FDPE core course

Principles of Econometrics II

"Model building and specific model types"

1) Heteroskedasticity and generalized least squares

Timo Kuosmanen

Professor

•Helsinki School of Economics

•MTT Agrifood Research Finland

Today's topics

- results of the 1st midterm exam
- schedule and topics of the spring term
- nonspherical disturbances
- causes of heteroskedasticity
- inefficiency of OLS
- generalized least squares (GLS)
- feasible generalized least squares (FGLS)
- testing for heteroskedasticity
- efficient GMM

Schedule – Spring term

Theme: Model building and specific model types

Date	Topic
Jan 14	**1.** Heteroskedasticity and generalized least squares (4h)
Jan 21	**2.** Serial correlation and lagged variables (4h)
Jan 28	**3.** Time series models (2h) + exerc. 1
Feb 1	**4.** Panel data models (4h)
Feb 18	**5.** Discrete response models (2h) + exerc. 2
Mar 4	**6.** Limited dependent variable and duration models (4h)
Mar 11	**7.** Systems of equations (2h) + exerc. 3
Mar 18	**8.** Modelling strategies (2h) + exerc. 4

2nd Midterm exam: April 16, 2010, from 12 p.m. to 4 p.m.

Classical linear regression model

A1. Linearity: There is a linear relationship between **y** and **X**.

A2. Full rank: columns of **X** are linearly independent and $n \geq K$.

A3. Exogeneity: $E(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \mathbf{0}$. Observations on **x** convey no information about the expected value of $\boldsymbol{\varepsilon}$.

A4. Spherical disturbances: $\mathrm{E}(\boldsymbol{\varepsilon\varepsilon}' \mid \mathbf{X}) = \sigma^2\mathbf{I}$

A4a. Homoskedasticity: $\mathrm{Var}(\varepsilon_i) = \sigma^2$ for all i

A4b. Nonautocorrelation: $\mathrm{Cov}(\varepsilon_h, \varepsilon_j \mid \mathbf{X}) = 0$ for all $i \neq h$

[A6. Normality: $\boldsymbol{\varepsilon} \mid \mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$**]**

Generalized linear regression model

A1. Linearity: There is a linear relationship between **y** and **X**.

A2. Full rank: columns of **X** are linearly independent and $n \geq K$.

A3. Exogeneity: $E(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \mathbf{0}$. Observations on **x** convey no information about the expected value of $\boldsymbol{\varepsilon}$.

A4. Nonspherical disturbances: $E(\boldsymbol{\varepsilon\varepsilon}' \mid \mathbf{X}) = \sigma^2\boldsymbol{\Omega}$

A4a. Heteroskedasticity allowed: $\mathrm{Var}(\varepsilon_i) = \sigma_i^2$

A4b. Autocorrelation allowed: $\mathrm{Cov}(\varepsilon_h, \varepsilon_j \mid \mathbf{X}) = \rho_{hi}$

[A6. Normality: $\boldsymbol{\varepsilon} \mid \mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$**]**

Covariance matrices

- Classical linear regression model

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' | \mathbf{X}) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 \mathbf{I}$$

Covariance matrices

- Generalized linear regression model

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' | \mathbf{X}) = \sigma^2 \begin{bmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1n} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{n1} & \omega_{n2} & \cdots & \omega_{nn} \end{bmatrix} = \sigma^2 \boldsymbol{\Omega}$$

Special cases

- Heteroskedasticity and no autocorrelation

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}) = \sigma^2 \begin{bmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_n \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

Special cases

- Autocorrelation and homoskedasticity

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' | \mathbf{X}) = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix}$$

Examples of heteroskedasticity

Arises in many types of applications

- Large firms typically exhibit greater variation in profit/output/productivity, even after accounting for the firm size
- High-income families exhibit greater variation in their expenditure on certain commodities

- Note: model misspecification might also show up as heteroskedasticity.

Estimating generalized regression model by OLS

Finite sample properties

- Is OLS still **unbiased?**
- Is OLS still **efficient?**

Asymptotic properties

- Is OLS still **consistent?**
- Is OLS still **asymptotically efficient?**

Estimating generalized regression model by OLS

Finite sample properties

- OLS is **unbiased:** If $E[\boldsymbol{\varepsilon}|\mathbf{X}]=\mathbf{0}$, then $E[\mathbf{b}] = \boldsymbol{\beta}$
  - recall 1 midterm exam

- OLS is **inefficient:** variance of OLS estimator is typically higher than the minimum variance

$$Var(\mathbf{b}|\mathbf{X}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}) (\mathbf{X}'\mathbf{X})^{-1}$$

Estimating generalized regression model by OLS

Note 1: Statistical inferences based on $s^2(\mathbf{X'X})^{-1}$ may be misleading:

- Even with a good estimator of σ^2, using $\sigma^2(\mathbf{X'X})^{-1}$ may under- or overestimate the true variance of **b.**

Estimating generalized regression model by OLS

Note 2: In the heteroskedastic case, White's (1980) estimator

$$Est.Asy.Var(\mathbf{b}) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\sum_{i=1}^{n} e_i^2 \mathbf{x}_i \mathbf{x}_i'\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

is consistent under very general conditions.

Thus, it is possible to make inferences based on OLS in the heteroskedastic case. Still, OLS remains inefficient

Estimating generalized regression model by OLS

Asymptotic properties

- OLS is **consistent:** If plim($\mathbf{X}'\mathbf{X}/n$) and plim($\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}/n$) are both positive definite matrices, then plim $\mathbf{b} = \boldsymbol{\beta}$.
- OLS is **asymptotically inefficient**

Note: In some cases, asymptotic normality of OLS estimator is preserved in the generalized model

  - E.g., heteroskedasticity with finite error variances

Violations of OLS assumptions => Consequences?

A1. Linearity: Relationship between **y** and **X** is nonlinear or unknown a priori => Model misspecification, coefficients meaningless?

A2. Full rank: Not enough observations with linearly independent data => OLS not computable, or coefficients are arbitrary

A3. Exogeneity: Endogeneity – Observed **X** contain information about the expected value of $\boldsymbol{\varepsilon}$. => OLS biased and inconsistent

A4. Spherical disturbances: Nonspherical disturbances: heteroskedasticity and autocorrelation => OLS inefficient

A6. Normality: $\boldsymbol{\varepsilon} \mid \mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ **:** Non-normal residuals (e.g. fat tails) => Conventional methods of statistical inference are invalid

Dealing with heteroskedasticity by weighted least squares (WLS)

- Suppose we know ε_i has a higher variance than ε_k.
- Therefore, it could make sense to assign residual e_k a larger weight than e_i in the least squares problem

=> Weighted least squares (WLS)

Which weights are appropriate?

Efficient GLS estimation

- Suppose matrix **Ω** is known
- Using Aitken's (1935) theorem, an efficient **generalized least squares** (GLS) estimator can be constructed

- In practice, GLS model can be estimated by first making certain **data transformations** to **y**,**X**, and then applying the standard OLS estimators to the resulting transformed data $\mathbf{y}_*,\mathbf{X}_*$
  - i.e., GLS model is simply $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

Efficient GLS estimation

Data transformations:

- Because **Ω** is a symmetric *n*x*n* positive definite matrix, it can be factored into

$$\boldsymbol{\Omega} = \mathbf{C\Lambda C'}$$

where

- **C** is a *nxn* matrix containing the **characteristic vectors** of **Ω** as its columns
- **Λ** is a *nxn* diagonal matrix with the **characteristic roots** of **Ω** arrayed on the diagonal ’
- see **Appendix A.6** for derivation of characteristic roots and vectors

Efficient GLS estimation

- Let $\mathbf{\Lambda}^{1/2}$ be a *nxn* diagonal matrix with square-roots of the characteristic roots as the diagonal elements
- Let $\mathbf{P'} = \mathbf{C\Lambda}^{1/2}$
- Thus, $\mathbf{\Omega}^{-1} = \mathbf{P'P}$

- Premultiply **y**, **X** by **P** to obtain the transformed data

  $\mathbf{y}_* = \mathbf{Py}$

  $\mathbf{X}_* = \mathbf{PX}$

Efficient GLS estimation

- In the heteroskedastic case,

$$\mathbf{\Omega}^{-1} = \begin{bmatrix} 1/\omega_1 & 0 & \cdots & 0 \\ 0 & 1/\omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\omega_n \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 1/\sqrt{\omega_1} & 0 & \cdots & 0 \\ 0 & 1/\sqrt{\omega_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sqrt{\omega_n} \end{bmatrix}$$

- Thus,

$$\mathbf{y}_* = \mathbf{P}\mathbf{y} = \begin{bmatrix} y_1/\sqrt{\omega_1} \\ y_2/\sqrt{\omega_2} \\ \vdots \\ y_n/\sqrt{\omega_n} \end{bmatrix}, \quad \mathbf{X}_* = \mathbf{P}\mathbf{X} = \begin{bmatrix} \mathbf{x}_1'/\sqrt{\omega_1} \\ \mathbf{x}_2'/\sqrt{\omega_2} \\ \vdots \\ \mathbf{x}_n'/\sqrt{\omega_n} \end{bmatrix}$$

Efficient GLS estimation

- GLS estimator in the heteroskedastic case,

$$\min \mathbf{e}'\mathbf{e}$$

$$s.t. \begin{bmatrix} y_1/\sqrt{\omega_1} \\ y_2/\sqrt{\omega_2} \\ \vdots \\ y_n/\sqrt{\omega_n} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1'/\sqrt{\omega_1} \\ \mathbf{x}_2'/\sqrt{\omega_2} \\ \vdots \\ \mathbf{x}_n'/\sqrt{\omega_n} \end{bmatrix}' \boldsymbol{\beta} + \mathbf{e}$$

- This is equivalent to the following weighted least squares estimator

$$\min \begin{bmatrix} e_1/\sqrt{\omega_1} \\ e_2/\sqrt{\omega_2} \\ \vdots \\ e_n/\sqrt{\omega_n} \end{bmatrix}' \begin{bmatrix} e_1/\sqrt{\omega_1} \\ e_2/\sqrt{\omega_2} \\ \vdots \\ e_n/\sqrt{\omega_n} \end{bmatrix}$$

$$s.t. \;\; \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

Properties of GLS

Finite sample properties

- GLS is **unbiased:** If $E[\boldsymbol{\varepsilon}|\mathbf{X}]=\mathbf{0}$, then $E[\mathbf{b}_{GLS}] = \boldsymbol{\beta}$
- GLS is **efficient:**

$$Var(\mathbf{b}_{GLS}|\mathbf{X}_*) = \sigma^2\left(\mathbf{X}_*'\mathbf{X}_*\right)^{-1} = \sigma^2\left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}$$

Aitken's theorem: the GLS estimator is the *minimum variance linear unbiased estimator* in the generalized regression model

Properties of GLS

Asymptotic properties

- GLS is **consistent:** If plim($\mathbf{X}_*'\mathbf{X}_*/n$) is a positive definite matrix, then plim $\mathbf{b}_{GLS} = \boldsymbol{\beta}$.
- GLS is **asymptotically efficient**
- GLS estimator is **asymptotically normally distributed**

In summary, all the results for the classical model, incl. the usual inference procedures, apply to the transformed model $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

Feasible Generalized Least Squares (FGLS)

- To apply GLS, matrix **Ω** must be known.
- This is rarely the case in practice: typically **Ω** must be estimated from data.
- To make GLS feasible, an estimator of **Ω** is needed.
- **Feasible GLS** (FGLS) uses an estimated **Ω** instead of assuming the true **Ω** is known.

Theorem: An **asymptotically efficient** FGLS estimator does not require an **efficient** estimator of **Ω**;

just a **consistent** one suffices to achieve full efficiency of the FGLS estimator

Feasible Generalized Least Squares (FGLS)

- If matrix $\mathbf{\Omega}$ consists of unknown paramaters, the symmetry of $\mathbf{\Omega}$ implies that there can be up to $n(n+1)/2$ unknowns. This is too many to be estimated with n observations.

- To estimate $\mathbf{\Omega}$, some structure must be imposed.

Models of heteroskedasticity

Harvey's model of multiplicative heteroskedasticity

$$\sigma_i^2 = \sigma^2 \exp(\mathbf{z}_i'\boldsymbol{\alpha})$$

Where $\mathbf{z}_i$ is a vector of variables that explain heteroskedasticity (e.g., income, firm size) and

$\boldsymbol{\alpha}$ is a parameter vector to be estimated

Models of heteroskedasticity

Model of groupwise heteroskedasticity

Sample partitioned into G groups, each with different error variance

$$\sigma_{ig}^2 = \sigma_g^2 \;\; \forall i \in g, g = 1, \ldots, G$$

Note: this is a special case of Harvey's model, where $\mathbf{z}_i$ is a set of group dummy variables.

Stepwise FGLS estimation

FGLS estimator can be computed in a stepwise fashion (here assuming Harvey's multiplicative model):

1) Estimate the model of interest by OLS, and save residuals e_i.
2) Estimate the following auxiliary model by OLS:

$$\ln e_i^2 = \mathrm{A} + \mathbf{z}_i'\boldsymbol{\alpha} + v_i$$

3) Compute the weights $\omega_i = \exp(\mathbf{z}_i'\boldsymbol{\alpha})$ and use them to transform data $\mathbf{y}$,$\mathbf{X}$ into $\mathbf{y}_*$,$\mathbf{X}_*$.
4) Apply OLS to the transformed data $\mathbf{y}_*$,$\mathbf{X}_*$.

Note: OLS estimator of $\boldsymbol{\alpha}$ in step 2 is consistent but likely inefficient. But consistency is all that is required for asymptotically efficient FGLS!

Testing for heteroskedasticity

White's general test

$$H_0 : \sigma_i^2 = \sigma^2 \ \forall i$$

$$H_1 : \sigma_i^2 \neq \sigma^2 \ \forall i$$

Test procedure: Use OLS to estimate the model

$$e_i^2 = \alpha_0 + \mathbf{x}_i'\boldsymbol{\alpha}_1 + \mathbf{x}_i'\boldsymbol{\alpha}_2\mathbf{x}_i + v_i$$

$$= \alpha_0 + \sum_{j=1}^{K} \alpha_j x_{ij} + \sum_{k=1}^{K}\sum_{j=1}^{K} \alpha_{jk} x_{ij} x_{ik} + v_i$$

Test statistic: nR^2

Under H_0, the test statistic has $\chi^2(P\text{-}1)$ distribution, where P is the no. of α parameters in the model.

Testing for heteroskedasticity

Breusch-Pagan LM test

$$H_0 : \sigma_i^2 = \sigma^2 \ \forall i$$

$$H_1 : \sigma_i^2 = \sigma^2 f(\alpha_0 + \mathbf{z}_i'\boldsymbol{\alpha}) \ \forall i$$

Test procedure: Use OLS to estimate the model

$$\frac{e_i^2}{\mathbf{e}'\mathbf{e}/n} = \alpha_0 + \mathbf{z}_i'\boldsymbol{\alpha} + v_i$$

Test statistic: ½ SSR

Under H_0, the test statistic has $\chi^2(P)$ distribution, where P is the no. of α parameters in the model.

GMM estimation

- Recall the GMM minimum distance estimator

$$\min \bar{\mathbf{m}}(\boldsymbol{\beta})'\mathbf{W}\bar{\mathbf{m}}(\boldsymbol{\beta})$$

- Analogous to GLS, the optimal weighting matrix **W** uses weights that are inversely proportional to the variances of the moments

- **W** allows different weights for different moments in the overidentified case: weights **W** have no effect in the exactly identified case

GMM estimation

- What if there is both heteroskedasticity and endogeneity?
- Using instruments **Z**, we can apply 2-step GMM estimation:

Step 1: Estimate the GMM model using **W** = **I**, and save residuals e_i. Apply White's heteroskedasticity consistent estimator and construct weights as

$$\hat{\mathbf{V}} = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \mathbf{z}_i \mathbf{z}_i'$$

$$\mathbf{W} = \hat{\mathbf{V}}^{-1}$$

Step 2: Use **W** to compute the GMM estimator

GMM estimation

- What is the GMM estimator in the heteroskedastic case if exogeneity holds?

- When the orthogonality conditions are defined in terms of regressors **x** (rather than instruments **z**), the GMM model is exactly identified.
- Thus, the weighting matrix **W** is irrelevant

=> The GMM estimator is identical to OLS, with White's heteroskedasticity consistent covariance estimator!

Conclusions

- In the heterorskedastic case, OLS remains unbiased and consistent, but becomes inefficient.
- GLS is efficient, but the true weights are typically unknown
- If the type of heteroskedasticity is known so that weights can be estimated, FGLS can improve efficiency

Next time…
21 January

Topic: Serial correlation and lagged variables

Read in advance:

- G07: Chapters 19, 20
- G03: Chapters 12, 19