

$$\begin{aligned} var\left(\hat{\beta}_1 \mid X\right) &= var\left(\beta_1 + \frac{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})\, u_t}{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2} \mid X\right) \\ &= \frac{var\left(\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})\, u_t \mid X\right)}{\left[\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2\right]^2} \end{aligned}$$

In General

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}(x_t-\bar{x})\,u_t \mid X\right)=\frac{1}{T^2}Var\left(\sum_{t=1}^{N}(x_t-\bar{x})\,u_t \mid X\right)$$

$$=\frac{1}{T^2}\sum_{t=1}^{N}Var\left((x_t-\bar{x})\,u_t \mid X\right)$$

$$+\frac{1}{T^2}2\sum_{t=1}^{T}\sum_{s=t+1}^{T}Cov\left((x_t-\bar{x})\,u_t,(x_s-\bar{x})\,u_s \mid X\right)$$

In the case we have been dealing with the data is independently distributed (no serial correlation) with $Var(u_t \mid x_t) = \sigma^2$.

The no serial correlation assumption means that

$$Cov\left(\left(x_t - \bar{x}\right) u_t, \left(x_s - \bar{x}\right) u_s \mid X\right) = 0$$

for all of the t and s

This gets rid of a ton of terms

The homoskedasticity assumption means that

$$\frac{1}{T^2}\sum_{t=1}^{T} Var\left(\left(x_t-\bar{x}\right)u_t\right) = \frac{1}{T^2}\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2 Var\left(u_t \mid X\right)$$

$$= \frac{\sigma^2}{T^2}\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2$$

Then

$$
\begin{aligned}
Var\left(\widehat{\beta}_1 \mid X\right) &= \frac{var\left(\frac{1}{T}\sum_{t=1}^{T}(x_t-\bar{x})\,u_t \mid X\right)}{\left[\frac{1}{T}\sum_{t=1}^{T}(x_t-\bar{x})^2\right]^2} \\
&= \frac{\frac{\sigma^2}{T^2}\sum_{t=1}^{T}(x_t-\bar{x})^2}{\left[\frac{1}{T}\sum_{t=1}^{T}(x_t-\bar{x})^2\right]^2} \\
&= \frac{\sigma^2}{\sum_{t=1}^{T}(x_t-\bar{x})^2}
\end{aligned}
$$

We can obtain a consistent estimate of this using

$$\sigma^2 \approx \frac{1}{T-2}\sum_{t=1}^{T}\widehat{u}_t^2$$

where $\widehat{u}_t$ is the residuals from the regression

What happens if no autocorrelation and/or homoskedasticity assumptions are violated?

1. The estimate is still unbiased (and consistent)
2. Our estimate of the standard errors are wrong
3. OLS is no longer BLUE

There are basically two different approaches we can take to deal with this

1. Continue to run OLS since it is consistent, but correct the standard errors to allow for heteroskedasticity or serial correlation (that is deal with 2 but not 3)
2. Run something other than OLS which is BLUE and figure out what the right standard errors are for that (that is deal with both 2 and 3)

If I taught this class 20 years ago I would probably only teach the second approach, however for heteroskedasticity people only tend to use the first

For serial correlation, both are used

Outline

Heteroskedasticity

First lets think about relaxing Heteroskedasticity but not the no autocorrelation assumption.

Everything here pertains to cross section data as well, not just time series.

Suppose that $Var(u_t)$ depends on X_t. However we will still assume that each individual is drawn at random.

Then

$$
\begin{aligned}
Var\left(\frac{1}{T}\sum_{t=1}^{T}(x_t-\bar{x})\,u_t \mid X\right) &= \frac{1}{T^2}\sum_{t=1}^{T} Var\left((x_t-\bar{x})\,u_t \mid X\right) \\
&= \frac{1}{T^2}\sum_{t=1}^{T} E\left((x_t-\bar{x})^2\,u_t^2 \mid X\right) \\
&= \frac{1}{T^2}\sum_{t=1}^{T} (x_t-\bar{x})^2\,E\left(u_t^2 \mid X\right) \\
&\approx \frac{E\left((x_t-\bar{x})^2\,E\left(u_t^2 \mid X\right)\right)}{T}
\end{aligned}
$$

We can just approximate this object as

$$E\left(\left(x_t-\bar{x}\right)^2 E\left(u_t^2 \mid X\right)\right) \approx \frac{1}{T}\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2 \widehat{u}_t^2$$

and use

$$\frac{\frac{1}{T^2}\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2 \widehat{u}_t^2}{\left[\frac{1}{T}\sum_{t=1}^{T}\left(x_t-\bar{x}\right)^2\right]^2}$$

this for standard errors in regression.

Something similar works in general for multiple regression

There is really no reason not to do this

In stata you just say

regress y x1, robust

You can't do F-tests in the simple way we learned before, but stata knows how to do it the more complicated way

In my experience this doesn't really matter much

Here are some examples

This is only the first of two approaches we talked about

This is OLS so the standard errors are right

However it is not BLUE

There are a bunch of different ways to come up with a BLUE estimate

Wooldridge talks about this if you are interested, but I don’t think it is that important so I am not going to worry about it

Outline

Relaxing the Serial Correlation Assumption

Can we do something similar to deal with serial correlation?

Since

$$var\left(\frac{1}{T}\sum_{t=1}^{T}(x_t-\bar{x})\,u_t \mid X\right)=\frac{1}{T^2}\sum_{t=1}^{N}Var\left((x_t-\bar{x})\,u_t \mid X\right)$$

$$+\frac{1}{T^2}2\sum_{t=1}^{T}\sum_{s=t+1}^{T}Cov\left((x_t-\bar{x})\,u_t,(x_s-\bar{x})\,u_s \mid X\right)$$

It sort of seems like we might be able to approximate this well as

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}(x_t - Ex)\,u_t \mid X\right) \approx \frac{1}{T^2}\sum_{t=1}^{T}(x_t - \bar{x})^2\,\widehat{u}_t^2$$
$$+\frac{1}{T^2}2\sum_{t=1}^{T}\sum_{s=t+1}^{T}(x_t - \bar{x})(x_t - \bar{x})\,\widehat{u}_t\widehat{u}_s$$

$$var\left(\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})\,u_t \mid X\right) \approx \frac{1}{T^2}\sum_{t=1}^{N}(x_t - \bar{x})\,\hat{u}_t^2$$
$$+\frac{1}{T^2}2\sum_{t=1}^{T}\sum_{s=t+1}^{T}(x_t - \bar{x})\,\hat{u}_t\,(x_s - \bar{x})\,\hat{u}_s$$

This does not work well at all (both in practice and for technical reasons)

The problem is that while for the first term there are T terms and we are dividing by T^2

For the second there are like T^2 terms and we are dividing by T^2

This turns out to be a problem both in defining the actual covariance and in approximating it

Essentially if

$$\frac{1}{T}2\sum_{t=1}^{T}\sum_{s=t+1}^{T} Cov\left(\left(x_t-\bar{x}\right)u_t,\left(x_s-\bar{x}\right)u_s \mid X\right)$$

blows up you have real problems

Relatedly, the estimator I suggested above will not settle down in the data

There are two different approaches to fix the problem.

AR Models

The first solution to this type of problem is to construct a model for the error terms

We can then estimate the parameters of the model and figure out the standard errors

The most common model for the error terms is called an AR(1)

Here we suppose that

$$u_t = \rho u_{t-1} + \varepsilon_t$$

where ε_t is iid (or white noise) with $E(\varepsilon_t) = 0$, and

$$-1 < \rho < 1.$$

Lets think about the properties of the AR(1)

Since ε_t is iid, u_t will be correlated with current and lagged values of ε_t, but not future values.

If the time series has been going on forever

$$\begin{aligned}
u_t &= \rho u_{t-1} + \varepsilon_t \\
&= \rho^2 u_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t \\
&= \rho^K u_{t-K} + \rho^{K-1}\varepsilon_{t-(K-1)} + \ldots + \varepsilon_t \\
&= \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}
\end{aligned}$$

But then

$$
\begin{aligned}
E(u_t) &= E\left(\sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}\right) \\
&= \sum_{j=0}^{\infty} \rho^j E(\varepsilon_{t-j}) \\
&= 0
\end{aligned}
$$

and

$$
\begin{aligned}
var(u_t) &= var\left(\sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}\right) \\
&= \sum_{j=0}^{\infty} \rho^{2j} var(\varepsilon_{t-j}) \\
&= var(\varepsilon_t) \sum_{j=0}^{\infty} \rho^{2j} \\
&= \frac{Var(\varepsilon_t)}{1-\rho^2}
\end{aligned}
$$

Under these conditions the model is “covariance stationary” because

$$\begin{aligned} Cov(u_t, u_{t+1}) &= Cov(u_t, \rho u_t + \varepsilon_{t+1}) \\ &= \rho Var(u_t) \end{aligned}$$

$$\begin{aligned} Cov(u_t, u_{t+2}) &= Cov(u_t, \rho u_{t+1} + \varepsilon_{t+2}) \\ &= Cov(u_t, \rho (\rho u_t + \varepsilon_{t+1}) + \varepsilon_{t+2}) \\ &= \rho^2 Var(u_t) \end{aligned}$$

More generally

$$Cov(u_t, u_{t+h}) = \rho^h Var(u_t)$$

It turns out that this solves the problem of the Variance blowing up described above (I will spare you the algebra, but its straightforward to show this)

This model is called $AR(1)$ for a simple reason:

there is 1 autoregressive term

It easily generalizes to an $AR(2)$,

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$$

with ε_t iid

Or even more generally an $AR(p)$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_p u_{t-p} + \varepsilon_t$$

MA Models

The other really common representation is what is called a moving average or *MA* process

In this case we can write

$$u_t = \varepsilon_t + \alpha\varepsilon_{t-1}$$

where ε_t is iid.

If $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = \sigma_\varepsilon^2$

$$\begin{aligned} E(u_t) &= E(\varepsilon_t) + \alpha E(\varepsilon_{t-1}) \\ &= 0 \end{aligned}$$

$$\begin{aligned} Var(u_t) &= Var(\varepsilon_t) + \alpha^2 Var(\varepsilon_{t-1}) \\ &= \left(1 + \alpha^2\right) \sigma_\varepsilon^2 \end{aligned}$$

$$\begin{aligned} Cov(u_t, u_{t+1}) &= Cov(\varepsilon_t + \alpha\varepsilon_{t-1}, \varepsilon_{t+1} + \alpha\varepsilon_t) \\ &= \alpha\sigma_\varepsilon^2 \end{aligned}$$

$$\begin{aligned} Cov(u_t, u_{t+2}) &= Cov(\varepsilon_t + \alpha\varepsilon_{t-1}, \varepsilon_{t+2} + \alpha\varepsilon_{t+1}) \\ &= 0 \end{aligned}$$

Thus the MA(1) is covariance stationary

This can be generalized to an $MA(2)$

$$u_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2}$$

and further to an $MA(q)$

$$u_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + ... + \alpha_q \varepsilon_{t-q}$$

The *MA* and *AR* specifications are not mutually exclusive

You can stick them together

An $ARMA(p, q)$ is written as

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_p u_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + ... + \alpha_q \varepsilon_{t-q}$$

with ε_t iid.

Using ARMA Models

In practice what do we do with this?

One possibility is to just run OLS and correct the model for the fact that the error terms are correlated

It turns out that there is a better (more efficient) way to estimate the parameters

It is called GLS and is discussed in Wooldridge

Stata does something quite similar to this, but not quite

It performs maximum likelihood which is similar to GLS but assumes that the error terms are normally distributed

Stata is not the best package for time series data, but will work

Lets look at some examples

Newey West Standard Errors

There is another approach one can take

Rather than trying to model the dependence, we can try to estimate the variance of $\widehat{\beta}$ directly

Lets go back to thinking about estimating

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}(x_t - Ex)\,\varepsilon_t \mid X\right) = \frac{1}{T^2}\sum_{t=1}^{T} Var\left((x_t - Ex)\,\varepsilon_t \mid X\right)$$

$$+\frac{1}{T^2}2\sum_{t=1}^{T}\sum_{s=t+1}^{T} Cov\left((x_t - Ex)\,\varepsilon_t, (x_s - Ex)\,\varepsilon_s \mid X\right)$$

directly

There are 2 problems

- There are too many terms as a result of the double sum which will mess things up (as I have said before)
- A practical problem is that the approximated terms might not match well together. In that case we might not be able to get reasonable standard errors

There turns out to be a fairly simple solution

- Don't use so many terms
- Weight in such a way that it works OK

Newey and West show that for some L you can approximate

$$Var(\hat{\beta}_1) = \frac{1}{T} \sum_{\ell=1}^{L} \sum_{t=\ell+1}^{T} W_\ell u_t u_{t-\ell} (\tilde{x}_t \tilde{x}_{t-\ell} - \tilde{x}_{t-\ell} \tilde{x}_t)$$

where

$$W_\ell = \frac{\ell}{L+1}$$

$$\tilde{x}_t = x_t - \bar{x}$$

The question is how do you pick L

This is pretty arbitrary

If your sample was really big, you would pick L to be really big

In Stata you just say

newey Y X, lag(L)

If you put in lag(0) this is equivalent to using Heteroskedasticity robust standard errors

Outline

Panel Data

Another type of data is panel data

- It is like cross-sectional data in that we assume there is a population and that we randomly sample people from the population
- However, we assume that we have more than one observation per individual

Here is an example

It is useful to use two subscripts so that Y_{it} means the outcome for individual i at time t.

We can write the model as

$$Y_{it} = \beta_0 + \beta_1 X_{it1} + ... + \beta_k X_{itk} + u_{it}$$

We can run a regression with all of the data as usual and under the standard assumption that $E(u_{it} \mid X_{it}) = 0$ we can obtain unbiased and consistent estimates.

The question is do we believe Assumption TS.5 which in this case would take the form:

Conditional on X, the errors for two different observations are uncorrelated:

$$Corr(u_{it}, u_{j\tau}) = 0$$

for all $(i, t) \neq (j, \tau)$.

There are really three different cases

- $i = j, t \neq \tau$:
$$Corr(u_{it}, u_{i\tau})$$
- $i \neq j, t = \tau$
$$Corr(u_{it}, u_{jt})$$
- $i \neq j, t \neq \tau$
$$Corr(u_{it}, u_{j\tau})$$

Invoking the assumption in the first case seems nuts as we probably think:
$$Corr(u_{it}, u_{i\tau}) > 0$$

The other two don't seem unreasonable

As in all the other cases there are two things to do:

- Run OLS and correct the standard errors (like robust and newey)
- Write down a model and do things more efficiently

The Cluster Command

There turns out to be a really nice way to do the first thing.

To use the "cluster" command in stata we need to assume that

$$Corr(u_{it}, u_{j\tau}) = 0$$

whenever $j \neq i$ for any t and τ

We don't need to make any assumption about $Corr(u_{it}, u_{i\tau})$

We also don't need to make any assumption about $var(u_{it})$ either so it is "heteroskeasticity robust" as well

The key to doing this in STATA is that you need some variable that uniquely identifies people (such as idcode)

You then would say:

```
regress y x, cluster(idcode)
```

Here are some examples.

Random Effect

The other common model people use is a "random effect" model.

We model the error term as

$$u_{it} = \theta_i + \varepsilon_{it}$$

Where

$$\begin{aligned} cov(\varepsilon_{it}, \varepsilon_{i\tau}) &= 0 \\ cov(\theta_i, \varepsilon_{i\tau}) &= 0 \end{aligned}$$

This means that

$$\begin{aligned} cov(u_{it}, u_{i\tau}) &= cov(\theta_i + \varepsilon_{it}, \theta_i + \varepsilon_{i\tau}) \\ &= cov(\theta_i, \theta_i) + cov(\theta_i, \varepsilon_{i\tau}) + cov(\varepsilon_{it}, \theta_i) + cov(\varepsilon_{it}, \varepsilon_{i\tau}) \\ &= var(\theta_i) \end{aligned}$$

We estimate this model using Generalized Least Squares which is more efficient than OLS.

In stata just say

```
xtreg y x, re i(idcode)
```

Lets see some examples

Fixed Effects

However there is something even cooler about this

Note that we can write the single regressor version of the model as

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \theta_i + \varepsilon_{it}$$

Suppose we only have two periods of data $t = 1, 0$ then notice that we can write

$$\begin{aligned}
\Delta Y_i =& Y_{i1} - Y_{i0} \\
=& \beta_0 + \beta_1 X_{i1} + \theta_i + \varepsilon_{i1} \\
& - (\beta_0 + \beta_1 X_{it0} + \theta_i + \varepsilon_{i0}) \\
=& \beta_1 \Delta X_i + \Delta \varepsilon_i
\end{aligned}$$

We can estimate this model by regression ΔY_i on ΔX_i

The really nice thing about this is that we didn't need to assume anything about the relationship between X and θ.

Here is a couple examples

More than 2 time periods

What do we do when we have more than 2 time periods?

We could still construct ΔY.That is if we had three periods we could construct $Y_{i1} - Y_{i0}$ and $Y_{i2} - Y_{i1}$.

It turns out that there is something that is often better.

Note that

$$\begin{aligned}
\bar{Y}_i &= \frac{1}{T}\sum_{t=1}^{T} Y_{it} \\
&= \frac{1}{T}\sum_{t=1}^{T} \left[\beta_0 + \beta_1 X_{it} + \theta_i + \varepsilon_{it}\right] \\
&= \beta_0 + \beta_1 \bar{X}_i + \theta_i + \bar{\varepsilon}_i
\end{aligned}$$

Then

$$
\begin{aligned}
Y_{it} - \bar{Y}_i =& \beta_0 + \beta_1 X_{it} + \theta_i + \varepsilon_{it} \\
& - \left(\beta_0 + \beta_1 \bar{X}_i + \theta_i + \bar{\varepsilon}_i\right) \\
=& \beta_1 \left(X_{it} - \bar{X}_i\right) + \varepsilon_{it} - \bar{\varepsilon}_i
\end{aligned}
$$

To estimate this model we just regress $Y_{it} - \bar{Y}_i$ on $X_{it} - \bar{X}_i$

This is what is typically referred to as “fixed effects”

In STATA

```
xtreg y x, fe i(idcode)
```

Lets see some examples

Thats all I have to say