Econometrics

Lecture 12

Nathaniel Higgins

ERS and JHU

- Random effects are the principal alternative to fixed effects
- We usually write a RE model *exactly the same way* that we write the FE model
- Recall the FE union example:

$$lnwage_{it} = \beta_0 + \alpha_i + Union_{it}\beta_1 + X's_{it}\beta_{2:k} + u_{it}$$

- Random effects are the principal alternative to fixed effects
- We usually write a RE model *exactly the same way* that we write the FE model
- Recall the FE union example:

$$lnwage_{it} = \beta_0 + \alpha_i + Union_{it}\beta_1 + X's_{it}\beta_{2:k} + u_{it}$$

- We write the RE model the exact same way, but we interpret the α_i's differently
- Now, instead of thinking of them as individual intercepts, we think of them as *part of the error term*

- Random effects are the principal alternative to fixed effects
- We usually write a RE model *exactly the same way* that we write the FE model
- Recall the FE union example:

$$lnwage_{it} = \beta_0 + \alpha_i + Union_{it}\beta_1 + X's_{it}\beta_{2:k} + u_{it}$$

- We write the RE model the exact same way, but we interpret the α_i's differently
- Now, instead of thinking of them as individual intercepts, we think of them as *part of the error term*
- Let's rewrite the model above so that the (subtle) difference is emphasized

$$lnwage_{it} = \beta_0 + Union_{it}\beta_1 + X's_{it}\beta_{2:k} + (\alpha_i + u_{it})$$

- From previous slide:

$$lnwage_{it} = \beta_0 + Union_{it}\beta_1 + X's_{it}\beta_{2:k} + (\alpha_i + u_{it})$$

- Note that the RE model is usually written a little differently
- You might see something like:

$$lnwage_{it} = \beta_0 + Union_{it}\beta_1 + X's_{it}\beta_{2:k} + \eta_{it}$$

where $\eta_{it} = \alpha_i + u_{it}$. This is just a cleaner way to write things.

- From previous slide:

$$lnwage_{it} = \beta_0 + Union_{it}\beta_1 + X's_{it}\beta_{2:k} + (\alpha_i + u_{it})$$

- Note that the RE model is usually written a little differently
- You might see something like:

$$lnwage_{it} = \beta_0 + Union_{it}\beta_1 + X's_{it}\beta_{2:k} + \eta_{it}$$

  where $\eta_{it} = \alpha_i + u_{it}$. This is just a cleaner way to write things.
- The important thing is to realize that the error term (i.e. all the unobservables) is now a composite error term made up of two parts:
  1. A time-invariant and individual-specific error term (what I call the α_i's)
  2. A pure random error (what I call the u_{it}'s)

- Now what? What's different? How does this change in the way that we view the α_i's affect the way we estimate the model?

- Now what? What's different? How does this change in the way that we view the α_i's affect the way we estimate the model?
- Recall that when we run a FE model, we interpret the α_i's as an individual-specific intercept
- This means that we insert an individual-specific dummy into the dset for each individual
- (or we transform the data in a way that is equivalent)

- With RE, if the model is:

$$lnwage_{it} = \beta_0 + Union_{it}\beta_1 + X's_{it}\beta_{2:k} + \eta_{it}$$

where $\eta_{it} = \alpha_i + u_{it}$.
- We do not insert individual-specific intercepts into the data
- Instead, we use the individual-specific "effects" to change the way we treat the error term

- So where are we?
- We have a model that looks pretty much like a standard linear model
- We're not going to alter the data by adding individual intercepts
- But we have a model where the error term is . . . *funny*
- So we are left with:
  - A pretty standard linear model
  - with a funky error term η

Take a step back. Let's talk about the error term.

- Errors (or unobservables) are the things that influence the outcome y but that aren't in our dset
- We don't observe them (of course) but we do assume we know something about them
- What we assume we know about them is important for our estimation strategy
- We haven't talked too much about the assumptions we usually make, but the assumptions we make do implicitly influence everything we've done so far

- Let's talk separately about two different types of unobservables
- We'll call the "regular" unobservables u
- We'll call the "funky" unobservables η

Regular unobservable term	Funky unobservable term
u_{it}	$\eta_{it} = \alpha_i + u_{it}$

Unobservables

The assumptions we have been making (even if you forgot):

1. $$E(u_{it}) = 0$$

   - Translation: we expect the unobservable to have a value of zero, on average

2. $$cov(u_{it}, u_{is}) = 0, t \neq s$$
   $$cov(u_{it}, u_{jt}) = 0, i \neq j$$

   - Translation: we don't expect any correlation between unobservable terms either over time, or between individuals

3. $$var(u_{it}) = \sigma_u^2 \text{ for all } i, t$$

   - Translation: the unobservable term has constant variability

- Set aside assumption #1 — let's say that this assumption is always satisfied (with both u and η)
- Let's look carefully at assumptions #2 and #3

- The definition of variance of a random variable is covariance with itself. That is:

$$v(u_{it}) = cov(u_{it}, u_{it}).$$

- Using this, we can write assumptions 2 and 3 this way:

$$cov(u_{it}, u_{is}) = 0, t \neq s \tag{2a}$$
$$cov(u_{it}, u_{jt}) = 0, i \neq j \tag{2b}$$
$$cov(u_{it}, u_{it}) = \sigma_u^2 \tag{3}$$

- Remember what these mean

Unobservables

The regular assumptions, re-written

- $$cov(u_{it}, u_{is}) = 0, t \neq s \quad (2a)$$
  $$cov(u_{it}, u_{jt}) = 0, i \neq j \quad (2b)$$
  $$cov(u_{it}, u_{it}) = \sigma_u^2 \quad (3)$$

- $$cov(u_{it}, u_{is}) = 0, t \neq s \tag{2a}$$
  $$cov(u_{it}, u_{jt}) = 0, i \neq j \tag{2b}$$
  $$cov(u_{it}, u_{it}) = \sigma_u^2 \tag{3}$$
- (3) means that we usually expect the unobservables to have a single common variance

Unobservables
The regular assumptions, re-written

- $$cov(u_{it}, u_{is}) = 0, t \neq s \tag{2a}$$
$$cov(u_{it}, u_{jt}) = 0, i \neq j \tag{2b}$$
$$cov(u_{it}, u_{it}) = \sigma_u^2 \tag{3}$$
- (3) means that we usually expect the unobservables to have a single common variance
- (2a) means that we don't expect any covariance (any correlation) between unobservable determinants of y for the same individual in different time periods

- $$cov(u_{it}, u_{is}) = 0, t \neq s \tag{2a}$$
  $$cov(u_{it}, u_{jt}) = 0, i \neq j \tag{2b}$$
  $$cov(u_{it}, u_{it}) = \sigma_u^2 \tag{3}$$
- (3) means that we usually expect the unobservables to have a single common variance
- (2a) means that we don't expect any covariance (any correlation) between unobservable determinants of y for the same individual in different time periods
- (2b) means that we don't expect any covariance (any correlation) between unobservable determinants of y for different individuals

- Think about what this means . . .
- The errors are just random “noise”
- If errors are nothing more than noise, then errors are unpredictable
- If errors are unpredictable, there is nothing about the errors that suggests we can make our inference any better

- Think about what this means . . .
- The errors are just random “noise”
- If errors are nothing more than noise, then errors are unpredictable
- If errors are unpredictable, there is nothing about the errors that suggests we can make our inference any better
- (Does this sound like our situation in the RE model?)

- Look again at the (basic) RE model

$$y_{it} = \beta_0 + x_{it}\beta_1 + \eta_{it}$$
$$\eta_{it} = \alpha_i + u_{it}$$

- Look again at the (basic) RE model

$$y_{it} = \beta_0 + x_{it}\beta_1 + \eta_{it}$$
$$\eta_{it} = \alpha_i + u_{it}$$

- Consider η_{it} and η_{is}. If t and s notation drive you up a wall, consider $\eta_{i,1995}$ and $\eta_{i,1997}$. Does it look like they have a common element?

- Look again at the (basic) RE model

$$y_{it} = \beta_0 + x_{it}\beta_1 + \eta_{it}$$
$$\eta_{it} = \alpha_i + u_{it}$$

- Consider η_{it} and η_{is}. If t and s notation drive you up a wall, consider $\eta_{i,1995}$ and $\eta_{i,1997}$. Does it look like they have a common element? Yes

- Look again at the (basic) RE model

$$y_{it} = \beta_0 + x_{it}\beta_1 + \eta_{it}$$
$$\eta_{it} = \alpha_i + u_{it}$$

- Consider η_{it} and η_{is}. If t and s notation drive you up a wall, consider $\eta_{i,1995}$ and $\eta_{i,1997}$. Does it look like they have a common element? Yes
- It looks like $cov(\eta_{it}, \eta_{is}) \neq 0$
- It looks like there is a relationship between the unobservables produced by a given individual i in two different time periods

- What does it mean if we assume that $cov(\eta_{i1}, \eta_{i2}) \neq 0$?
- If it's hard to think about $cov(\eta_{i1}, \eta_{i2})$, give the i a name and the t a value:

$$cov(\eta_{John,1998}, \eta_{John,1999})$$

- How can we use this information?
- Develop a story supposing that $cov(\eta_{i1}, \eta_{i2}) > 0$

- When we say that $cov(u_{i1}, u_{i2}) > 0$, we are saying that we assume there is positive covariance between the first two error terms (the first two unobservables)

- When we say that $cov(u_{i1}, u_{i2}) > 0$, we are saying that we assume there is positive covariance between the first two error terms (the first two unobservables)
- This suggests that when we are estimating our model, if we get a positive error for the first observation (the observed value of y is greater than the predicted value of y for the first observation), then we are *more likely* to get a positive observation for the second observation

- When we say that $cov(u_{i1}, u_{i2}) > 0$, we are saying that we assume there is positive covariance between the first two error terms (the first two unobservables)
- This suggests that when we are estimating our model, if we get a positive error for the first observation (the observed value of y is greater than the predicted value of y for the first observation), then we are *more likely* to get a positive observation for the second observation
- If we know that we are likely to get a positive error for the second observation, what does this suggest about our prediction for the second observation?

- It suggests that our model is predicting a value for the second observation that is *too low*
- If we know this to begin with, then we should make our prediction for the second observation *lower*
- How do we do this?

- It suggests that our model is predicting a value for the second observation that is *too low*
- If we know this to begin with, then we should make our prediction for the second observation *lower*
- How do we do this?
- As long as we have $cov(\eta_{it}, \eta_{is}) \neq 0$, we'll know that we could improve our estimates by taking this into account

- It suggests that our model is predicting a value for the second observation that is *too low*
- If we know this to begin with, then we should make our prediction for the second observation *lower*
- How do we do this?
- As long as we have $cov(\eta_{it}, \eta_{is}) \neq 0$, we'll know that we could improve our estimates by taking this into account
- We need to take our model with $cov(\eta_{it}, \eta_{is}) \neq 0$ and transform it into a model with new errors $\eta*$ such that $cov(\eta*_{it}, \eta*_{is}) = 0$

- It suggests that our model is predicting a value for the second observation that is *too low*
- If we know this to begin with, then we should make our prediction for the second observation *lower*
- How do we do this?
- As long as we have $cov(\eta_{it}, \eta_{is}) \neq 0$, we'll know that we could improve our estimates by taking this into account
- We need to take our model with $cov(\eta_{it}, \eta_{is}) \neq 0$ and transform it into a model with new errors $\eta*$ such that $cov(\eta*_{it}, \eta*_{is}) = 0$
- We do this with a procedure called Generalized Least Squares (GLS)

GLS

- Just like OLS on the FE model gives us FE estimates, GLS on the RE model gives us RE estimates

GLS

- Just like OLS on the FE model gives us FE estimates, GLS on the RE model gives us RE estimates
- GLS is similar to OLS, but with a twist
- We use our knowledge about the relationship between the errors to improve our inference
- We improve our inference by transforming the data until the errors are “regular”

GLS

- Just like OLS on the FE model gives us FE estimates, GLS on the RE model gives us RE estimates
- GLS is similar to OLS, but with a twist
- We use our knowledge about the relationship between the errors to improve our inference
- We improve our inference by transforming the data until the errors are “regular”
- Let’s compare the standard model to the funky-error model to see if we can figure out how to transform the data

GLS

- Our "regular" model is characterized by

$$y_{it} = \beta_0 + x_{it}\beta_1 + u_{it} \quad \text{(reg)}$$
$$E(u_{it}) = 0 \quad (1)$$
$$cov(u_{it}, u_{is}) = 0, t \neq s \quad \text{(2a)}$$
$$cov(u_{it}, u_{jt}) = 0, i \neq j \quad \text{(2b)}$$
$$cov(u_{it}, u_{it}) = \sigma_u^2 \quad (3)$$

- Our funky-error model is characterized by

$$y_{it} = \beta_0 + x_{it}\beta_1 + \eta_{it} \quad \text{(funk)}$$
$$E(\eta_{it}) = 0 \quad (1)$$
$$cov(\eta_{it}, \eta_{is}) = \sigma_\alpha^2, t \neq s \quad \text{(2a')}$$
$$cov(\eta_{it}, \eta_{jt}) = 0, i \neq j \quad \text{(2b)}$$
$$cov(\eta_{it}, \eta_{it}) = \sigma_\alpha^2 + \sigma_u^2 \quad (3)$$

GLS with RE model

- In this case, the form of the correlation is based on our assumption that there is a constant bit of the unobservable for each individual (the α_i)
- With this form, *correlation*$(\eta_{it}, \eta_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_u^2}, t \neq s$
- Let's call the following awful thing γ:

$$\gamma = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + T\sigma_u^2}$$

where T is the number of time periods

GLS with RE model

- We transform every piece of data in the model by subtracting the within-individual mean and adding the within-individual mean times γ

$$(y_{it} - \overline{y_i} + \gamma \overline{y_i}) = \beta_0 \gamma + \beta_1 (x_{it} - \overline{x_i} + \overline{x_i}\gamma) + (\eta_{it} - \overline{\eta_i} + \overline{\eta_i}\gamma)$$

- This model's error term has the "regular" properties
- Re-write for a cleaner look:

GLS with RE model

- We transform every piece of data in the model by subtracting the within-individual mean and adding the within-individual mean times γ

$$(y_{it} - \overline{y_i} + \gamma\overline{y_i}) = \beta_0\gamma + \beta_1(x_{it} - \overline{x_i} + \overline{x_i}\gamma) + (\eta_{it} - \overline{\eta_i} + \overline{\eta_i}\gamma)$$

- This model's error term has the "regular" properties
- Re-write for a cleaner look:

$$y*_{it} = \beta_0 + \beta_1 x *_{it} + \eta*_{it}$$

where $\eta*_{it}$ has all the regular properties (non-funky properties)
- OLS run on this model is GLS! OLS run on this model is known as the Random Effects (RE) model. If we do this, we now have an *efficient* estimator — an estimator that incorporates all the information we have about the data into the estimation procedure

GLS

- Complicated-alert! All you need to understand is this:
  1. When we assume that part of the unobservables is constant, we can derive a formula for the covariance between observations
  2. When we can derive a formula, we can use this to transform the data
- This all seems pretty good
- We can perform GLS by transforming the data
- We transform the data using simple means for each individual (we can calculate those) and γ

GLS

- Complicated-alert! All you need to understand is this:
  1. When we assume that part of the unobservables is constant, we can derive a formula for the covariance between observations
  2. When we can derive a formula, we can use this to transform the data
- This all seems pretty good
- We can perform GLS by transforming the data
- We transform the data using simple means for each individual (we can calculate those) and γ
- But where the heck do we get γ?

GLS

- We can't conjure γ out of thin air
- So we need some way to estimate γ
- If we have an estimate of γ, then we can estimate the model:

$$y*_{it} = \beta_0 + \beta_1 x *_{it} + \eta*_{it}$$

- Here's how we get the two estimates that we need to finish the RE model:
  1. Estimate a model that gets rid of all the individual-specific effects. This gives us errors that only consist of u's
  2. Estimate a full model, which gives us errors that have a variance $\sigma_\alpha^2 + \sigma_u^2$
  3. Subtract from this estimate the first estimate to get an estimate of σ_α^2

Good news!

- You don't have to do any of this nonsense
- You do:

Stata code

```
xtreg lwage union, re
```

Huh?

- You: So why did you make me listen to all that, jerk?

Huh?

- You: So why did you make me listen to all that, jerk?
- It's not because I'm a sadist
- There are some serious differences between FE and RE that are difficult to appreciate if you don't know *how* each of the models are calculated
- Even if you don't have to do the calculations, understanding how the calculations are done gives you lots of intuition for the difference between the models

RE and FE

- Both FE and RE assume that differences between individuals (individual-specific heterogeneity) are unobservable
- Both FE and RE assume that the differences between individuals can be incorporated into our estimation

RE and FE

- Both FE and RE assume that differences between individuals (individual-specific heterogeneity) are unobservable
- Both FE and RE assume that the differences between individuals can be incorporated into our estimation
- RE assumes that differences between individuals in a panel dset are part of the error term
  - RE assumes that $E(\eta_{it}) = E(\alpha_i + u_{it}) = 0$
  - We don't need to estimate the α_i's
  - We only need to account for their impact on the correlation between predictions

RE and FE

- Both FE and RE assume that differences between individuals (individual-specific heterogeneity) are unobservable
- Both FE and RE assume that the differences between individuals can be incorporated into our estimation
- RE assumes that differences between individuals in a panel dset are part of the error term
  - RE assumes that $E(\eta_{it}) = E(\alpha_i + u_{it}) = 0$
  - We don't need to estimate the α_i's
  - We only need to account for their impact on the correlation between predictions
- FE assumes that differences between individuals in a panel dset are "fixed" and can be controlled for by *removing their effect* from the data

RE and FE

- What is the import of these differences?

RE and FE

- What is the import of these differences?
- FE works great if what you want to do is remove ALL time-invariant heterogeneity between individuals
- FE allows us to estimate a time-varying effect consistently under a pretty wide range of assumptions

RE and FE

- What is the import of these differences?
- FE works great if what you want to do is remove ALL time-invariant heterogeneity between individuals
- FE allows us to estimate a time-varying effect consistently under a pretty wide range of assumptions
- BUT! FE throws away lots of variation
- Remember the FE John/Jane example: John might as well have not been in the dset
- Any difference between John and Jane was totally ignored

RE and FE

- What is the import of these differences?
- FE works great if what you want to do is remove ALL time-invariant heterogeneity between individuals
- FE allows us to estimate a time-varying effect consistently under a pretty wide range of assumptions
- BUT! FE throws away lots of variation
- Remember the FE John/Jane example: John might as well have not been in the dset
- Any difference between John and Jane was totally ignored
- RE uses this variation!
- RE can estimate the impact of time-invariant differences between individuals
- BUT! RE is not necessarily unbiased if the unobservables are correlated with x (because RE leaves these unobservables in the data rather than removing them)
- That is, RE is subject to endogeneity concerns

RE and FE

- An analogy that might be useful:
  - Fixed effects is a hatchet
  - Random effects is a paring knife
- If the unobservable individual-specific heterogeneity is causing bias (endogeneity), we want to remove it
- We remove it by controlling for it
- We control for it by including it specifically in the regression: α_i

RE and FE

- An analogy that might be useful:
  - Fixed effects is a hatchet
  - Random effects is a paring knife
- If the unobservable individual-specific heterogeneity is causing bias (endogeneity), we want to remove it
- We remove it by controlling for it
- We control for it by including it specifically in the regression: α_i
- If the unobservable individual-specific heterogeneity is not causing bias, then we want to account for it without removing all the variation from the regression
- We do this by accounting for "who" generates each observation, transforming the data so that observations generated by the same *i* (person, country, etc.) are more similar

RE and FE

- How to decide?
- Well, it’s complicated. But the short answer is that the Hausman test is often used.
- So if you find yourself in a situation such that you are comparing the two models and you look back at your notes, you’ll have to look up the details of the Hausman test
- The Stata command is . . . you guessed it: `hausman`

- Discussion of RE is a good primer for discussing heteroskedasticity
- . . . as long as you remember what heteroskedasticity is

- Discussion of RE is a good primer for discussing heteroskedasticity
- . . . as long as you remember what heteroskedasticity is
- Heteroskedasticity is when the variance of our unobservables are not constant

- Discussion of RE is a good primer for discussing heteroskedasticity
- ...as long as you remember what heteroskedasticity is
- Heteroskedasticity is when the variance of our unobservables are not constant
- There skedasticity brothers:
  1. Homo:
  $$Var(u_i) = \sigma_u^2$$
  2. Hetero:
  $$Var(u_i) = \sigma_{u_i}^2$$
  (variance of u is different for different i)

- If variance is not constant, we've got problems
- What problems?

- If variance is not constant, we've got problems
- What problems?
- No problems with bias — heteroskedasticity has nothing to do with the coefficient estimates

- If variance is not constant, we've got problems
- What problems?
- No problems with bias — heteroskedasticity has nothing to do with the coefficient estimates
- Where does the variance of the unobservables show up?

- If variance is not constant, we've got problems
- What problems?
- No problems with bias — heteroskedasticity has nothing to do with the coefficient estimates
- Where does the variance of the unobservables show up?
- In the calculation of our standard errors
- If the s.e.'s are no good, the our t-stats and our inference is no good (we can't trust it)

- If variance is not constant, we've got problems
- What problems?
- No problems with bias — heteroskedasticity has nothing to do with the coefficient estimates
- Where does the variance of the unobservables show up?
- In the calculation of our standard errors
- If the s.e.'s are no good, the our t-stats and our inference is no good (we can't trust it)
- So we need to "fix" heteroskedasticity

- There are two common fixes for HDicity
- Based on what we just did with RE, any ideas on one way to fix HDicity?

- There are two common fixes for HDicity
- Based on what we just did with RE, any ideas on one way to fix HDicity?
- If we know how the unobservables vary, we can specify it
- If we can specify it, we can do GLS

- There are two common fixes for HDicity
- Based on what we just did with RE, any ideas on one way to fix HDicity?
- If we know how the unobservables vary, we can specify it
- If we can specify it, we can do GLS
- Example:

$$Var(u_i) = \sigma_u^2 x_i$$

- There are two common fixes for HDicity
- Based on what we just did with RE, any ideas on one way to fix HDicity?
- If we know how the unobservables vary, we can specify it
- If we can specify it, we can do GLS
- Example:

$$Var(u_i) = \sigma_u^2 x_i$$

- Then we transform the data so that $Var(u*_i) = \sigma_u^2$

- Model is:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$
$$\text{where } E(u_i) = 0$$
$$\text{and } V(u_i) = \sigma_u^2 x_i$$

- Can you figure out how to run GLS on this model? Can you figure out how to transform the data?
- When reviewing your notes, try before advancing beyond this point.
- Hint: $V(u_i) = E[(u_i - E(u_i))^2] = \sigma_u^2 x_i$

- Model is:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$
$$\text{where } E(u_i) = 0$$
$$\text{and } V(u_i) = \sigma_u^2 x_i$$

- Model is:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$
$$\text{where } E(u_i) = 0$$
$$\text{and } V(u_i) = \sigma_u^2 x_i$$

- We need to transform the data so that $V(u*_i) = \sigma_u^2$ instead of $\sigma_u^2 x_i$

- Model is:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$
$$\text{where } E(u_i) = 0$$
$$\text{and } V(u_i) = \sigma_u^2 x_i$$

- We need to transform the data so that $V(u*_i) = \sigma_u^2$ instead of $\sigma_u^2 x_i$
- Try:

$$u*_i = \frac{u_i}{\sqrt{x_i}}$$

- If we do this simple transformation ...

$$
\begin{aligned}
V(u*_i) &= E[((u*_i) - E(u*_i))^2] \\
&= E[(u*_i)^2 - 2(u*_i)E(u*_i) + E(u*_i)^2] \\
&= E[(u*_i)^2] \\
&= E[\left(\frac{u_i}{\sqrt{x_i}}\right)^2] \\
&= E[\frac{u_i^2}{x_i}] \\
&= \frac{1}{x_i}(\sigma_u^2 x_i) \\
&= \sigma_u^2
\end{aligned}
$$

- So the GLS model would be OLS performed on the transformed model

$$y*_i = \beta_0 + \beta_1 x *_i + u*_i$$

where *starred* variables are divided by $\sqrt{x_i}$:

$$\frac{y_i}{\sqrt{x_i}} = \beta_0 \frac{1}{\sqrt{x_i}} + \beta_1 \frac{x_i}{\sqrt{x_i}} + \frac{u_i}{\sqrt{x_i}}$$

- So the GLS model would be OLS performed on the transformed model

$$y*_i = \beta_0 + \beta_1 x *_i + u*_i$$

where *starred* variables are divided by $\sqrt{x_i}$:

$$\frac{y_i}{\sqrt{x_i}} = \beta_0 \frac{1}{\sqrt{x_i}} + \beta_1 \frac{x_i}{\sqrt{x_i}} + \frac{u_i}{\sqrt{x_i}}$$

- This is one way to deal with heteroskedasticity

- The second way to deal with heteroskedasticity is, in some ways, simpler
- It is also much more common. Because:
  1. Because it is physically easier to do
  2. Because it is often hard to know the exact form of the heteroskedasticity

- The second way to deal with heteroskedasticity is, in some ways, simpler
- It is also much more common. Because:
  1. Because it is physically easier to do
  2. Because it is often hard to know the exact form of the heteroskedasticity
- The first way is based on transforming the data to get *rid* of the HDicity
- The second way is based on leaving the HDicity alone (since HDicity doesn't influence coefficient estimates), but taking it into account when we calculate the standard errors

- The regular coefficient s.e.s are

$$s.e.(\widehat{\beta}_j) =$$

- The regular coefficient s.e.s are

$$s.e.(\widehat{\beta}_j) = \frac{\sigma_u}{\sqrt{SST_j(1 - R_j^2)}}$$

if $V(u_i) = \sigma_u^2$
ref p. 102

Heteroskedasticity
two fixes: #2

- The regular coefficient s.e.s are

$$s.e.(\widehat{\beta}_j) = \frac{\sigma_u}{\sqrt{SST_j(1 - R_j^2)}}$$

if $V(u_i) = \sigma_u^2$
ref p. 102

- Heteroskedasticity-robust s.e.s are

$$\sqrt{\frac{\sum_i \widehat{r_{ij}^2}\widehat{u_i^2}}{SSR_j^2}}$$

if $V(u_i) = \sigma_{u_i}^2$
ref p. 267

- In Stata:

Stata code

```
reg y x, robust
```

1. Heteroskedasticity is a problem because it affects our standard errors
2. If we know the *form* of the heteroskedasticity, we can specify it and use it to perform GLS
   - This is the way to go if we have a reason to suspect a particular form
3. If we don't know the form of the heteroskedasticity, we can be conservative in our calculation of s.e.s

- (TS econometrics will not be on the final)

- Modern econometrics is broken into two (nearly unrelated) branches
  1. Microeconometrics (what we've done so far)
  2. Time series econometrics (aka macroeconometrics)

- TS lesson #1: You can use OLS to estimate TS data

- TS lesson #1: You can use OLS to estimate TS data
- This is legitimate:

$$infl_t = \beta_0 + \beta_1 unem_t + u_t$$

- Unlike with micro data, often in macro it is totally unclear whether inflation causes unemployment or unemployment causes inflation
- Often it is safe to consider *both* variables to be endogenous (they are co-determined)

- Unlike with micro data, often in macro it is totally unclear whether inflation causes unemployment or unemployment causes inflation
- Often it is safe to consider *both* variables to be endogenous (they are co-determined)
- Unlike with cross-section data, there is a clear ordering of data so that:
  $unem_{t-1}$ was clearly determined *before* $infl_t$

- Unlike with micro data, often in macro it is totally unclear whether inflation causes unemployment or unemployment causes inflation
- Often it is safe to consider *both* variables to be endogenous (they are co-determined)
- Unlike with cross-section data, there is a clear ordering of data so that:
  $unem_{t-1}$ was clearly determined *before* $infl_t$
- Two legitimate models might be

$$infl_t = \beta_0 + \beta_1 unem_t + \beta_2 unem_{t-1} + u_t$$
$$infl_t = \beta_0 + \beta_2 unem_{t-1} + u_t$$

- Also legitimate to suppose that the effect of unemployment *persists* over multiple periods
- This leads to models like

$$infl_t = \beta_0 + \beta_1 unem_t + \beta_2 unem_{t-1} + \beta_3 unem_{t-2} + u_t$$

- In this case, unemployment can have an effect on inflation for three periods
- Totally legitimate to have a model with more lags

- Also legitimate to suppose that the effect of unemployment *persists* over multiple periods
- This leads to models like

$$infl_t = \beta_0 + \beta_1 unem_t + \beta_2 unem_{t-1} + \beta_3 unem_{t-2} + u_t$$

- In this case, unemployment can have an effect on inflation for three periods
- Totally legitimate to have a model with more lags
- In this type of model (known as an FDL model; p. 342):
  - β_1 is known as the instantaneous impact of *unem* on *infl*
  - $\beta_1 + \beta_2 + \beta_3$ is known as the long-run impact

- Synonyms
  - instantaneous impact, impact multiplier, impact propensity
  - long run impact, long run multiplier, long run propensity

- Much of basic TS econometrics is about how to deal with the concerns inherent to TS data when we want to run these basic models, e.g.:
  1. Seasonality of data
  2. Trends in data
  3. *Stationarity* of data

- Because ...
  1. It is often hard to tell which variables are exogenous *except* for lagged variables, and
  2. Lagged values of the dependent variable are often *excellent* predictors of future values
- ... a very popular approach in TS is to estimate y_t using lagged values of y_t

- Because . . .
  1. It is often hard to tell which variables are exogenous *except* for lagged variables, and
  2. Lagged values of the dependent variable are often *excellent* predictors of future values
- . . . a very popular approach in TS is to estimate y_t using lagged values of y_t
- This is an ARIMA model

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

- There has to be an initial period. Call this period 0

$$y_0 = \beta_0 + u_0$$

- Then what is the model of period 1 y?

$$y_1 = \beta_0 + \beta_1 y_0 + u_1$$

substitute to get

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 y_0 + u_1 \\ &= \beta_0 + \beta_1 (\beta_0 + u_0) + u_1 \\ &= \beta_0 + \beta_1 \beta_0 + \beta_1 u_0 + u_1 \end{aligned}$$

- This sort of awfulness goes on and on

$$\begin{aligned} y_2 &= \beta_0 + \beta_1 y_1 + u_2 \\ &= \beta_0 + \beta_1(\beta_0 + \beta_1\beta_0 + \beta_1 u_0 + u_1) + u_2 \\ &= \beta_0 + \beta_1\beta_0 + \beta_1^2\beta_0 + \beta_1^2 u_0 + \beta_1 u_1 + u_2 \end{aligned}$$

- What do you notice?

- This sort of awfulness goes on and on

$$\begin{aligned} y_2 &= \beta_0 + \beta_1 y_1 + u_2 \\ &= \beta_0 + \beta_1(\beta_0 + \beta_1\beta_0 + \beta_1 u_0 + u_1) + u_2 \\ &= \beta_0 + \beta_1\beta_0 + \beta_1^2\beta_0 + \beta_1^2 u_0 + \beta_1 u_1 + u_2 \end{aligned}$$

- What do you notice?
- No y left on the RHS
- It's all random stuff (u)

- A good portion of advanced econometrics is dealing with models that are highly dependent on the specification of the randomness in the model
- A lot less of this type of econometrics is based on intuition, causal models, and the experimental-type of reasoning that microeconometrics relies upon
- TS econometrics is essentially its own discipline
- I could go the rest of my career without running an ARIMA model

Exam

- 14 May from 6p to 9p in