Econometrics Part 4

Heteroskedasticity and Autocorrelation

Jose Angel Garcia Sanchez

jagarsanc@gmail.com

Sorbonne Data Analytics

November 2025

Outline

1. Generalized least squares
2. Heteroskedasticity
3. Autocorrelation

Outline

1. Generalized least squares
2. Heteroskedasticity
3. Autocorrelation

What if assumptions do not hold?

- Heteroskedasticity: error terms do not have the same variance (ex: size effect for firms) $\rightarrow$ mostly cross sections
- Autocorrelation: error term from period t is correlated to error term of period $t-1$, and/or $t-2$ ... $\rightarrow$ mostly time series

In that case, $V(u) \neq \sigma^2 I_N$: OLS are still unbiased but less precise. Furthermore, all the tests we derived are not valid any more since they rely on homoskedasticity assumptions. We thus have to modify the estimation to make a reliable assessment on parameter values.

The problem with OLS

Let's use the following model: $y = Xb + u$

- Assume that $V(u) = \sigma^2\Omega$, with Ω any suitable matrix (must be symmetric and positive definite)
- $E(\hat{b}_{ols}) = b$: still unbiased
- But $V(\hat{b}_{ols}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$
- If we run an OLS regression with a software, we will get $\hat{b}$ computed as $\hat{b} = (X'X)^{-1}X'y$ (right) but $\hat{V}(\hat{b})$ computed as $\hat{V}(\hat{b}) = \hat{\sigma}^2(X'X)^{-1}$ (wrong)
- The software will thus give us unbiased estimates, but wrong variances, so all the tests we may run are wrong (T-test, F-test)

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

How can we correct for these problems?

- We need to transform the model so as to get homoskedastic and independent error terms
- We can use *generalized* least squares, that will replace *ordinary* least squares
- We will eventually end up re-weighting the model with the help of the variance matrix of error terms

Generalized least squares

Let's use the following model: $y = Xb + u$

- Assume that $V(u) = \sigma^2 \Omega$
- Ω is symmetric and positive definite, so Ω^{-1} exists and $\Omega^{-1/2}$ exists as well
- $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$
- Let's multiply the model on the left handside by $\Omega^{-1/2}$: $\Omega^{-1/2}y = \Omega^{-1/2}Xb + \Omega^{-1/2}u$
- The error terms of this "new" model are homoskedastic and uncorrelated: $V(\Omega^{-1/2}u) = \sigma^2 I_N$
- We can use the transformed model for the estimation of b and most of all for inference

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

The GLS estimator

- Using OLS on the transformed model, we get the GLS estimator
- $\hat{b}_{gls} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$
- It is obtained by minimizing the sum of squared residuals *weighted* by Ω^{-1}
- GLS are unbiased
- The "true" variance that can be used for tests is the following: $V(\hat{b}_{gls}) = \sigma^2(X'\Omega^{-1}X)^{-1}$
- Notice that the transformed model does not necessarily have a meaning: it is only useful for the estimation procedure

Using GLS

Since GLS amounts to OLS computed on a model that has nice properties, we can implement all the tests we described earlier (T-test, F-test)

The FGLS estimator

- GLS have nice properties
- The problem is however that we do not know the true Ω: we thus have to use an estimator of Ω and we thus turn to the *feasible* generalized least squares FGLS
- We cannot compute all the variances and covariances for the N error terms, since we have N observations
- We'll have to estimate these variances and covariances in some way (see later)

Remarks

- FGLS is an estimate of true GLS
- When sample size goes to infinity, estimation becomes almost perfect
- So FGLS and GLS are asymptotically equivalent
- FGLS have good properties only asymptotically
- FGLS are usually biased with a small sample
- Thus, in small samples, we cannot be sure that FGLS outperform OLS

Outline

1 Generalized least squares

2 Heteroskedasticity

3 Autocorrelation

Heteroskedasticity

- Heteroskedasticity usually occurs with cross sections
- Let's assume error terms are uncorrelated but do not have the same variance (ex: households with very different income levels)
- The variance matrix of error terms is a diagonal matrix $\sigma^2\Omega$
- Is it thus easy to compute Ω^{-1}
- This is sometimes called the Weighted Least Squares estimator (WLS) because each observation is weighted proportionally to the inverse of the standard error of its corresponding error term
- It is as if we put a smaller weight on high-income households

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Re-weighting matrix

$$\Omega = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & a_N \end{pmatrix} \text{ so } \Omega^{-1} = \begin{pmatrix} 1/a_1 & 0 & \cdots & 0 \\ 0 & 1/a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1/a_N \end{pmatrix}$$

And the re-weighting matrix is:

$$\Omega^{-1/2} = \begin{pmatrix} 1/\sqrt{a_1} & 0 & \cdots & 0 \\ 0 & 1/\sqrt{a_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1/\sqrt{a_N} \end{pmatrix}$$

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Testing for heteroskedasticity

- First, simply looking at residuals
- Then, various tests are available
- *Breusch-Pagan* test: regress the squared residuals on all the explanatory variables
- *White* test: regress the squared residuals on all the explanatory variables raw, squared, and their cross-products
- If we find some parameters significantly different from 0, then we conclude there is heteroskedasticity
- We will eventually use these regressions to estimate Ω and lead the weighted regression

Weighted least squares in practice (1)

Let's consider a heteroskedastic model $y = Xb + u$

- The variance matrix of error terms is a diagonal with N unknown elements $V(u_i)$
- Assume that $V(u)$ depends on variables X: $V(u) = \sigma^2 \cdot \exp(X\gamma)$ (we drop i's)
- The exponential ensures everything is always positive
- The squared residuals are the empirical counterparts of the error variances because $V(u_i) = E(u_i^2)$
- So this equation becomes empirically: $\hat{u}^2 = \sigma^2 \cdot \exp(X\gamma)$ so that $\log(\hat{u}^2) = X\gamma + constant$
- The error variances can now be estimated

Weighted least squares in practice (2)

What to do:

1. Estimate the original regression $y = Xb + u$ with OLS, save residuals, compute their square
2. Using plots or any other method, pick the right X's for the auxiliary regression: $\log(\hat{u}^2) = X\gamma + constant$
3. Compute the *predicted* residual squared from the auxiliary regression (need to get back to levels from logs)
4. For each individual i, divide every term of the original model by the square root of that prediction
5. Run OLS on transformed data

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Should we always use FGLS?

- FGLS do not have nice small sample properties, while OLS do
- But using OLS gives us wrong standard errors for estimates
- We could use OLS and compute corrected standard errors for them
- So we can keep OLS estimates because they are unbiased and consistent, and use the White (1980) matrix to run tests only
- That's what most people do

The White matrix

White (1980) proved that a consistent estimator for the variance matrix of parameters is:

$$\hat{V}(\hat{b}_{ols}) = (X'X)^{-1}(\sum_{t=1}^{N} \hat{u}_t^2 X_t X_t')(X'X)^{-1}$$

Where the $\hat{u}$'s are the residuals from the OLS regression.
This matrix can be used as an estimate of the true variance of OLS estimators, and thus for tests. They are called "White" or "robust" standard errors (because they are robust to heteroskedasticity), and can be computed directly by softwares (option "robust").

Outline

1. Generalized least squares
2. Heteroskedasticity
3. Autocorrelation

Autocorrelation

- Error term from period t is correlated to error term of period $t-1$, and/or $t-2$ ...
- Focus here on first-order autocorrelation: $u_t = \rho u_{t-1} + \varepsilon_t$
- Where $|\rho| < 1$ and ε is a "white noise"
- "White noise" means $\forall t$, $E(\varepsilon_t) = 0$ *and* $V(\varepsilon_t) = \sigma_\varepsilon$, all ε_t uncorrelated to each other and to u's

First-order autocorrelation

The variance of error terms $\sigma^2\Omega$ is no longer a diagonal matrix because Ω is no longer the identity matrix:

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{pmatrix}$$

And we have: $E(u_t) = 0$, $V(u_t) = \sigma^2 = \frac{\sigma_\varepsilon^2}{1-\rho^2}$.
The correlation coefficient between u_{t-h} and u_t is $\rho^{|h|}$.

Testing for first-order autocorrelation

Durbin-Watson test: $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$
Test statistic:

$$DW = \frac{\sum_{t=2}^{N}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{N} \hat{u}_t^2}$$

And

$$DW \approx 2(1 - \hat{\rho})$$

The Durbin-Watson table gives us two values d_1 and d_2 such that:
$DW \in [0, d_1] \Rightarrow \rho > 0$
$DW \in [d_1, d_2] \Rightarrow \textit{inconclusive}$
$DW \in [d_2, 4 - d_2] \Rightarrow \rho = 0$
$DW \in [4 - d_2, 4 - d_1] \Rightarrow \textit{inconclusive}$
$DW \in [4 - d_1, 4] \Rightarrow \rho < 0$
Requisites: the model needs a constant term; in the explanatory variables, we can't have the dependent variable lagged

The Breusch-Godfrey test

- This test is used to detect autocorrelation > 1 and works even when the dependent variable lagged belongs to the explanatory variables
- Run OLS on the model
- Save residuals e_t
- Estimate the auxiliary regression explaining e_t as a linear function of explanatory variables $+p$ previous residuals
- Under the null (parameters of previous residuals simultaneously equal to zero), $(N - p) \cdot R^2$ follows a χ^2 distribution of parameter p
- In the test, the R^2 used is the one from the auxiliary regression

What to do if we find autocorrelation

- The finding of autocorrelation often means that the model is misspecified
- We thus should not directly switch from OLS to FGLS, but rather try to change the model
- Ex: include as explanatory variable the dependent variable lagged (y from previous period), switch from x to $\log(x)$, use variables expressed as growth rates instead of levels, etc.

FGLS in the case of autocorrelation

- If modifying the model doesn't help, we have to correct for autocorrelation using FGLS, just as we did in the case of heteroskedasticity
- We will use "quasi-differencing"

Reminder: first-order autocorrelation

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{pmatrix}$$

So we know the form of Ω^{-1}:

$$\Omega^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots -\rho & 1 \end{pmatrix}$$

The FGLS transforming matrix

We can find a matrix Γ such that $(1-\rho^2)\Omega^{-1} = \Gamma'\Gamma$ with:

$$\Gamma = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots - \rho & 1 \end{pmatrix}$$

And the FGLS transformed model is written as: $\Gamma y = \Gamma X b + \Gamma u$.

The transformed model

- The FGLS transformed model is written as: $\Gamma y = \Gamma X b + \Gamma u$
- In the new model, y_1 is replaced by $(\sqrt{1-\rho^2})y_1$ and all the other outcome variables y_i are replaced by $y_i - \rho y_{i-1}$
- Same for each independent variable x_j including the constant
- Except for the 1st observation, all observations are "quasi-differenced"
- $y_i - \rho y_{i-1} = b_0(1-\rho) + b_1(x_{1,i} - \rho x_{1,i-1}) + ... + \varepsilon_i$
- $\forall i > 1$, $u_i - \rho u_{i-1} = \varepsilon_i$ which is a white noise: now regular OLS can be used on this new model
- The last thing we need to run the regression is an estimate of ρ, $\hat{\rho}$

Estimation: the two-step method

1. Run original OLS regression $y = Xb + u$,
2. Save residuals $\hat{u}_i$
3. Estimate ρ by $\hat{\rho} = \frac{\sum \hat{u}_i \hat{u}_{i-1}}{\sum \hat{u}_{i-1}^2}$
4. Quasi-difference the model using $\hat{\rho}$
5. Then run OLS on the transformed model

Estimation: the iterative method

Cochrane-Orcutt or *Prais-Winsten* method

1. Run the two-step method once, save $\hat{\rho}_1$ and $\hat{b}_1$
2. From OLS on the (already) transformed model, save estimated parameters and compute residuals
3. Use these residuals to compute a new $\hat{\rho}$, $\hat{\rho}_2$
4. Repeat until two successive $\hat{\rho}$'s are sufficiently close to each other

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

Last, what if we have both autocorrelation and heteroskedasticity?

- We may use a generalization of the White variance matrix if we have both heteroskedasticity and autocorrelation that could not be corrected for
- It is used when we assume that the correlation between the u's is zero after some lag
- We get the so-called *Newey-West standard errors*, that are heteroskedasciticy and autocorrelation consistent