Regression analysis with cross-sectional data: Heteroskedasticity

Bernard Lejeune

Supplemental lecture notes III

Advanced Econometrics
HEC-University of Liège
Academic year 2021-2022

- These lecture notes restate, in matrix form and with more details, the main results of Sections 8-1, 8-2 and 8-4 of Wooldridge (2016).

1. Consequences of heteroskedasticity for OLS

- From Property 1 and Property 6 in the supplemental lecture notes I (hereafter SLN-I)[1], we know that the OLS estimator $\hat{\beta}$ is unbiased under assumptions E.1 – E.3, as well as also consistent under the seminal assumptions MLR.1 – MLR.4., i.e., without having to assume homoskedasticity.

- If the presence of heteroskedasticity does not cause the OLS estimator to be biased or inconsistent, it has other consequences:

  (1) As the homoskedasticity assumption is part of the Gauss-Markov assumptions, the OLS estimator $\hat{\beta}$ is no longer the best linear unbiased estimator (BLUE) of β.

  (2) The usual – exact in finite sample under normality and asymptotically valid without normality – inference procedures are no longer valid, basically because the usual formulas for the exact and asymptotic variance-covariance of the OLS estimator $\hat{\beta}$ are no longer correct. As a matter of fact, from Section 3.1.3 in SLN-I, under assumptions E.1 – E.3, we have:

$$\hat{\beta} = \beta + (X'X)^{-1} X'u \quad \text{and} \quad E(\hat{\beta}|X) = \beta$$

[1] Regression analysis with cross-sectional data: specification, estimation and inference.

so that:

$$\begin{aligned} V(\hat{\beta}|X) &= E\left[(\hat{\beta}-E(\hat{\beta}|X))(\hat{\beta}-E(\hat{\beta}|X))'|X\right] \\ &= E\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'|X\right] \\ &= E\left[(X'X)^{-1}X'uu'X\,(X'X)^{-1}|X\right] \\ &= (X'X)^{-1}X'E(uu'|X)X\,(X'X)^{-1} \\ &= (X'X)^{-1}X\Sigma X\,(X'X)^{-1} \end{aligned} \tag{1}$$

where, maintaining the no correlation part of assumption E.4, $\Sigma = E(uu'|X) = V(u|X) = V(Y|X)$ is a $n \times n$ diagonal matrix:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{bmatrix} \tag{2}$$

with diagonal elements σ_i^2 equal to the conditional variances[2] $Var(u_i|X) = Var(y_i|X)$, $i = 1, ..., n$. The same is true (the usual formula is no longer correct) for the asymptotic variance $Avar(\hat{\beta})$ of $\hat{\beta}$ (see below).

- The fact that the OLS estimator is no longer BLUE may be addressed by deriving a more efficient estimator – called the weighted least squares estimator – which explicitly takes into account the presence of heteroskedasticity. This however requires that the form of heteroskedasticity is known, or can be estimated. On the other hand, it is possible to address the fact that the usual inference procedures are no longer valid by deriving inference procedures which are valid in the presence of heteroskedasticity of unknown form, i.e., which are robust to arbitrary form of heteroskedasticity.

2. Heteroskedasticity robust inference after OLS estimation

- Heteroskedasticity robust inference procedures are only asymptotically valid, i.e. approximately valid for n sufficiently large. They basically rely on the following property, which outlines the asymptotic properties of the OLS estimator $\hat{\beta}$ under the minimal set of assumptions MLR.1 – MLR.4 (linearity in parameters, random sampling, no perfect colinearity and zero conditional mean), i.e., without both the homoskedasticity and the normality assumptions MLR.5 and MLR.6.:

  Property 9 Asymptotic properties of $\hat{\beta}$ without homoskedasticity and normality
  Under assumptions MLR.1 – MLR.4, the OLS estimator $\hat{\beta}$ is a consistent estimator of β:

$$\hat{\beta} \xrightarrow{p} \beta$$

[2] Under homoskedasticity, $\sigma_i^2 = \sigma^2$(a constant) for all $i = 1, ..., n$, so that $\Sigma = \sigma^2 I$, and $V(\hat{\beta}|X)$ can be simplified to yield the usual formula $V(\hat{\beta}|X) = \sigma^2\,(X'X)^{-1}$.

and is asymptotically normally distributed as:

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(0, A^{-1}BA^{-1}) \tag{3}$$

where:

$$A = E(X_i'X_i) \quad \text{and} \quad B = E(u_i^2 X_i'X_i)$$

The fact that $\hat{\beta}$ is consistent for β was already outlined by Property 6 in SLN-I. A sketch of the proof of the limiting distribution result (3) is actually the same as the sketch of the proof of Property 7 in SLN-I. Hereafter is a concise version of it. Under assumptions MLR.1 – MLR.3, we have:

$$\sqrt{n}(\hat{\beta}-\beta) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(n^{-\frac{1}{2}}\sum_{i=1}^{n} X_i'u_i\right)$$

Under random sampling, both $X_i'X_i$ and $X_i'u_i$ are i.i.d. across i. From the law of large numbers[3] (LLN), we have $\frac{1}{n}\sum_{i=1}^{n} X_i'X_i \xrightarrow{p} E(X_i'X_i) = A$, and we can write:

$$\sqrt{n}(\hat{\beta}-\beta) \overset{as}{=} A^{-1}\left(n^{-\frac{1}{2}}\sum_{i=1}^{n} X_i'u_i\right)$$

so that $\sqrt{n}(\hat{\beta}-\beta)$ has asymptotically the same distribution as $A^{-1}\left(n^{-\frac{1}{2}}\sum_{i=1}^{n} X_i'u_i\right)$. Under the zero conditional mean assumption MLR.4, we have $E\left(X_i'u_i\right) = 0$. From the central limit theorem[4] (CLT), we thus have:

$$n^{-\frac{1}{2}}\sum_{i=1}^{n} X_i'u_i \xrightarrow{d} N(0,B), \quad \text{where } B = V(X_i'u_i) = E(u_i^2 X_i'X_i)$$

so that:

$$A^{-1}\left(n^{-\frac{1}{2}}\sum_{i=1}^{n} X_i'u_i\right) \xrightarrow{d} N(0, A^{-1}BA^{-1})$$

and thus we finally have[5]:

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(0, A^{-1}BA^{-1})$$

- As the limiting distributional result of Property 7 under assumptions MLR.1 – MLR.5 in SLN-I, the limiting distributional result (3) provides an approximate finite sample distribution for the OLS estimator $\hat{\beta}$, which can be used – when n is sufficiently large – for performing inference (confidence interval, hypothesis testing) without having to rely on neither the homoskedasticity assumption MLR.5 nor the normality assumption MLR.6. From (3), in terms of approximation, we have:

$$\sqrt{n}(\hat{\beta}-\beta) \approx N(0, A^{-1}BA^{-1})$$

[3] If $\{Z_i\colon i = 1, ..., n\}$ are i.i.d. random variables with $E(Z_i) = m$, then by the LLN we have: $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i \xrightarrow{p} m$.

[4] if $\{Z_i\colon i = 1, ..., n\}$ are i.i.d. $(k \times 1)$ random vectors with $E(Z_i) = m$ and $V(Z_i) = \Sigma$, then by the CLT we have: $\sqrt{n}(\bar{Z}_n - m) = n^{-\frac{1}{2}}\sum_{i=1}^{n}(Z_i - m) \xrightarrow{d} N(0,\Sigma)$.

[5] As a reminder, under the homoskedasticity assumption MLR.5, $B = V(X_i'u_i) = E(u_i^2 X_i'X_i) = \sigma^2 A$, so that $A^{-1}BA^{-1} = \sigma^2 A^{-1}$.

so that:

$$\hat{\beta} \approx N(\beta, A^{-1}BA^{-1}/n) \tag{4}$$

i.e., for n sufficiently large, $\hat{\beta}$ can be treated as approximately normal with mean β and asymptotic variance-covariance matrix $Avar(\hat{\beta}) = A^{-1}BA^{-1}/n$. Note that $Avar(\hat{\beta}) \to 0$ as $n \to \infty$. Note further that if we replace $A = E(X_i'X_i)$ by its consistent estimator $\frac{1}{n}\sum_{i=1}^n X_i'X_i = X'X/n$, remark that[6] $B = E(u_i^2X_i'X_i) = E\left(\sigma_i^2X_i'X_i\right)$ where $\sigma_i^2 = E(u_i^2|X_i) = Var(u_i|X_i)$, and likewise replace $E\left(\sigma_i^2X_i'X_i\right)$ by its consistent estimator[7] $\frac{1}{n}\sum_{i=1}^n \sigma_i^2X_i'X_i = X'\Sigma X/n$, where Σ is as defined in (2), then $Avar(\hat{\beta})$ becomes:

$$\begin{aligned} Avar(\hat{\beta}) &= \frac{A^{-1}BA^{-1}}{n} \\ &\approx \frac{(X'X/n)^{-1}\, X'\Sigma X/n\ (X'X/n)^{-1}}{n} = (X'X)^{-1}\, X\Sigma X\, (X'X)^{-1} \end{aligned}$$

This is the same as the exact in finite sample variance-covariance matrix $V(\hat{\beta}|X)$ obtained in (1) above.

- For inference based on the limiting distributional result (3), or equivalently on the approximate distributional result (4), we need an estimator of $Avar(\hat{\beta}) = A^{-1}BA^{-1}/n$. This requires consistent estimators of A and B. We already know that $X'X/n$ is a consistent estimator of A (this directly follows from the LLN). A consistent estimator of $B = E(u_i^2X_i'X_i)$ is simply given by $\frac{1}{n}\sum_{i=1}^n \hat{u}_i^2X_i'X_i$, where obviously $\hat{u}_i = y_i - X_i\hat{\beta}$. This is formalized in the following property[8]:

  Property 10 Consistent estimator of B
  Under MLR.1 – MLR.4, $\frac{1}{n}\sum_{i=1}^n \hat{u}_i^2X_i'X_i$ is a consistent estimator of B:

  $$\frac{1}{n}\sum_{i=1}^n \hat{u}_i^2X_i'X_i \xrightarrow{p} B$$

  Here is the intuition of this property: from the LLN, $\frac{1}{n}\sum_{i=1}^n u_i^2X_i'X_i \xrightarrow{p} E(u_i^2X_i'X_i) = B$. As $\hat{u}_i$ converges to u_i (because $\hat{\beta} \xrightarrow{p} \beta$), we also have $\frac{1}{n}\sum_{i=1}^n \hat{u}_i^2X_i'X_i \xrightarrow{p} B$.

- With $X'X/n$ and $\frac{1}{n}\sum_{i=1}^n \hat{u}_i^2X_i'X_i$ as consistent estimators of A and B, an estimator of $Avar(\hat{\beta}) = A^{-1}BA^{-1}/n$ is given by:

  $$\hat{V}_{HC}(\hat{\beta}) = (X'X)^{-1}\left(\sum_{i=1}^n \hat{u}_i^2X_i'X_i\right)(X'X)^{-1} \tag{5}$$

---

[6] By the law of iterated expectation, we have: $B = E(u_i^2X_i'X_i) = E\left[E(u_i^2X_i'X_i|X_i)\right] = E\left[E(u_i^2|X_i)X_i'X_i\right] = E\left(\sigma_i^2X_i'X_i\right)$.

[7] This is not an operational estimator because the σ_i^2's are unknown. An operational estimator is given below in Proterty 10.

[8] A variant of this estimator, with a degrees of freedom adjustment, is outlined in Wooldridge (2016), Appendix E-4. For a detailed proof (under weaker assumptions than MLR.1 – MLR.4), see Hayashi (2000) p. 123-124. You may also see Wooldridge (2010), Chapter 4. Whether or not a degrees of freedom adjustment is used does not matter asymptotically.

where the subscript 'HC' stands for 'Heteroskedasticity Consistent'. As usual, the diagonal elements $V\hat{a}r_{HC}(\hat{\beta}_j)$ of the $k \times k$ matrix estimator $\hat{V}_{HC}(\hat{\beta})$ being the estimators of the variance $Avar(\hat{\beta}_j)$ of the estimator $\hat{\beta}_j$ of the different parameters β_j $(j = 1, ..., k)$, natural estimators of the asymptotic standard error $As.e.(\hat{\beta}_j) = \sqrt{Avar(\hat{\beta}_j)}$ of the estimator $\hat{\beta}_j$ of the different parameters β_j, as well as a natural estimator of the asymptotic standard error $As.e.(R_0\hat{\beta}) = \sqrt{Avar(R_0\hat{\beta})} = \sqrt{R_0 Avar(\hat{\beta})R_0'}$ of the estimator $R_0\hat{\beta}$ of a single linear combination $R_0\beta$ of β, are likewise given by:

$$s.\hat{e}._{HC}(\hat{\beta}_j) = \sqrt{V\hat{a}r_{HC}(\hat{\beta}_j)}, \quad j = 1, ..., k \tag{6}$$

and:

$$s.\hat{e}._{HC}(R_0\hat{\beta}) = \sqrt{R_0\hat{V}_{HC}(\hat{\beta})R_0'} \tag{7}$$

where R_0 is a $1 \times k$ (row) vector of constants.

- The limiting distributional result (3), or equivalently the approximate distributional result (4), and the heteroskedasticity consistent (or heteroskedasticity robust) estimators $\hat{V}_{HC}(\hat{\beta})$, $s.\hat{e}._{HC}(\hat{\beta}_j)$ and $s.\hat{e}._{HC}(R_0\hat{\beta})$ given above in respectively (5), (6) and (7), provide all which is needed for performing robust inference after OLS estimation. Following exactly the same reasoning as in Section 4.3 and Section 4.4.2 of SLN-I[9], it may readily be checked that if in all the usual exact in finite sample inference procedures – confidence interval for β_j or a single linear combination $R_0\beta$, two-sided and one-sided t-tests of β_j or a single linear combination $R_0\beta$, F-test (or Wald test) of multiple linear restrictions – we replace the usual estimators $\hat{V}(\hat{\beta})$, $s.\hat{e}.(\hat{\beta}_j)$ and $s.\hat{e}.(R_0\hat{\beta})$ by their heteroskedasticity consistent (or heteroskedasticity robust) versions $\hat{V}_{HC}(\hat{\beta})$, $s.\hat{e}._{HC}(\hat{\beta}_j)$ and $s.\hat{e}._{HC}(R_0\hat{\beta})$, then we obtain inference procedures that are asymptotically valid – i.e., approximately valid for n sufficiently large – under only assumptions MLR.1 – MLR.4, i.e., without having to rely on neither the homoskedasticity assumption MLR.5 nor the normality assumption MLR.6.

- Remarks:

  - All modern econometric software optionally provide heteroskedasticity robust standard errors and allow to perform heteroskedasticity robust tests. In applied works, it is not uncommon that only the heteroskedasticity robust standard errors and tests are considered and reported.

  - The usual sum of squared residuals form of the F-test statistic, which is exact under assumptions MLR.1 – MLR.6 (and assumptions E.1 – E.5) and asymptotically valid under assumptions MLR.1 – MLR.5, i.e. without the normality assumption MLR.6, is no longer valid without the homoskedas-

[9] As a reminder, starting from the similar limiting distributional result of Property 7 and using the standard estimators $\hat{V}(\hat{\beta})$, $s.\hat{e}.(\hat{\beta}_j)$ and $s.\hat{e}.(R_0\hat{\beta})$, Section 4.3 and Section 4.4.2 in SLN-I show that the usual exact in finite sample inference procedures remain asymptotically valid – i.e., approximately valid for n sufficiently large – without the normality assumption.

ticity assumption MLR.5, and there exist no such form of the F-test statistic which is heteroskedasticity robust.

– The presence of heteroskedasticity may easily be checked by running an auxiliary regression of the form:

$$\hat{u}_i^2 = \delta_1 + \delta_2 x_{i2} + \delta_3 x_{i3} + ... + \delta_k x_{ik} + v_i, \quad i = 1, ..., n \tag{8}$$

An usual F-test of $\mathrm{H}_0: \ \delta_2 = ... = \delta_k = 0$ provides an asymptotically valid test of the null hypothesis of homoskedasticity. An asymptotically equivalent test may be performed using the LM statistic[10]:

$$LM = n \cdot R^2_{\hat{u}^2}$$

where $R^2_{\hat{u}^2}$ is the R-squared from the auxiliary regression (8). This test is known as the Breush-Pagan test for heteroskedasticity. A variant of the above tests adds the squares and cross-products of the explanatory variables in (8). Another variant considers instead the fitted values $\hat{y}_i$, $\hat{y}_i^2$, ... as explanatory variables in (8). See Wooldridge (2016) Section 8-3 for details.

3. Weighted least squares estimation

3.1. The generalized least squares estimator

- The weighted least squares (WLS) estimator is a special case of a more general estimator, called the generalized least squares (GLS) estimator. We first derive the GLS estimator.

- Consider a regression model satisfying the usual matrix assumptions E.1 – E.3 – i.e., linearity in parameters, no perfect colinearity and zero conditional mean – but where the homoskedasticity and no correlation assumption E.4 is relaxed and replaced by the fully non restrictive assumption:

  E.4bis Arbitrary heteroskedasticity & correlation:

$$V(u|X) = \sigma^2 \Omega \quad \Leftrightarrow \quad V(Y|X) = \sigma^2 \Omega$$

  where Ω is an arbitrary symmetric positive definite $n \times n$ matrix[11] whose elements may depend on X, so that we may have both non constant (conditional) variances and non zero (conditional) covariances across i, possibly depending of the value the explanatory variables.

- As assumptions E.1 – E.3 are maintained, we know from Property 1 in SLN-I

[10] The decision rule of this LM test is: reject H_0 if $LM > \chi^2_{k-1;1-\alpha}$ and do not reject otherwise, where $\chi^2_{k-1;1-\alpha}$ is the quantile of order $1-\alpha$ of the $\chi^2(k-1)$ distribution. The p-value of the test, for a value LM^* of the test statistic obtained in a particular sample, is given by: $p_{LM} = \mathbb{P}(v > LM^*)$, where $v \sim \chi^2(k-1)$.

[11] A variance-covariance matrix is always a symmetric and positive definite (or at least positive semidefinite) matrix.

that the OLS estimator $\hat{\beta}$ is still an unbiased estimator of β, but because of the violation of assumptions E.4, it is no longer the best linear unbiased estimator (BLUE) of β. We are looking for a better – i.e., more efficient, with a smaller variance-covariance matrix – estimator.

- Let $\Omega^{\frac{1}{2}}$ denote a symmetric positive definite $n \times n$ matrix[12] such that $\Omega^{\frac{1}{2}}\Omega^{\frac{1}{2}} = \Omega$. Intuitively, $\Omega^{\frac{1}{2}}$ may be viewed as the square root (in a matrix sense) of Ω. Let further $\Omega^{-\frac{1}{2}}$ denote the inverse of $\Omega^{\frac{1}{2}}$. We have:

$$Y = X\beta + u$$

so that, premultiplying both sides by $\Omega^{-\frac{1}{2}}$, we have:

$$\begin{aligned} \Omega^{-\frac{1}{2}}Y &= \Omega^{-\frac{1}{2}}X\beta + \Omega^{-\frac{1}{2}}u \\ \Leftrightarrow \quad Y^* &= X^*\beta + u^* \end{aligned} \tag{9}$$

where $Y^* = \Omega^{-\frac{1}{2}}Y$ and $X^* = \Omega^{-\frac{1}{2}}X$ are just transformations of the original variables Y and X.

- Under assumptions E.1 – E.3 and assumption E.4bis, the transformed model (9) is linear in parameters and has no perfect colinearity[13], and we have:

$$E(u^*|X) = E(\Omega^{-\frac{1}{2}}u|X) = \Omega^{-\frac{1}{2}}E(u|X) = \Omega^{-\frac{1}{2}} \cdot 0 = 0$$

and:

$$\begin{aligned} V(u^*|X) &= V(\Omega^{-\frac{1}{2}}u|X) = \Omega^{-\frac{1}{2}}V(u|X)\Omega^{-\frac{1}{2}\prime} = \Omega^{-\frac{1}{2}}(\sigma^2\Omega)\Omega^{-\frac{1}{2}} \\ &= \sigma^2\Omega^{-\frac{1}{2}}\Omega^{\frac{1}{2}}\Omega^{\frac{1}{2}}\Omega^{-\frac{1}{2}} = \sigma^2 I \end{aligned}$$

so that, as X^* is a function of X, by the law of iterated expectations[14], we also have:

$$E(u^*|X^*) = E\left[E(u^*|X)|X^*\right] = E\left[0|X^*\right] = 0$$

and:

$$\begin{aligned} V(u^*|X^*) &= E(u^*u^{*\prime}|X^*) = E\left[E(u^*u^{*\prime}|X)|X^*\right] \\ &= E\left[V(u^*|X)|X^*\right] = E\left[\sigma^2 I|X^*\right] = \sigma^2 I \end{aligned}$$

In words, the transformed model (9) actually satisfies the usual Gauss-Markov assumptions E.1 – E.4, so that, from the Gauss-Markov theorem (Property 3 in SLN-I), the OLS estimator applied to transformed model (9) is the best linear estimator of β under assumptions E.1 – E.3 and assumption E.4bis in the original model. The OLS estimator of the transformed model (9) is called the generalized least squares (GLS) estimator of β. It can be written as[15]:

$$\hat{\beta}_{GLS} = (X^{*\prime}X^*)^{-1} X^{*\prime}Y^* = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y$$

---

[12] Because Ω is symmetric positive definite, it may be shown that $\Omega^{\frac{1}{2}}$ always exists and is unique.

[13] If X is full rank (i.e., $\text{rank}(X) = k$), because $\Omega^{-\frac{1}{2}}$ is non-singular, then X^* is also full rank.

[14] In its most general form, the law of iterated expectations states that, if x is a function of w, then $E(y|x) = E\left[E(y|w)|x\right]$. See Wooldridge (2010) p. 19 for details.

[15] Note that by definition: $\hat{\beta}_{GLS} = \text{Argmin}_\beta (Y^* - X^*\beta)'(Y^* - X^*\beta) = \text{Argmin}_\beta (Y - X\beta)'\Omega^{-1}(Y - X\beta)$.

and its variance-covariance matrix can accordingly be written as :

$$V(\hat{\beta}|X) = \sigma^2 \left(X^{*\prime}X^*\right)^{-1} = \sigma^2 \left(X'\Omega^{-1}X\right)^{-1}$$

- We have just established the following property[16] :

  Property 11 The generalized least squares of β
  Under assumptions E.1 − E.3 and assumption E.4bis, the best linear unbiased estimator (BLUE) of β is the GLS estimator :

  $$\hat{\beta}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y$$

  and its the variance-covariance matrix is given by $V(\hat{\beta}|X) = \sigma^2 \left(X'\Omega^{-1}X\right)^{-1}$.

3.2. Weighted least squares estimation when the heteroskedasticity is known up to a multiplicative constant

- The generalized least squares theory developed above may readily be used to address our question of finding a more efficient estimator than the OLS estimator which explicitly takes into account the presence of heteroskedasticity, i.e., which yield a more efficient estimator than the OLS estimator when the seminal homoskedasticity assumption MLR.5 is violated.

- To replace the seminal homoskedasticity assumption MLR.5, we consider the following assumption, which assumes that the conditional variance is known up to a multiplicative constant :

  MLR.5' Heteroskedasticity known up to a multiplicative constant
  The variance of u given any values of $(x_2, ..., x_k)$, and equivalently the variance of y given any values of $(x_2, ..., x_k)$, are given by :

  $$Var(u|x_2, ..., x_k) = \sigma^2 h(x_2, ..., x_k) \;\Leftrightarrow\; Var(y|x_2, ..., x_k) = \sigma^2 h(x_2, ..., x_k)$$

  where σ^2 is an unknown parameter and $h(x_2, ..., x_k)$ is known and strictly positive for any value of $(x_2, ..., x_k)$.

  In matrix form, Assumption MLR.5' can be rewritten as follows :

  E.4' Heteroskedasticity & no correlation : $V(u|X) = \sigma^2\Omega \Leftrightarrow V(Y|X) = \sigma^2\Omega$

  where Ω is here a diagonal $n \times n$ matrix with diagonal elements equal to $h_i = h(x_{i2}, ..., x_{ik})$, $i = 1, ..., n$:

  $$\Omega = \begin{bmatrix} h_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & h_n \end{bmatrix}$$

[16] For more details, see Hayashi (2000) p. 54-57.

so that :

$$\Omega^{-1} = \begin{bmatrix} \frac{1}{h_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{h_n} \end{bmatrix} \quad \text{and} \quad \Omega^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{h_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{h_n}} \end{bmatrix}$$

- From Property 11, under assumptions E.1 – E.3 and assumption E.4', the best linear unbiased estimator (BLUE) of β is given by the special case of the GLS estimator, called the weighted least squares (WLS) estimator[17] :

$$\begin{aligned} \hat{\beta}_{WLS} &= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y \\ &= \left(\sum_{i=1}^{n} \frac{1}{h_i} X_i' X_i\right)^{-1} \sum_{i=1}^{n} \frac{1}{h_i} X_i' y_i \end{aligned} \tag{10}$$

  which by definition is the OLS estimator of the transformed model :

$$Y^* = X^*\beta + u^*, \quad \text{where } Y^* = \Omega^{-\frac{1}{2}}Y \text{ and } X^* = \Omega^{-\frac{1}{2}}X$$

  i.e., in detailed form :

$$y_i^* = \beta_1 x_{i1}^* + \beta_2 x_{i2}^* + ... + \beta_k x_{ik}^* + u_i^*, \quad i = 1, ..., n \tag{11}$$

  where the transformed variables are[18] :

$$y_i^* = \tfrac{y_i}{\sqrt{h_i}}, \ \ x_{i1}^* = \tfrac{1}{\sqrt{h_i}}, \ \ x_{i2}^* = \tfrac{x_{i2}}{\sqrt{h_i}}, ..., \ \ x_{ik}^* = \tfrac{x_{ik}}{\sqrt{h_i}}, \quad i = 1, ..., n \tag{12}$$

- As the weighted least squares estimator (10) is nothing but the OLS estimator of the linear regression (11) with the variables transformed as indicated in (12), it has all the usual finite sample and asymptotic properties associated with standard OLS estimation. More specifically :
  - Under the full set of seminal assumptions MLR.1 – MLR.4, assumption MLR.5' and an appropriately restated normality assumption[19] MLR.6', which implies the full set of assumptions E.1 – E.3, assumption E.4' and an appropriately restated normality assumption[20] E.5', the transformed model (11) satisfies the full set of usual assumptions[21] MLR.1 – MLR.6 and E.1 – E.5, so that the WLS estimator $\hat{\beta}_{WLS}$ is both the best linear unbiased estimator and a consistent estimator for β, and all the usual inference procedures[22] computed after OLS estimation of the transformed

---

[17] Note that by definition : $\hat{\beta}_{WLS} = \text{Argmin}_\beta (Y - X\beta)'\Omega^{-1}(Y - X\beta) = \text{Argmin}_\beta \sum_{i=1}^{n} \frac{1}{h_i}(y_i - X_i\beta)^2$, hence the name 'weighted least squares'.

[18] Note that the variable corresponding to the intercept (i.e., x_{i1}^*) is no longer equal to 1, but equal to $1/\sqrt{h_i}$.

[19] MLR.6' Normality
The distribution of u given any values of $(x_2, ..., x_k)$, and equivalently the distribution of y given any value of $(x_2, ..., x_k)$, are normal and given by : $u|x_2,..., x_k \sim N(0, \sigma^2 h(x_2,..., x_k)) \Leftrightarrow y|x_2,..., x_k \sim N(\beta_1 + \beta_2 x_2 + ... + \beta_k x_k, \sigma^2 h(x_2,..., x_k))$.

[20] E.5' Normality : $u|X \sim N(0, \sigma^2\Omega) \Leftrightarrow Y|X \sim N(X\beta, \sigma^2\Omega)$.

[21] Normality follows from the fact that u^* (resp. Y^*) is a linear function of u (resp. Y).

[22] i.e., confidence intervals for β_j or a single linear combination $R_0\beta$, two-sided and one-sided t-tests of β_j

model (11) are valid and exact in finite sample.

– Under the seminal assumptions MLR.1 – MLR.4 and assumption MLR.5', which implies assumptions E.1 – E.3 and assumption E.4', i.e., without having to rely on any normality assumption, the transformed model (11) satisfies the usual seminal assumptions MLR.1 – MLR.5 and assumptions E.1 – E.4, so that the WLS estimator $\hat{\beta}_{WLS}$ is still both the best linear unbiased estimator and a consistent estimator for β, and likewise all usual inference procedures computed after OLS estimation of the transformed model (11) are still valid, but only asymptotically, i.e., are still approximately valid for n sufficiently large.

– Finally, and importantly, under the seminal assumptions MLR.1 – MLR.4, which implies assumptions E.1 – E.3, i.e., without assuming neither that the heteroskedasticity known up to a multiplicative constant assumption is correct nor that any normality assumption holds, the transformed model (11) still satisfies the seminal assumptions MLR.1 – MLR.4 and assumptions E.1 – E.3, so that the WLS estimator $\hat{\beta}_{WLS}$ is still unbiased – but no longer BLUE – and consistent for β, and asymptotically valid – i.e., approximately valid for n sufficiently large – inference procedures are provided by the heteroskedasticity robust inference procedures outlined above in Section 2 computed after OLS estimation of the transformed model (11).

• Remarks :

– Multiplying the weights $w_i = 1/h_i$ – which are inversely proportional to the (conditional) variance of y_i – used by the WLS estimator $\hat{\beta}_{WLS}$ by any positive constant has no effect on both the value of $\hat{\beta}_{WLS}$ and the outcome of the (usual or heteroskedasticity robust) inference procedures computed after OLS estimation of the transformed model (11). This means in particular that h_i may always be replaced by $h_i^* = ah_i$, where a is any positive constant, without affecting neither the value of the WLS estimator $\hat{\beta}_{WLS}$ nor the outcome of the (usual or heteroskedasticity robust) inference procedures.

– All modern econometric software provide built-in routines for performing WLS estimation and inference, so that in practice it is not required to transform the variables and run an OLS regression using the transformed variables to perform WLS estimation and inference.

– In practice, there are few situations where the heteroskedasticity is known up to a multiplicative constant. This typically happens when, instead of using individual-level data, we only have averages of data, or per capita data, across some group or geographical region (see Wooldridge (2016) p. 258 for details).

---

or a single linear combination $R_0\beta$, F-test of multiple linear restrictions.

3.3. Weighted least squares estimation when the heteroskedasticity function must be estimated

- When, as it is usually the case, the heteroskedasticity function is not known, it may be estimated from the data. This requires to specify a model for the conditional variance $Var(u|x_2, ..., x_k) = Var(y|x_2, ..., x_k)$.

- A popular and fairly flexible model for the conditional variance is given by the so-called multiplicative heteroskedasticity model :

  MLR.5" Multiplicative heteroskedasticity

  The variance of u given any values of $(x_2, ..., x_k)$, and equivalently the variance of y given any values of $(x_2, ..., x_k)$, are given by :

$$Var(u|x_2, ..., x_k) = \exp(\delta_1 + \delta_2 x_2 + ... + \delta_k x_k)$$

$$\Leftrightarrow \; Var(y|x_2, ..., x_k) = \exp(\delta_1 + \delta_2 x_2 + ... + \delta_k x_k)$$

  where $(\delta_1, ..., \delta_k)$ are unknown parameters.

- Assumption MLR.5" is just a special case of assumption MLR.5' where :

$$\sigma^2 = \exp(\delta_1) \quad \text{and} \quad h(x_2, ..., x_k) = \exp(\delta_2 x_2 + ... + \delta_k x_k)$$

  If the parameters $(\delta_2, ..., \delta_k)$ were known, we could just apply WLS, as outlined in the previous section. When, as it is usually the case, $(\delta_2, ..., \delta_k)$ are not known, they may be consistently estimated from the data under mild conditions.

  Let for conciseness $x = (x_2, ..., x_k)$ denote the vector of explanatory variables $(x_2, ..., x_k)$ and $\sigma^2_{u|x} = Var(u|x)$ the conditional variance $Var(u|x) = E(u^2|x)$. Under assumption MLR.5", we can write :

$$u^2 = \exp(\delta_1 + \delta_2 x_2 + ... + \delta_k x_k) \cdot v$$

  where by construction $v = \frac{u^2}{\sigma^2_{u|x}}$ and $E(v|x) = \frac{1}{\sigma^2_{u|x}} E(u^2|x) = \frac{\sigma^2_{u|x}}{\sigma^2_{u|x}} = 1$. Taking the log, we can further write :

$$\log(u^2) = \delta_1 + \delta_2 x_2 + ... + \delta_k x_k + r$$

  where $r = \log(v)$. If v is independent[23] of x, then $r = \log(v)$ is also independent of x, and $E(r|x)$ is constant (i.e., does not depend of x), say $E(r|x) = \mu_r$. Letting $e = r - \mu_r$ denote the centered value of r, we can finally write :

$$\log(u^2) = \alpha_1 + \delta_2 x_2 + ... + \delta_k x_k + e \qquad (13)$$

  where by construction $\alpha_1 = \delta_1 + \mu_r$ and $E(e|x_2, ..., x_k) = 0$.

[23] It is for example the case under normality, i.e., if $u|x \sim N(0, \sigma^2_{u|x})$, so that $\frac{u}{\sigma_{u|x}} \sim N(0,1)$. It will likewise be the case whenever the standardized error $\frac{u}{\sigma_{u|x}}$ is distributed according to a distribution which does not depend on x, so that $v = (\frac{u}{\sigma_{u|x}})^2$ is also distributed according to a distribution which does not depend on x (and is thus independent of x).

Model (13) satisfies the usual Gauss-Markov assumptions[24], so that a regression (including an intercept) of $\log(u_i^2)$ on $x_{i2}, ..., x_{ik}$ would provide unbiased and consistent estimators of $(\alpha_1, \delta_2, ..., \delta_k)$. It may be shown that if the unknown errors u_i are replaced by their consistent estimates $\hat{u}_i = y_i - X_i\hat{\beta}$ obtained from the OLS estimation of β, then the regression (including an intercept) of:

$$\log(\hat{u}_i^2) \text{ on } x_{i2}, ..., x_{ik} \tag{14}$$

likewise provides consistent (but no longer unbiased) estimators $(\hat{\alpha}_1, \hat{\delta}_2, ..., \hat{\delta}_k)$ of $(\alpha_1, \delta_2, ..., \delta_k)$.

- Based on the above estimators of the multiplicative heteroskedasticity parameters, from (10), a feasible weighted least squares estimator – more usually called feasible generalized least squares (FGLS) estimator – of β is given by:

$$\hat{\beta}_{FGLS} = \left(\sum_{i=1}^{n} \frac{1}{\hat{h}_i} X_i' X_i\right)^{-1} \sum_{i=1}^{n} \frac{1}{\hat{h}_i} X_i' y_i$$

which by definition is the OLS estimator of the transformed model:

$$\check{y}_i^* = \beta_1 \check{x}_{i1}^* + \beta_2 \check{x}_{i2}^* + ... + \beta_k \check{x}_{ik}^* + u_i^*, \quad i = 1, ..., n \tag{15}$$

where:

$$\check{y}_i^* = \frac{y_i}{\sqrt{\hat{h}_i}}, \quad \check{x}_{i1}^* = \frac{1}{\sqrt{\hat{h}_i}}, \quad \check{x}_{i2}^* = \frac{x_{i2}}{\sqrt{\hat{h}_i}}, ..., \quad \check{x}_{ik}^* = \frac{x_{ik}}{\sqrt{\hat{h}_i}}, \quad i = 1, ..., n$$

and, because $\hat{h}_i$ may be multiplied by any positive constant without affecting neither the value of the estimator nor the outcome of the (usual or heteroskedasticity robust) inference procedures, $\hat{h}_i$ may simply be taken as:

$$\hat{h}_i = \exp(\hat{g}_i), \quad \text{where } \hat{g}_i = \hat{\alpha}_1 + \hat{\delta}_2 x_2 + ... + \hat{\delta}_k x_k$$

i.e., as the exponential of the fitted values $\hat{g}_i$ of the OLS regression (14).

- It may be shown that the replacement of h_i by its estimator $\hat{h}_i$ (up to a multiplicative constant), which turns the WLS estimator into a feasible estimator, does not change its asymptotic properties[25]. More specifically:

  – Under the seminal assumptions MLR.1 – MLR.4 and assumption MLR.5", the FGLS estimator $\hat{\beta}_{FGLS}$ is consistent for β and asymptotically more efficient than the OLS estimator[26], and all usual inference procedures computed after OLS estimation of the transformed model (15) are still asymptotically valid, i.e., approximately valid for n sufficiently large.

  – Likewise, under only the seminal assumptions MLR.1 – MLR.4, i.e., without assuming that the multiplicative heteroskedasticity assumption MLR.5" is correct, both the FGLS estimator $\hat{\beta}_{FGLS}$ is still consistent for β and the

[24] If r is independent of x, not only $E(r|x)$, but also $Var(r|x)$ is constant (i.e., does not depend of x), so that not only $E(e|x) = 0$, but also $Var(e|x)$ is constant.

[25] For a discussion, see Hayashi (2000), p. 133-137.

[26] This is just the asymptotic analog of the BLUE property of the WLS estimator.

heteroskedasticity robust inference procedures outlined above in Section 2 computed after OLS estimation of the transformed model (15) are still asymptotically valid, i.e., approximately valid for n sufficiently large.

- Remarks :
  - Due to the estimation of the heteroskedasticity function, the FGLS estimator $\hat{\beta}_{FGLS}$ has no longer exact in finite sample properties : it is not unbiased, and the usual inference procedures computed after OLS estimation of the transformed model (15) are not exact even if an appropriately stated normality assumption is assumed to hold.
  - The computation of the FGLS estimator $\hat{\beta}_{FGLS}$ entails three steps : (1) preliminary estimation of the model by OLS to obtain the residuals $\hat{u}_i = y_i - X_i\hat{\beta}$, (2) estimation of the conditional variance parameters by OLS through the regression (14) to obtain the estimated variances (up to a multiplicative constant) $\hat{h}_i$, (3) re-estimation of the model by the FGLS estimator $\hat{\beta}_{FGLS}$, either through the OLS estimation of the transformed model (15) or using the built-in routines for performing WLS provided by modern econometric software. Note that if there is a very large difference between the estimated value of β in step (1) and step (3), we should be suspicious : this may indicate a functional misspecification of the conditional mean – i.e., $E(y|x_2, ..., x_k)$ – of the model (see Wooldridge (2016) p. 262 for details).
  - If the assumed form of heteroskedasticity is not correct, although still consistent[27], the FGLS estimator is no longer necessarily more efficient than the OLS estimator. This is the same as when h_i is supposed to be known : if h_i is actually not correct, although still unbiased and consistent, the WLS estimator is no longer BLUE, and thus not necessarily better than the OLS estimator. In practice however, in cases of strong heteroskedasticity, even if the specification of the heteroskedasticity is roughly approximate, applying FGLS / WLS along with heteroskedasticity robust inference procedures will often yield a more precise estimate of β than ignoring heteroskedasticity altogether and using OLS likewise along with heteroskedasticity robust inference procedures (see Wooldridge (2016) p. 263-264 for details).

Reference

Hayashi F. (2000), *Econometrics*, Princeton University Press.

Wooldridge J.M. (2010), *Econometric Analysis of Cross-Section and Panel Data*, Second Edition, MIT Press.

Wooldridge J.M. (2016), *Introductory Econometrics : A Modern Approach*, 6th Edition, Cengage Learning.

[27] provided of course that assumptions MLR.1 – MLR.4 hold.