Chapter 4

WLS and Generalized Least Squares

4.1 Random Vectors

The concepts of a random vector, the expected value of a random vector and the covariance of a random vector are needed before covering generalized least squares. Recall that for random variables Y_i and Y_j, the covariance of Y_i and Y_j is $\text{Cov}(Y_i, Y_j) \equiv \sigma_{i,j} = E[(Y_i - E(Y_i))(Y_j - E(Y_j))] = E(Y_i Y_j) - E(Y_i)E(Y_j)$ provided the second moments of Y_i and Y_j exist.

Definition 4.1. $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$ is an $n \times 1$ **random vector** if Y_i is a random variable for $i = 1, ..., n$. $\boldsymbol{Y}$ is a discrete random vector if each Y_i is discrete and $\boldsymbol{Y}$ is a continuous random vector if each Y_i is continuous. A random variable Y_1 is the special case of a random vector with $n = 1$.

Definition 4.2. The *population mean* of a random $n \times 1$ vector $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$ is
$$E(\boldsymbol{Y}) = (E(Y_1), ..., E(Y_n))^T$$
provided that $E(Y_i)$ exists for $i = 1, ..., n$. Otherwise the expected value does not exist. The $n \times n$ *population covariance matrix*
$$\text{Cov}(\boldsymbol{Y}) = E[(\boldsymbol{Y} - E(\boldsymbol{Y}))(\boldsymbol{Y} - E(\boldsymbol{Y}))^T] = ((\sigma_{i,j}))$$
where the ij entry of $\text{Cov}(\boldsymbol{Y})$ is $\text{Cov}(Y_i, Y_j) = \sigma_{i,j}$ provided that each $\sigma_{i,j}$ exists. Otherwise $\text{Cov}(\boldsymbol{Y})$ does not exist.

The covariance matrix is also called the variance–covariance matrix and variance matrix. Sometimes the notation $\mathrm{Var}(\boldsymbol{Y})$ is used. Note that $\mathrm{Cov}(\boldsymbol{Y})$ is a symmetric positive semidefinite matrix. If $\boldsymbol{Z}$ and $\boldsymbol{Y}$ are $n \times 1$ random vectors, $\boldsymbol{a}$ a conformable constant vector and $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable constant matrices, then

$$E(\boldsymbol{a} + \boldsymbol{Y}) = \boldsymbol{a} + E(\boldsymbol{Y}) \text{ and } E(\boldsymbol{Y} + \boldsymbol{Z}) = E(\boldsymbol{Y}) + E(\boldsymbol{Z}) \tag{4.1}$$

and

$$E(\boldsymbol{A}\boldsymbol{Y}) = \boldsymbol{A}E(\boldsymbol{Y}) \text{ and } E(\boldsymbol{A}\boldsymbol{Y}\boldsymbol{B}) = \boldsymbol{A}E(\boldsymbol{Y})\boldsymbol{B}. \tag{4.2}$$

Also

$$\mathrm{Cov}(\boldsymbol{a} + \boldsymbol{A}\boldsymbol{Y}) = \mathrm{Cov}(\boldsymbol{A}\boldsymbol{Y}) = \boldsymbol{A}\mathrm{Cov}(\boldsymbol{Y})\boldsymbol{A}^T. \tag{4.3}$$

Example 4.1. Consider the OLS model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where the e_i are iid with mean 0 and variance σ^2. Then $\boldsymbol{Y}$ and $\boldsymbol{e}$ are random vectors while $\boldsymbol{a} = \boldsymbol{X}\boldsymbol{\beta}$ is a constant vector. Notice that $E(\boldsymbol{e}) = \boldsymbol{0}$. Thus

$$E(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta} + E(\boldsymbol{e}) = \boldsymbol{X}\boldsymbol{\beta}.$$

Since the e_i are iid,

$$\mathrm{Cov}(\boldsymbol{Y}) = \mathrm{Cov}(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_n \tag{4.4}$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. This result makes sense because the Y_i are independent with $Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i$. Hence $\mathrm{VAR}(Y_i) = \mathrm{VAR}(e_i) = \sigma^2$.

Recall that $\hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y}$. Hence

$$E(\hat{\boldsymbol{\beta}}_{OLS}) = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T E(\boldsymbol{Y}) = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{\beta} = \boldsymbol{\beta}.$$

That is, $\hat{\boldsymbol{\beta}}_{OLS}$ is an unbiased estimator of $\boldsymbol{\beta}$. Using (4.3) and (4.4),

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{OLS}) = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \mathrm{Cov}(\boldsymbol{Y}) \boldsymbol{X} (\boldsymbol{X}^T \boldsymbol{X})^{-1}$$

$$= \sigma^2 (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{X} (\boldsymbol{X}^T \boldsymbol{X})^{-1} = \sigma^2 (\boldsymbol{X}^T \boldsymbol{X})^{-1}.$$

Recall that $\hat{\boldsymbol{Y}}_{OLS} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{OLS} = \boldsymbol{X}(\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y} = \boldsymbol{H}\boldsymbol{Y}$. Hence

$$E(\hat{\boldsymbol{Y}}_{OLS}) = \boldsymbol{X}(\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T E(\boldsymbol{Y}) = \boldsymbol{X}(\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta} = E(\boldsymbol{Y}).$$

Using (4.3) and (4.4),

$$\mathrm{Cov}(\hat{\boldsymbol{Y}}_{OLS}) = \boldsymbol{H}\mathrm{Cov}(\boldsymbol{Y})\boldsymbol{H}^T = \sigma^2 \boldsymbol{H}$$

since $\boldsymbol{H}^T = \boldsymbol{H}$ and $\boldsymbol{H}\boldsymbol{H} = \boldsymbol{H}$.

Recall that the vector of residuals $\boldsymbol{r}_{OLS} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}_{OLS}$. Hence $E(\boldsymbol{r}_{OLS}) = E(\boldsymbol{Y}) - E(\hat{\boldsymbol{Y}}_{OLS}) = E(\boldsymbol{Y}) - E(\boldsymbol{Y}) = \boldsymbol{0}$. Using (4.3) and (4.4),

$$\text{Cov}(\hat{\boldsymbol{r}}_{OLS}) = (\boldsymbol{I} - \boldsymbol{H})\text{Cov}(\boldsymbol{Y})(\boldsymbol{I} - \boldsymbol{H})^T = \sigma^2(\boldsymbol{I} - \boldsymbol{H})$$

since $\boldsymbol{I} - \boldsymbol{H}$ is symmetric and idempotent: $(\boldsymbol{I} - \boldsymbol{H})^T = \boldsymbol{I} - \boldsymbol{H}$ and $(\boldsymbol{I} - \boldsymbol{H})(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{I} - \boldsymbol{H}$.

4.2   GLS, WLS and FGLS

Definition 4.3. Suppose that the response variable and at least one of the predictor variables is quantitative. Then the *generalized least squares* (GLS) model is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{4.5}$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors. Also $E(\boldsymbol{e}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{e}) = \sigma^2 \boldsymbol{V}$ where $\boldsymbol{V}$ is a known $n \times n$ positive definite matrix.

Definition 4.4. The *GLS estimator*

$$\hat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{Y}. \tag{4.6}$$

The fitted values are $\hat{\boldsymbol{Y}}_{GLS} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}$.

Definition 4.5. Suppose that the response variable and at least one of the predictor variables is quantitative. Then the *weighted least squares* (WLS) model with weights $w_1, ..., w_n$ is the special case of the GLS model where $\boldsymbol{V}$ is diagonal: $\boldsymbol{V} = \text{diag}(v_1, ..., v_n)$ and $w_i = 1/v_i$. Hence

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{4.7}$$

$E(\boldsymbol{e}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{e}) = \sigma^2 \text{diag}(v_1, ..., v_n) = \sigma^2 \text{diag}(1/w_1, ..., 1/w_n)$.

Definition 4.6. The *WLS estimator*

$$\hat{\boldsymbol{\beta}}_{WLS} = (\boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{Y}. \tag{4.8}$$

The fitted values are $\hat{\boldsymbol{Y}}_{WLS} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{WLS}$.

Definition 4.7. The *feasible generalized least squares* (FGLS) model is the same as the GLS estimator except that $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\theta})$ is a function of an unknown $q \times 1$ vector of parameters $\boldsymbol{\theta}$. Let the estimator of $\boldsymbol{V}$ be $\hat{\boldsymbol{V}} = \boldsymbol{V}(\hat{\boldsymbol{\theta}})$. Then the FGLS estimator

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\boldsymbol{X}^T \hat{\boldsymbol{V}}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \hat{\boldsymbol{V}}^{-1} \boldsymbol{Y}. \tag{4.9}$$

The fitted values are $\hat{\boldsymbol{Y}}_{FGLS} = \boldsymbol{X} \hat{\boldsymbol{\beta}}_{FGLS}$. The *feasible weighted least squares* (FWLS) estimator is the special case of the FGLS estimator where $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\theta})$ is diagonal. Hence the estimated weights $\hat{w}_i = 1/\hat{v}_i = 1/v_i(\hat{\boldsymbol{\theta}})$. The FWLS estimator and fitted values will be denoted by $\hat{\boldsymbol{\beta}}_{FWLS}$ and $\hat{\boldsymbol{Y}}_{FWLS}$, respectively.

Notice that the ordinary least squares (OLS) model is a special case of GLS with $\boldsymbol{V} = \boldsymbol{I}_n$, the $n \times n$ identity matrix. It can be shown that the GLS estimator minimizes the GLS criterion

$$Q_{GLS}(\boldsymbol{\eta}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\eta})^T \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\eta}).$$

Notice that the FGLS and FWLS estimators have $p + q + 1$ unknown parameters. These estimators can perform very poorly if $n < 10(p + q + 1)$.

The GLS and WLS estimators can be found from the OLS regression (without an intercept) of a transformed model. Typically there will be a constant in the model: the first column of $\boldsymbol{X}$ is a vector of ones. Following Seber and Lee (2003, p. 66-68), there is a nonsingular $n \times n$ matrix $\boldsymbol{K}$ such that $\boldsymbol{V} = \boldsymbol{K}\boldsymbol{K}^T$. Let $\boldsymbol{Z} = \boldsymbol{K}^{-1}\boldsymbol{Y}$, $\boldsymbol{U} = \boldsymbol{K}^{-1}\boldsymbol{X}$ and $\boldsymbol{\epsilon} = \boldsymbol{K}^{-1}\boldsymbol{e}$. This method uses the Cholesky decomposition and is numerically unstable.

Proposition 4.1 a)
$$\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{4.10}$$
follows the OLS model since $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n$.

b) The GLS estimator $\hat{\boldsymbol{\beta}}_{GLS}$ can be obtained from the OLS regression (without an intercept) of $\boldsymbol{Z}$ on $\boldsymbol{U}$.

c) For WLS, $Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i$. The corresponding OLS model $\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ is equivalent to $Z_i = \boldsymbol{u}_i^T \boldsymbol{\beta} + \epsilon_i$ for $i = 1, ..., n$ where $\boldsymbol{u}_i^T$ is the ith row of $\boldsymbol{U}$. Then $Z_i = \sqrt{w_i}\, Y_i$ and $\boldsymbol{u}_i = \sqrt{w_i}\, \boldsymbol{x}_i$. Hence $\hat{\boldsymbol{\beta}}_{WLS}$ can be obtained from the OLS regression (without an intercept) of $Z_i = \sqrt{w_i}\, Y_i$ on $\boldsymbol{u}_i = \sqrt{w_i}\, \boldsymbol{x}_i$.

Proof. a) $E(\boldsymbol{\epsilon}) = \boldsymbol{K}^{-1}E(\boldsymbol{e}) = \boldsymbol{0}$ and

$$\mathrm{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{K}^{-1}\mathrm{Cov}(\boldsymbol{e})(\boldsymbol{K}^{-1})^T = \sigma^2 \boldsymbol{K}^{-1}\boldsymbol{V}(\boldsymbol{K}^{-1})^T$$

$$= \sigma^2 \boldsymbol{K}^{-1} \boldsymbol{K} \boldsymbol{K}^T (\boldsymbol{K}^{-1})^T = \sigma^2 \boldsymbol{I}_n.$$

Notice that OLS without an intercept needs to be used since $\boldsymbol{U}$ does not contain a vector of ones. The first column of $\boldsymbol{U}$ is $\boldsymbol{K}^{-1}\boldsymbol{1} \neq \boldsymbol{1}$.

b) Let $\hat{\boldsymbol{\beta}}_{ZU}$ denote the OLS estimator obtained by regressing $\boldsymbol{Z}$ on $\boldsymbol{U}$. Then

$$\hat{\boldsymbol{\beta}}_{ZU} = (\boldsymbol{U}^T\boldsymbol{U})^{-1}\boldsymbol{U}^T\boldsymbol{Z} = (\boldsymbol{X}^T(\boldsymbol{K}^{-1})^T\boldsymbol{K}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{K}^{-1})^T\boldsymbol{K}^{-1}\boldsymbol{Y}$$

and the result follows since $\boldsymbol{V}^{-1} = (\boldsymbol{K}\boldsymbol{K}^T)^{-1} = (\boldsymbol{K}^T)^{-1}\boldsymbol{K}^{-1} = (\boldsymbol{K}^{-1})^T\boldsymbol{K}^{-1}$.

c) The result follows from b) if $Z_i = \sqrt{w_i}\, Y_i$ and $\boldsymbol{u}_i = \sqrt{w_i}\, \boldsymbol{x}_i$. But for WLS, $\boldsymbol{V} = \text{diag}(v_1, ..., v_n)$ and hence $\boldsymbol{K} = \boldsymbol{K}^T = \text{diag}(\sqrt{v_1}, ..., \sqrt{v_n})$. Hence

$$\boldsymbol{K}^{-1} = \text{diag}(1/\sqrt{v_1}, ..., 1/\sqrt{v_n}) = \text{diag}(\sqrt{w_1}, ..., \sqrt{w_n})$$

and $\boldsymbol{Z} = \boldsymbol{K}^{-1}\boldsymbol{Y}$ has ith element $Z_i = \sqrt{w_i}\, Y_i$. Similarly, $\boldsymbol{U} = \boldsymbol{K}^{-1}\boldsymbol{X}$ has ith row $\boldsymbol{u}_i^T = \sqrt{w_i}\, \boldsymbol{x}_i^T$.   QED

Following Johnson and Wichern (1988, p. 51) and Freedman (2005, p. 54), there is a symmetric, nonsingular $n \times n$ matrix $\boldsymbol{R}$ such that $\boldsymbol{V} = \boldsymbol{R}\boldsymbol{R}$. Let $\boldsymbol{Z} = \boldsymbol{R}^{-1}\boldsymbol{Y}$, $\boldsymbol{U} = \boldsymbol{R}^{-1}\boldsymbol{X}$ and $\boldsymbol{\epsilon} = \boldsymbol{R}^{-1}\boldsymbol{e}$. This method uses the spectral theorem (singular value decomposition) and has better computational properties than transformation based on the Cholesky decomposition.

Proposition 4.2 a)
$$\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{4.11}$$

follows the OLS model since $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n$.

b) The GLS estimator $\hat{\boldsymbol{\beta}}_{GLS}$ can be obtained from the OLS regression (without an intercept) of $\boldsymbol{Z}$ on $\boldsymbol{U}$.

c) For WLS, $Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$. The corresponding OLS model $\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ is equivalent to $Z_i = \boldsymbol{u}_i^T\boldsymbol{\beta} + \epsilon_i$ for $i = 1, ..., n$ where $\boldsymbol{u}_i^T$ is the ith row of $\boldsymbol{U}$. Then $Z_i = \sqrt{w_i}\, Y_i$ and $\boldsymbol{u}_i = \sqrt{w_i}\, \boldsymbol{x}_i$. Hence $\hat{\boldsymbol{\beta}}_{WLS}$ can be obtained from the OLS regression (without an intercept) of $Z_i = \sqrt{w_i}\, Y_i$ on $\boldsymbol{u}_i = \sqrt{w_i}\, \boldsymbol{x}_i$.

Proof. a) $E(\boldsymbol{\epsilon}) = \boldsymbol{R}^{-1}E(\boldsymbol{e}) = \boldsymbol{0}$ and

$$\text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{R}^{-1}\text{Cov}(\boldsymbol{e})(\boldsymbol{R}^{-1})^T = \sigma^2 \boldsymbol{R}^{-1}\boldsymbol{V}(\boldsymbol{R}^{-1})^T$$

$$= \sigma^2 \boldsymbol{R}^{-1}\boldsymbol{R}\boldsymbol{R}(\boldsymbol{R}^{-1}) = \sigma^2 \boldsymbol{I}_n.$$

Notice that OLS without an intercept needs to be used since $\boldsymbol{U}$ does not contain a vector of ones. The first column of $\boldsymbol{U}$ is $\boldsymbol{R}^{-1}\boldsymbol{1} \neq \boldsymbol{1}$.

b) Let $\hat{\boldsymbol{\beta}}_{ZU}$ denote the OLS estimator obtained by regressing $\boldsymbol{Z}$ on $\boldsymbol{U}$. Then

$$\hat{\boldsymbol{\beta}}_{ZU} = (\boldsymbol{U}^T\boldsymbol{U})^{-1}\boldsymbol{U}^T\boldsymbol{Z} = (\boldsymbol{X}^T(\boldsymbol{R}^{-1})^T\boldsymbol{R}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{R}^{-1})^T\boldsymbol{R}^{-1}\boldsymbol{Y}$$

and the result follows since $\boldsymbol{V}^{-1} = (\boldsymbol{R}\boldsymbol{R})^{-1} = \boldsymbol{R}^{-1}\boldsymbol{R}^{-1} = (\boldsymbol{R}^{-1})^T\boldsymbol{R}^{-1}$.

c) The result follows from b) if $Z_i = \sqrt{w_i}\, Y_i$ and $\boldsymbol{u}_i = \sqrt{w_i}\, \boldsymbol{x}_i$. But for WLS, $\boldsymbol{V} = \text{diag}(v_1, ..., v_n)$ and hence $\boldsymbol{R} = \text{diag}(\sqrt{v_1}, ..., \sqrt{v_n})$. Hence

$$\boldsymbol{R}^{-1} = \text{diag}(1/\sqrt{v_1}, ..., 1/\sqrt{v_n}) = \text{diag}(\sqrt{w_1}, ..., \sqrt{w_n})$$

and $\boldsymbol{Z} = \boldsymbol{R}^{-1}\boldsymbol{Y}$ has ith element $Z_i = \sqrt{w_i}\, Y_i$. Similarly, $\boldsymbol{U} = \boldsymbol{R}^{-1}\boldsymbol{X}$ has ith row $\boldsymbol{u}_i^T = \sqrt{w_i}\, \boldsymbol{x}_i^T$.   QED

Remark 4.1. Standard software produces WLS output and the ANOVA F test and Wald t tests are performed using this output.

Remark 4.2. The FGLS estimator can also be found from the OLS regression (without an intercept) of $\boldsymbol{Z}$ on $\boldsymbol{U}$ where $\boldsymbol{V}(\hat{\boldsymbol{\theta}}) = \boldsymbol{R}\boldsymbol{R}$. Similarly the FWLS estimator can be found from the OLS regression (without an intercept) of $Z_i = \sqrt{\hat{w}_i}Y_i$ on $\boldsymbol{u}_i = \sqrt{\hat{w}_i}\boldsymbol{x}_i$. But now $\boldsymbol{U}$ is a random matrix instead of a constant matrix. Hence these estimators are highly nonlinear. OLS output can be used for exploratory purposes, but the p–values are generally not correct.

Under regularity conditions, the OLS estimator $\hat{\boldsymbol{\beta}}_{OLS}$ is a consistent estimator of $\boldsymbol{\beta}$ when the GLS model holds, but $\hat{\boldsymbol{\beta}}_{GLS}$ should be used because it generally has higher efficiency.

Definition 4.8. Let $\hat{\boldsymbol{\beta}}_{ZU}$ be the OLS estimator from regressing $\boldsymbol{Z}$ on $\boldsymbol{U}$. The vector of fitted values is $\hat{\boldsymbol{Z}} = \boldsymbol{U}\hat{\boldsymbol{\beta}}_{ZU}$ and the vector of residuals is $\boldsymbol{r}_{ZU} = \boldsymbol{Z} - \hat{\boldsymbol{Z}}$. Then $\hat{\boldsymbol{\beta}}_{ZU} = \hat{\boldsymbol{\beta}}_{GLS}$ for GLS, $\hat{\boldsymbol{\beta}}_{ZU} = \hat{\boldsymbol{\beta}}_{FGLS}$ for FGLS, $\hat{\boldsymbol{\beta}}_{ZU} = \hat{\boldsymbol{\beta}}_{WLS}$ for WLS and $\hat{\boldsymbol{\beta}}_{ZU} = \hat{\boldsymbol{\beta}}_{FWLS}$ for FWLS. For GLS, FGLS, WLS and FWLS, a *residual plot* is a plot of $\hat{Z}_i$ versus $r_{ZU,i}$ and a *response plot* is a plot of $\hat{Z}_i$ versus Z_i.

Notice that the residual and response plots are based on the OLS output from the OLS regression without intercept of $\boldsymbol{Z}$ on $\boldsymbol{U}$. If the model is good,
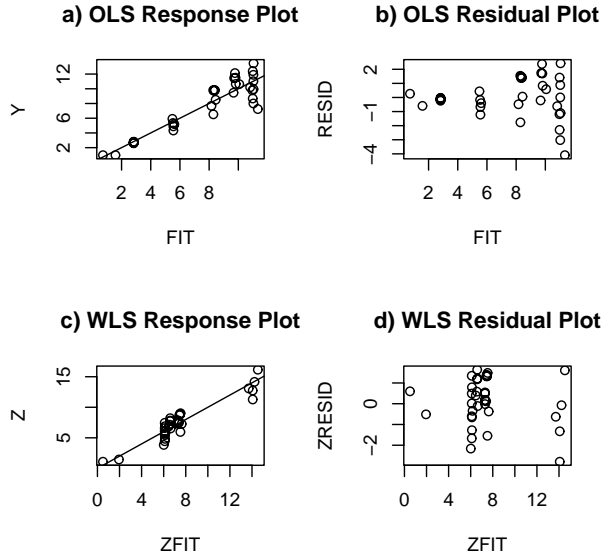
Figure 4.1: Plots for Draper and Smith Data

then the plotted points in the response plot should follow the identity line in an evenly populated band while the plotted points in the residual plot should follow the line $r_{ZU,i} = 0$ in an evenly populated band (at least if the distribution of ϵ is not highly skewed).

Plots based on $\hat{Y}_{GLS} = X\hat{\beta}_{ZU}$ and on $r_{i,GLS} = Y_i - \hat{Y}_{i,GLS}$ should be similar to those based on $\hat{\beta}_{OLS}$. Although the plot of $\hat{Y}_{i,GLS}$ versus Y_i should be linear, the plotted points will not scatter about the identity line in an evenly populated band. Hence this plot can not be used to check whether the GLS model with V is a good approximation to the data. Moreover, the $r_{i,GLS}$ and $\hat{Y}_{i,GLS}$ may be correlated and usually do not scatter about the $r = 0$ line in an evenly populated band. The plots in Definition 4.8 are both a check on linearity and on whether the model using V (or $\hat{V}$) gives a good approximation of the data, provided that $n > k(p + q + 1)$ where $k \geq 5$ and preferably $k \geq 10$.

For GLS and WLS (and for exploratory purposes for FGLS and FWLS), plots and model building and variable selection should be based on Z and U. Form Z and U and then use OLS software for model selection and variable selection. If the columns of X are $x^1, ..., x^p$, then the columns of

U are $U_1, ..., U_p$ where $U_j = \boldsymbol{R}^{-1}\boldsymbol{x}^j$ corresponds to the jth predictor x_j. For example, the analog of the OLS residual plot of jth predictor versus the residuals is the plot of the jth predictor U_j versus r_{ZU}. The notation is confusing but the idea is simple: form $\boldsymbol{Z}$ and $\boldsymbol{U}$, then use OLS software and the OLS techniques from Chapters 2 and 3 to build the model.

Example 4.2. Draper and Smith (1981, p. 112-114) presents a FWLS example with $n = 35$ and $p = 2$. Hence $Y = \beta_1 + \beta_2 x + e$. Let $\hat{v}_i = v_i(\hat{\boldsymbol{\theta}}) = 1.5329 - 0.7334x_i + 0.0883x_i^2$. Thus $\hat{\boldsymbol{\theta}} = (1.5329, -0.7334, 0.0883)^T$. Figure 4.1a and b show the response and residual plots based on the OLS regression of Y on x. The residual plot has the shape of the right opening megaphone, suggesting that the variance is not constant. Figure 4.1c and d show the response and residual plots based on FWLS with weights $\hat{w}_i = 1/\hat{v}_i$. See Problem 4.2 to reproduce these plots. Software meant for WLS needs the weights. Hence FWLS can be computed using WLS software with the estimated weights, but the software may print WLS instead of FWLS, as in Figure 4.1c and d.

Warning. A problem with the response and residual plots for GLS and FGLS given in Definition 4.8 is that some of the transformed cases $(Z_i, \boldsymbol{u}_i^T)^T$ can be outliers or high leverage points.

Remark 4.3. If the response Y_i is the sample mean or sample median of n_i cases where the n_i are not all equal, then use WLS with weights $w_i = n_i$. See Sheather (2009, p. 121).

4.3 Inference for GLS

Inference for the GLS model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ can be performed by using the partial F test for the equivalent no intercept OLS model $\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Following Section 2.10, create $\boldsymbol{Z}$ and $\boldsymbol{U}$, fit the full and reduced model using the "no intercept" or "intercept = F" option.

The 4 step partial F test of hypotheses: i) State the hypotheses Ho: the reduced model is good Ha: use the full model
ii) Find the test statistic $F_R =$

$$\left[\frac{SSE(R) - SSE(F)}{df_R - df_F}\right]/MSE(F)$$

iii) Find the p–value $= P(F_{df_R - df_F, df_F} > F_R)$. (On exams often an F table is used. Here $df_R - df_F = p - q =$ number of parameters set to 0, and $df_F = n - p$.)

iv) State whether you reject Ho or fail to reject Ho. Reject Ho if the p–value $< \delta$ and conclude that the full model should be used. Otherwise, fail to reject Ho and conclude that the reduced model is good.

Assume that the GLS model contains a constant β_1. The GLS ANOVA F test of $Ho : \beta_2 = \cdots = \beta_p$ versus Ha: not Ho uses the reduced model that contains the first column of $\boldsymbol{U}$. The GLS ANOVA F test of $Ho : \beta_i = 0$ versus $Ho : \beta_i \neq 0$ uses the reduced model with the ith column of $\boldsymbol{U}$ deleted. For the special case of WLS, the software will often have a `weights` option that will also give correct output for inference.

Example 4.3. Suppose that the data from Example 4.2 has valid weights, so that WLS can be used instead of FWLS. The *R/Splus* commands below perform WLS.

```
> ls.print(lsfit(dsx,dsy,wt=dsw))
Residual Standard Error=1.137
R-Square=0.9209
F-statistic (df=1, 33)=384.4139
p-value=0

          Estimate Std.Err t-value Pr(>|t|)
Intercept  -0.8891  0.3004 -2.9602   0.0057
X           1.1648  0.0594 19.6065   0.0000
```

Alternative *R/Splus* commands given below produce similar output.

```
zout<-lm(dsy~dsx,weights=dsw)
summary(zout)
anova(zout)
zoutr<-lm(dsy~1,weights=dsw)
anova(zoutr,zout)
```

The F statistic 384.4139 tests $Ho : \beta_2 = 0$ since weights were used. The WLS ANOVA F test for $Ho : \beta_2 = 0$ can also be found with the no intercept model by adding a column of ones to x, form $\boldsymbol{U}$ and $\boldsymbol{Z}$ and compute the partial F test where the reduced model uses the first column of $\boldsymbol{U}$. Notice

189

that the "intercept=F" option needs to be used to fit both models. The residual standard error = RSE = $\sqrt{MSE}$. Thus SSE = $(n - k)(RSE)^2$ where $n - k$ is the denominator degrees of freedom for the F test and k is the numerator degrees of freedom = number of variables in the model. The column of ones *xone* is counted as a variable. The last line of output computes the partial F statistic and is again ≈ 384.4.

```
> xone <- 1 + 0*1:35
>   x <- cbind(xone,dsx)
> z <- as.vector(diag(sqrt(dsw))%*%dsy)
>   u <- diag(sqrt(dsw))%*%x
> ls.print(lsfit(u,z,intercept=F))
Residual Standard Error=1.137
R-Square=0.9817
F-statistic (df=2, 33)=886.4982
p-value=0


      Estimate Std.Err t-value Pr(>|t|)
xone   -0.8891  0.3004 -2.9602   0.0057
dsx     1.1648  0.0594 19.6065   0.0000

> ls.print(lsfit(u[,1],z,intercept=F))
Residual Standard Error=3.9838
R-Square=0.7689
F-statistic (df=1, 34)=113.1055
p-value=0


  Estimate Std.Err t-value Pr(>|t|)
X   4.5024  0.4234 10.6351        0


> ((34*(3.9838)^2-33*(1.137)^2)/1)/(1.137)^2
[1] 384.4006
```

The WLS t-test for this data has $t = 19.6065$ which corresponds to $F = t^2 = 384.4$ since this test is equivalent to the WLS ANOVA F test when there is only one predictor. The WLS t-test for the intercept has $F = t^2 = 8.76$. This test statistic can be found from the no intercept OLS model by leaving the first column of $\boldsymbol{U}$ out of the model, then perform the partial F test as shown below.

```
> ls.print(lsfit(u[,2],z,intercept=F))
Residual Standard Error=1.2601
F-statistic (df=1, 34)=1436.300

  Estimate Std.Err t-value Pr(>|t|)
X   1.0038  0.0265 37.8985        0

> ((34*(1.2601)^2-33*(1.137)^2)/1)/(1.137)^2
[1] 8.760723
```

4.4   Complements

The theory for GLS and WLS is similar to the theory for the OLS MLR model, but the theory for FGLS and FWLS is often lacking or huge sample sizes are needed. However, FGLS and FWLS are often used in practice because usually V is not known and $\hat{V}$ must be used instead. Kariya and Kurata (2004) is a PhD level text covering FGLS.

Shi and Chen (2009) describe numerical diagnostics for GLS. Long and Ervin (2000) discuss methods for obtaining standard errors when the constant variance assumption is violated.

Following Sheather (2009, ch. 9, ch. 10) many linear models with serially correlated errors (eg AR(1) errors) and many linear mixed models can be fit with FGLS. Both Sheather (2009) and Houseman, Ryan and Coull (2004) use the Cholesky decomposition and make the residual plots based on the Cholesky residuals $Z - \hat{Z}$ where $V(\hat{\theta}) = KK^T$. Plots should be based on $Z - \hat{Z}$ where $V(\hat{\theta}) = RR$. In other words, use transformation corresponding to Proposition 4.2 instead of the transformation corresponding to Proposition 4.1.

4.5   Problems

Problems with an asterisk * are especially important.

R/Splus Problems

Use the command *source("A:/regpack.txt")* **to download the functions** and the command *source("A:/regdata.txt")* **to download the data.**

See Preface or Section 17.1. Typing the name of the `regpack` function, eg *wlsplot*, will display the code for the function. Use the `args` command, eg *args(wlsplot)*, to display the needed arguments for the function.

4.1. Generalized and weighted least squares are each equivalent to a least squares regression without intercept. Let $\boldsymbol{V} = \text{diag}(1, 1/2, 1/3, ..., 1/9) = \text{diag}(1/w_i)$ where $n = 9$ and the weights $w_i = i$ for $i = 1, ..., 9$. Let $\boldsymbol{x}^T = (1, x_1, x_2, x_3)$. Then the weighted least squares with weight vector $\boldsymbol{w}^T = (1, 2, ..., 9)$ should be equivalent to the OLS regression of $\sqrt{w_i}\, Y_i = Z_i$ on $\boldsymbol{u}$ where $\boldsymbol{u}^T = \sqrt{w_i}\boldsymbol{x} = (\sqrt{w_i}, \sqrt{w_i}x_1, \sqrt{w_i}x_2, \sqrt{w_i}x_3)$. There is no intercept because the vector of ones has been replaced by a vector of the $\sqrt{w_i}$'s. Type the following commands in *R/Splus* and include the output from both *lsfit* commands. The coefficients from both *lsfit* commands should be the same. The commands can also be copied and pasted from (www.math.siu.edu/olive/reghw.txt).

```
e <- rnorm(9)
x <- matrix(rnorm(27),nrow=9,ncol=3)
sqrtv <- sqrt(diag(1/1:9))
Y <- 4 + x%*%c(1,2,3) + sqrtv%*%e
wtt <- 1:9
lsfit(x,Y,wtt)$coef
kinv <- sqrt(diag(1:9))
Z <- kinv%*%Y
X <- 1 + 0*1:9
X <- cbind(X,x)
U <- kinv%*%X
lsfit(U,Z,int=F)$coef
```

4.2. Download the `wlsplot` function and the Draper and Smith (1981) data `dsx, dsy, dsw`.

a) Enter the *R/Splus* command **wlsplot(x=dsx, y = dsy, w = dsw)** to reproduce Figure 4.1. Once you have the plot you can print it out directly, but it will generally save paper by placing the plots in the *Word* editor.

b) Activate *Word* (often by double clicking on a *Word* icon). Click on the screen and type "Problem 4.2." In *R/Splus,* click on the plot and then press the keys *Ctrl* and *c* simultaneously. This procedure makes a temporary copy of the plot. In *Word,* move the pointer to *Edit* and hold down the leftmost mouse button. This will cause a menu to appear. Drag the pointer down to

Paste. In the future, these menu commands will be denoted by "Edit>Paste." The plot should appear on the screen. To save your output on your diskette, use the *Word* menu commands "File > Save as." In the **Save in** box select "3 1/2 Floppy(A:)" and in the *File name* box enter HW4d2.doc. To exit from *Word,* click on the "X" in the upper right corner of the screen. In *Word* a screen will appear and ask whether you want to save changes made in your document. Click on *No.* To exit from *R/Splus,* type "q()" or click on the "X" in the upper right corner of the screen and then click on *No.*

4.3. Download the `fwlssim` function. This creates WLS data if "type" is 1 or 3 and FWLS data if "type" is 2 or 4. Let the sufficient predictor $SP = 25 + 2x_2 + \cdots + 2x_p$. Then $Y = SP + |SP - 25k|\sigma e$ where the x_{ij} and e_i are iid $N(0, 1)$. Thus $Y|SP \sim N(SP, (SP - 25k)^2\sigma^2)$. If "type" is 1 or 2 then $k = 1/5$, but $k = 1$ if "type" is 3 or 4. The default has $\sigma^2 = 1$.

The function creates the OLS response and residual plots and the FWLS (or WLS) response and residual plots.

a) Type the following command several times. The OLS and WLS plots tend to look the same.

```
fwlssim(type=1)
```

b) Type the following command several times. Now the FWLS plots often have outliers.

```
fwlssim(type=2)
```

c) Type the following command several times. The OLS residual plots have a saddle shape, but the WLS plots tend to have highly skewed fitted values.

```
fwlssim(type=3)
```

d) Type the following command several times. The OLS residual plots have a saddle shape, but the FWLS plots tend to have outliers and highly skewed fitted values.

```
fwlssim(type=4)
```