Contents

1 Alternatives to LS

1.1 Generalized Least Square (GLS)

1.1.1 Definition

Definition (Generalized Least Squares). From Alexander Aitkin 1935, Let a model $y_i = \alpha + \beta x_i + \varepsilon_i$ we are going to relax two assumptions of the Markov-Gauss theorem:

$$\begin{aligned} &\text{(Serial correlation)} \quad cov(\varepsilon_i, \varepsilon_j) \neq 0 \\ &\text{(Heteroskedastic errors)} \quad Var(\varepsilon_i \mid x_i) = f(x_i) \end{aligned}$$

The GLS framework enables to take into account those alteration of the model for inference: estimating the α and β.

The basic idea of GLS is to normalize our data using the known covariance such that the normalized data adheres to the OLS assumptions. We can then apply the OLS estimator, which is BLUE, to these transformed data.

In generalized least squares, we assume the following model:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \\ \text{(Strong exogeneity)} \quad \mathbb{E}[\varepsilon \mid \mathbf{X}] &= \mathbf{0}, \\ \text{(Heteroskedastic errors)} \quad \mathbb{V}[\varepsilon \mid \mathbf{X}] &= \sigma^2 \boldsymbol{\Omega} = \boldsymbol{\Sigma}. \end{aligned}$$

Here, $\boldsymbol{\Omega}$ is a positive definite matrix, and we assume it is known. This is called generalized least squares because OLS is just a special case of this model, when $\boldsymbol{\Omega} = \mathbf{I}$. Note recollect that the Homoskedastic errors states that there is no covariance $\mathbb{V}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \sigma^2 \mathbf{I}$.

Algebric definition. The GLS objective is to estimate linear coefficients β that minimizes the SSR. It is defined as a Gaussian Kernel (or Mahalanobis distance)

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{GLS}} &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \ ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||_{\Omega^{-1}} \\ &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \end{aligned}$$

Where $\Omega_{n \times n}$ is a <u>covariance matrix of the error term</u> $\mathbb{V}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \sigma^2 \boldsymbol{\Omega}$.

- GLS is homoscedastic. Using the property that Ω is psd, it has a matrix inverse which a Cholesky decomposition of the form

$$\boldsymbol{\Omega}^{-1} = \mathbf{L}\mathbf{L}^\top.$$

And we can rewrite the Gaussian kernel in Equation 2 as

$$\begin{aligned} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{L}\mathbf{L}^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \left(\mathbf{L}^\top \mathbf{y} - \mathbf{L}^\top \mathbf{X}\boldsymbol{\beta}\right)^\top \left(\mathbf{L}^\top \mathbf{y} - \mathbf{L}^\top \mathbf{X}\boldsymbol{\beta}\right). \end{aligned}$$

In other words, the objective with transformed data $\mathbf{L}^\top \mathbf{y}$ and $\mathbf{L}^\top \mathbf{X}$ is just the OLS objective, and the transformed errors are homoscedastic,

$$\begin{aligned} \mathbb{E}\left[\mathbf{L}^\top \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top \mathbf{L} \mid \mathbf{L}^\top \mathbf{X}\right] &= \mathbf{L}^\top \mathbb{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top \mid \mathbf{L}^\top \mathbf{X}\right] \mathbf{L} \\ &= \mathbf{L}^\top \left(\sigma^2 \boldsymbol{\Omega}\right) \mathbf{L} \\ &= \sigma^2 \mathbf{I}. \end{aligned}$$

Why does this work? Since the univariate analog to $\mathbf{\Sigma}$ is σ^2, then $\mathbf{L}^\top$ is the multivariate analog to $1/\sigma$. In other words, $\mathbf{L}^\top$ represents a linear transformation to our data such that the new data's variance is normalized. Furthermore, note that $\mathbf{L}^\top\mathbf{y}$ and $\mathbf{L}^\top\mathbf{X}$ adhere to the standard assumptions of OLS : $\mathbf{L}^\top\mathbf{X}$ is nonrandom since $\mathbf{X}$ is non-random and $\mathbf{L}$ is non-random (since $\mathbf{\Omega}$ is known), $\mathbf{L}^\top\mathbf{X}$ is full-rank since $\mathbf{X}$ is full-rank, and the conditional mean of our observations $\mathbf{L}^\top\mathbf{y}$ is what we would expect, i.e.

$$\mathbb{E}\left[\mathbf{L}^\top\mathbf{y} \mid \mathbf{L}^\top\mathbf{X}\right] = \mathbf{L}^\top\mathbf{X}\boldsymbol{\beta}$$

So GLS is OLS but with $\mathbf{y}$ and $\mathbf{X}$ mapped by the linear function $\mathbf{L}^\top$. To make this point clear, it is common to use the following notation:

$$\mathbf{y}_* = \mathbf{L}^\top\mathbf{y}$$
$$\mathbf{X}_* = \mathbf{L}^\top\mathbf{X}$$

and to rewrite the GLS objective in Equation 2 as

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{(\mathbf{y}_* - \mathbf{X}_*\boldsymbol{\beta})^\top (\mathbf{y}_* - \mathbf{X}_*\boldsymbol{\beta})\right\}.$$

Since the OLS estimator,

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = \left(\mathbf{X}_*^\top\mathbf{X}_*\right)^{-1}\mathbf{X}_*^\top\mathbf{y}_*$$

is BLUE, this implies the GLS estimator, which can just Equation 10 but un-transformed, is also BLUE:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\text{GLS}} &= \left(\mathbf{X}^\top\mathbf{L}\mathbf{L}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{L}\mathbf{L}^\top\mathbf{y} \\ &= \left(\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{y}.\end{aligned}$$

So the GLS estimator is BLUE in the presence of heteroscedastic errors, governed by $\mathbf{\Omega}$. The trick was to normalize the data by the known covariance and then apply the OLS estimator to these transformed data.

- Special case: OLS. If we let $\mathbf{\Omega}^{-1} = I$ we get back the OLS

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \; ||\mathbf{X}\boldsymbol{\beta} - \boldsymbol{y}||$$

Since $e^Te = ||e||^2$.

- Special case: WLS. It assumes heteroscedasticity but with <u>uncorrelated errors</u>, i.e. the cross-covariance terms in $\mathbf{\Omega}$ are zero (the matrix is diagonal). Here, each observation is assigned a weight w_n that scales the squared residual error. When $w_n = 1/\sigma_n^2$

$$\hat{\boldsymbol{\beta}}_{\text{WLS}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top\boldsymbol{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

Alternatively we can rewrite it in simple LR format:

$$J(\boldsymbol{\beta}) = \sum_{n=1}^{N} w_n \left(y_n - \mathbf{x}_n^\top\boldsymbol{\beta}\right)^2$$

Thus if we know σ_n^2 for each sample and use weights that are the reciprocal of the variances, the WLS is the best linear unbiased estimator (BLUE), since the GLS estimator is BLUE.

1.1.2 Weighted Least Squares

Definition. WLS is a subcase of the GLS where violate solely the Homoskedastic hypothesis of the Gauss-Markov theorem. That is alternatively we define the

$$\begin{aligned}&\text{(Strong exogeneity)}\quad \mathbb{E}[\varepsilon \mid \mathbf{X}] = \mathbf{0},\\ &\text{(Heteroskedastic errors)}\quad Var(\varepsilon_i \mid x_i) = f(x_i)\end{aligned}$$

Definition.

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\text{WLS}} &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\ ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||_W \\ &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{e^T W e\right\} \\ &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{\sum_{n=1}^{N} w_n \left(y_n - \mathbf{x}_n^\top \boldsymbol{\beta}\right)^2\right\}.\end{aligned}$$

We can visual the weight matrix as, note that when $w_n = 1/\sigma_n^2$ we get a special case of GLS with uncorrelated errors.

$$W = \begin{bmatrix} w_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & w_n \end{bmatrix}; \quad w_i > 0, \forall i$$

We can turn this problem into an OLS by performing the steps:

$$\begin{aligned}\{e^T W e\} &= \{e^T \sqrt{W} \sqrt{W} e\} \\ &= \{(\sqrt{W} e)^T (\sqrt{W} e)\} \\ &= \{\widetilde{e}^T \widetilde{e}\} = ||\widetilde{e}||^2 \\ &= ||\underbrace{\sqrt{W} X}_{\widetilde{X}} \beta - \underbrace{\sqrt{W} y}_{\widetilde{y}}||^2\end{aligned}$$

Thus we are back to our original

$$\begin{aligned}\hat{\beta}_{GLS} &= (\widetilde{X}^T \widetilde{X})^{-1} \widetilde{X}^T y \\ &= \left[(\sqrt{W} X)^\top (\sqrt{W} X)\right]^{-1} \cdot (\sqrt{W} X)^\top \cdot \sqrt{W} y \\ &= \boxed{\left(X^\top W X\right)^{-1} \cdot X^\top W y = \hat{\beta}_{GLS}}\end{aligned}$$

1.2 Régression quadratique avec moindre carrés

The height of the rocket at fixed times is recorded in the Table. The time t is recorded in seconds, the heights y in meters.

t	5	10	15	20	25	30
y	722	1073	1178	1117	781	102

After looking at the data we propose to fit the equation $\boxed{y = a + bt + ct^2}$ which in matrix form:

$$\begin{pmatrix} 1 & 5 & (5)^2 \\ 1 & 10 & (10)^2 \\ 1 & 15 & (15)^2 \\ 1 & 20 & (20)^2 \\ 1 & 25 & (25)^2 \\ 1 & 30 & (30)^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 722 \\ 1073 \\ 1178 \\ 1117 \\ 781 \\ 102 \end{pmatrix} \iff X\beta = y$$

Solving the problem by the usual LS estimator we obtain:

$$\hat{\beta} = (X^T X)^{-1} X^T y \iff \hat{\beta} = (80.2, 149.78, -4.9385)^T$$

That we can plot:

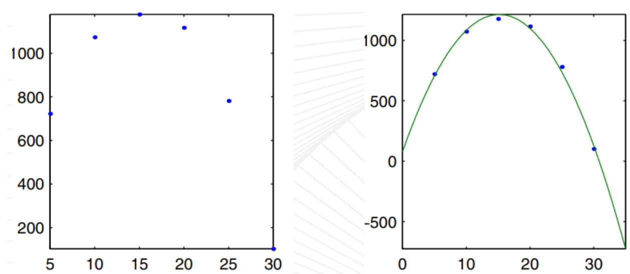

Figure 1: Fitting a parabola to the rocket data.

Bibliographie:

- best course

1.3 Restricted Least Squares

From the book An introduction to classsical econometric theory by Ruud et al. 2000. Recollect that we solve:

$$\beta_{OLS} = \underset{\beta \in B}{argmin} \; ||y - X\beta||_2^2$$

In the restricted least-square setting we are interested in adding constraints to the β's. Such restriction can always be written in the form:

$$\underset{K\times 1}{\boldsymbol{\beta}} = \underset{K\times M}{\mathbf{S}} \underset{M\times 1}{\boldsymbol{\gamma}} + \underset{K\times 1}{\mathbf{s}}$$

- Example (Exclusion Restrictions). The simplest linear restrictions are exclusion, or "zero" restrictions. In this case $\mathbf{s} = 0$. If we restrict the coefficient of a RHS variable to zero, then we are excluding it from the OLS fit. Suppose that $K = 5$ and $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_K]$ so that by partition LS:

$$\mathbf{X}\boldsymbol{\beta} \equiv \beta_1 + \mathbf{X}_2\beta_2 + \mathbf{X}_3\beta_3 + \mathbf{X}_4\beta_4 + \mathbf{X}_5\beta_5$$

The restrictions

$$\beta_2 = 0; \quad \beta_4 = 0$$

simplify the RHS to

$$\begin{aligned} \mathbf{X}\beta &= \beta_1 + 0 \cdot \mathbf{X}_2 + \mathbf{X}_3\beta_3 + 0 \cdot \mathbf{X}_4 + \mathbf{X}_5\beta_5 \\ &= \beta_1 + \mathbf{X}_3\beta_3 + \mathbf{X}_5\beta_5 \end{aligned}$$

thereby excluding x_2 and x_4. We can write

$$\underset{K\times 1}{\boldsymbol{\beta}} = \underset{K\times M}{\mathbf{S}} \underset{M\times 1}{\boldsymbol{\gamma}} + \underset{K\times 1}{\mathbf{s}} \iff \beta = \begin{bmatrix} \beta_1 \\ 0 \\ \beta_3 \\ 0 \\ \beta_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_3 \\ \beta_5 \end{bmatrix}$$

so that two restrictions make five parameters a linear function of only three. In our notation,

$$\boldsymbol{\gamma} \equiv \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_3 \\ \beta_5 \end{bmatrix}$$

RLS coefficient. The RLS coefficient vector is :

$$\begin{aligned} \hat{\beta}_R &= \underset{\{\beta | \beta = \mathbf{S}\mathbf{y} + \mathbf{s}\}}{argmin} \; ||\mathbf{y} - \mathbf{X}\beta||^2 \\ &= \mathbf{S}\left[(\mathbf{XS})'\mathbf{XS}\right]^{-1}(\mathbf{XS})'(y - \mathbf{XS}) + \mathbf{s} \end{aligned}$$

- Proof. We prove this proposition by substituting the restrictions direclly into the objcctive function and deriving an unconstrained minimum in $\boldsymbol{\gamma}$. The restrictions imply that

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{X}(\mathbf{S}\boldsymbol{\gamma} + \mathbf{s}) = \mathbf{XS}\boldsymbol{\gamma} + \mathbf{Xs}$$

yielding the unconstrained minimization

$$\hat{\gamma} = \underset{\gamma}{\operatorname{argmin}} \|(\mathbf{y} - \mathbf{Xs}) - (\mathbf{XS})\boldsymbol{\gamma}\|^2$$

By choosing a new matrix of RHS variables $\mathbf{X}_R = \mathbf{XS}$ and a new LHS vector $\mathbf{y}_R = \mathbf{y} - \mathbf{Xs}$, we can find the solution with two equations: the OLS fit of y_R to $\mathbf{X}_R$,

$$
\begin{aligned}
\hat{\gamma} &= \left(\mathbf{X}_{\mathrm{R}}'\mathbf{X}_{\mathrm{R}}\right)^{-1}\mathbf{X}_{\mathrm{R}}'\mathbf{y}_{\mathrm{R}} \\
&= \left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{X}'(\mathbf{y}-\mathbf{X}\mathbf{s})
\end{aligned}
$$

and the restrictions expressing $\boldsymbol{\beta}$ as a function of γ,

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathrm{R}} &= \mathbf{S}\hat{\boldsymbol{\gamma}}+\mathbf{s} \\
&= \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{X}'(\mathbf{y}-\mathbf{X}\mathbf{s})+\mathbf{s}
\end{aligned}
$$

Partition.

$$
\begin{aligned}
\mathbf{y} = \underset{n\times d}{\mathbf{X}}\;\underset{d\times 1}{\beta} &= \begin{pmatrix} X_1 & X_2 \end{pmatrix}\begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix} \\
&= \underset{n\times d_1}{\mathbf{X}_1}\;\underset{d_1\times 1}{\beta_1} + \underset{n\times d_2}{\mathbf{X}_2}\;\underset{d_2\times 1}{\beta_2}
\end{aligned}
$$

For example. We can decompose the full matrix

$$
\mathbf{y} = \begin{array}{c c}
 & \begin{array}{cccc} \text{FR} & \text{UK} & \text{Oil} & \text{Bank} \end{array} \\
\begin{array}{c} \text{asset 1} \\ \text{asset 2} \\ .. \\ \text{asset n} \end{array} & \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}
\end{array}
\begin{pmatrix}\underline{\beta_1} \\ \underline{\beta_2}\end{pmatrix}
$$

$$
= \begin{array}{c c}
 & \begin{array}{cc} \text{FR} & \text{UK} \end{array} \\
\begin{array}{c} \text{asset 1} \\ \text{asset 2} \\ .. \\ \text{asset n} \end{array} & \begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}
\end{array}
\underline{\beta_1} +
\begin{array}{c c}
 & \begin{array}{cc} \text{Oil} & \text{Bank} \end{array} \\
\begin{array}{c} \text{asset 1} \\ \text{asset 2} \\ .. \\ \text{asset n} \end{array} & \begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}
\end{array}
\underline{\beta_2}
$$

1.4 Robust least square