# Chapter 8
# MLR with Heterogeneity

A multiple linear regression model with heterogeneity is

$$Y_i = \beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i \tag{8.1}$$

for $i = 1, ..., n$ where the $e_i$ are independent with $E(e_i) = 0$ and $V(e_i) = \sigma_i^2$. In matrix form, this model is

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

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 $\text{Cov}(\boldsymbol{e}) = \boldsymbol{\Sigma}_{\boldsymbol{e}} = diag(\sigma_i^2) = diag(\sigma_1^2, ..., \sigma_n^2)$ is an $n \times n$ positive definite matrix. In chapters 2 and 3, the constant variance assumption was used: $\sigma_i^2 = \sigma^2$ for all $i$. Hence heterogeneity means that the constant variance assumption does not hold. A common assumption is that the $e_i = \sigma_i \epsilon_i$ where the $\epsilon_i$ are independent and identically distributed (iid) with $V(\epsilon_i) = 1$.

Weighted least squares (WLS) would be useful if the $\sigma_i^2$ were known. Since the $\sigma_i^2$ are not known, ordinary least squares (OLS) is often used, but the large sample theory differs from that given in Chapter 2.

## 8.1 OLS Large Sample Theory

The OLS theory for MLR with heterogeneity often assume iid cases. For the following theorem, see Romano and Wolf (2017), Freedman (1981), and White (1980).

**Theorem 8.1.** Assume $Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ for $i = 1, ..., n$ where the cases $(Y_i, \boldsymbol{x}_i^T)^T$ are iid with "fourth moments," $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$, the $e_i = e_i(\boldsymbol{x}_i)$ are independent, $E[e_i|\boldsymbol{x}_i] = 0$, $\boldsymbol{V}^{-1} = E[\boldsymbol{x}_i\boldsymbol{x}_i^T]$, $E[e_i^2|\boldsymbol{x}_i] = v(\boldsymbol{x}_i) = \sigma_i^2$, $Cov[\boldsymbol{e}|\boldsymbol{X}] = diag(v(\boldsymbol{x}_1), ..., v(\boldsymbol{x}_n))$ and $\boldsymbol{\Omega} = E[v(\boldsymbol{x}_i)\boldsymbol{x}_i\boldsymbol{x}_i^T] = E[e_i^2\boldsymbol{x}_i\boldsymbol{x}_i^T]$.

Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\mathbf{0}, \boldsymbol{V\Omega V}). \tag{8.2}$$

**Remark 8.1.** a) White (1980) showed that the iid cases assumption can be weakened. Assume the cases are independent,

$$\boldsymbol{V}_n = \frac{1}{n}\sum_{i=1}^{n} E[\boldsymbol{x}_i \boldsymbol{x}_i^T] \xrightarrow{P} \boldsymbol{V}^{-1},$$

and

$$\boldsymbol{\Omega}_n = \frac{1}{n}\sum_{i=1}^{n} E[e_i^2 \boldsymbol{x}_i \boldsymbol{x}_i] \xrightarrow{P} \boldsymbol{\Omega}.$$

Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\mathbf{0}, \boldsymbol{V\Omega V}).$$

b) Under the assumptions of Theorem 8.1,

$$\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T \xrightarrow{P} \boldsymbol{V}^{-1}.$$

Let $\boldsymbol{D} = diag(\sigma_1^2, ..., \sigma_n^2) = \boldsymbol{\Sigma_e}$ and $\hat{\boldsymbol{D}} = diag(r_1^2, ..., r_n^2)$ where $r_i^2$ is the $i$th residual from OLS regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$. Then $\hat{\boldsymbol{D}}$ is not a consistent estimator of $\boldsymbol{D}$. The following theorem, due to White (1980), shows that $\hat{\boldsymbol{D}}$ can be used to get a consistent estimator of $\boldsymbol{\Omega}$. This result leads to the sandwich estimators given in the following section.

**Theorem 8.2.** Under strong regularity conditions,

$$\frac{1}{n}(\boldsymbol{X}^T\hat{\boldsymbol{D}}\boldsymbol{X}) \xrightarrow{P} \boldsymbol{\Omega} \text{ and } \frac{1}{\mathrm{n}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{X}) \xrightarrow{\mathrm{P}} \boldsymbol{\Omega}.$$

Hence

$$n(\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\hat{\boldsymbol{D}}\boldsymbol{X})(\boldsymbol{X}^T\boldsymbol{X})^{-1} \xrightarrow{P} \boldsymbol{V\Omega V}.$$

## 8.2 Bootstrap Methods and Sandwich Estimators

Under regularity conditions, the OLS estimator $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ can be shown to be a consistent estimator of $\boldsymbol{\beta}$ with $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $\text{Cov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma_e}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$. See, for example, White (1980). Assume $n\text{Cov}(\hat{\boldsymbol{\beta}}) \to \boldsymbol{V\Omega V}$ as $n \to \infty$. Assume $\boldsymbol{X}^T\boldsymbol{X}/n \to \boldsymbol{V}^{-1}$ and $\boldsymbol{X}^T\boldsymbol{\Sigma_e}\boldsymbol{X}/n \to \boldsymbol{\Omega}$ where convergence in probability is used if the $\boldsymbol{x}_i$ are random vectors. See Theorem 8.2. We assume that a constant $\beta_1$ corresponding to $x_1 \equiv 1$ is in the model so that the OLS residuals sum to 0.

A sandwich estimator is $\widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}_{OLS}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{D}}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$. Often $\hat{\boldsymbol{D}}$ is not a consistent estimator of $\boldsymbol{D} = \boldsymbol{\Sigma}_{\boldsymbol{e}}$, but often $\boldsymbol{X}^T\hat{\boldsymbol{D}}\boldsymbol{X}/n \xrightarrow{P} \boldsymbol{\Omega}$ under regularity conditions. For the wild bootstrap, we will use $\hat{\boldsymbol{D}}_W = n\ diag(r_1^2, ..., r_n^2)/(n-p)$ where the $r_i$ are the OLS residuals. Often $\hat{\boldsymbol{D}} = diag(d_i^2 r_i^2)$, where $\hat{\boldsymbol{D}}_W$ uses $d_i^2 = n/(n-p)$.

The *nonparametric bootstrap = pairs bootstrap* samples the cases $(Y_i, \boldsymbol{x}_i)$ with replacement, and uses

$$\boldsymbol{Y}^* = \boldsymbol{X}^*\hat{\boldsymbol{\beta}} + \boldsymbol{e}^*$$

with $\boldsymbol{e}^* = \boldsymbol{r}^*$ where $(Y_i, \boldsymbol{x}_i, r_i)$ are selected with replacement to form $\boldsymbol{Y}^*$, $\boldsymbol{X}^*$, and $\boldsymbol{r}^*$. Then $\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{Y}^* = \hat{\boldsymbol{\beta}} + (\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{r}^* = \hat{\boldsymbol{\beta}} + \boldsymbol{b}^*$ is obtained from the OLS regression of $\boldsymbol{Y}^*$ on $\boldsymbol{X}^*$. Thus $E(\hat{\boldsymbol{\beta}}^*) = \hat{\boldsymbol{\beta}} + E[(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{r}^*] = \hat{\boldsymbol{\beta}} + \boldsymbol{b}$ where the expectation is with respect to the bootstrap distribution and the bias vector $\boldsymbol{b} = E(\boldsymbol{b}^*)$. Freedman (1981) showed that the nonparametric bootstrap can be useful for model (8.1) with the $e_i$ independent, suggesting that $\boldsymbol{b}^* = o_p(n^{-1/2})$ or $\boldsymbol{b}^* = O_p(n^{-1/2})$. With respect to the bootstrap distribution, $\text{Cov}(\hat{\boldsymbol{\beta}}^*) = \text{Cov}[(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{r}^*] = E[(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{r}^*\boldsymbol{r}^{*T}\boldsymbol{X}^*(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}] - \boldsymbol{b}\boldsymbol{b}^T$. This result suggests that $\text{Cov}(\hat{\boldsymbol{\beta}}^*)$ is estimating the sandwich estimator

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{r}\boldsymbol{r}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1},$$

which replaces $diag(r_i^2)$ by $\boldsymbol{r}\boldsymbol{r}^T$. Also, with respect to the bootstrap distribution, the cases $(Y_i^*, \boldsymbol{x}_i^{*T})^T$ are iid with $V(e_i^*) = V(r_i^*)$ depending on $\boldsymbol{x}_i^*$.

A version of the *wild bootstrap* uses

$$\boldsymbol{Y}^* = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}^*$$

with $e_i^* = W_i c_n r_i$ where $P(W_i = \pm 1) = 0.5$, $E(W_i) = 0$, $V(W_i) = 1$ and $c_n = \sqrt{n/(n-p)}$. Note that $W_i = 2Z_i - 1$ where $Z_i \sim \text{binomial}(m = 1, p = 0.5) \sim \text{Bernoulli}(p = 0.5)$. See Flachaire (2005). With respect to the bootstrap distribution, the $c_n r_i$ are constants, and the $e_i^*$ are independent with $E(e_i^*) = E(W_i)c_n r_i = 0$, and $V(e_i^*) = E(e_i^{*2}) = E(W_i^2)c_n^2 r_i^2 = c_n^2 r_i^2$. Thus $E(\boldsymbol{e}^*) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{e}^*) = \hat{\boldsymbol{D}}_W$. Then $\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}^*$ with $E(\hat{\boldsymbol{\beta}}^*) = \hat{\boldsymbol{\beta}}$ and $\text{Cov}(\hat{\boldsymbol{\beta}}^*) = \widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}_{OLS}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{D}}_W\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$, a sandwich estimator. Note that $\text{Cov}(\hat{\boldsymbol{\beta}}^*) = \text{Cov}(\hat{\boldsymbol{\beta}}) + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T[\hat{\boldsymbol{D}}_W - \boldsymbol{\Sigma}_{\boldsymbol{e}}]\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

The following method is due to Rajapaksha and Olive (2022). For the OLS model of chapter 2, $V(e_i) = V(Y_i|\boldsymbol{x}_i) = V(Y_i|\boldsymbol{x}_i^T\boldsymbol{\beta}) = \sigma^2$. Hence $Y_i = Y_i|\boldsymbol{x}_i = Y_i|\boldsymbol{x}_i^T\boldsymbol{\beta} = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ with $V(e_i) = \sigma^2$. For model (8.1), $Y_i = Y_i|\boldsymbol{x}_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ with $V(e_i) = \sigma_i^2$, while $Y_i = Y_i|\boldsymbol{x}_i^T\boldsymbol{\beta} = \boldsymbol{x}_i^T\boldsymbol{\beta} + \epsilon_i$ with $V(\epsilon_i) = \tau_i^2$. The $\tau_i^2$ can be estimated as follows. Make the residual plot of $\hat{Y}_i = \boldsymbol{x}_i\hat{\boldsymbol{\beta}}$ versus $r_i$ on the vertical axis. Divide the ordered $\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$ into $m_s$ slices each containing approximately $n/m_s$ cases, and find the variance of the residuals $v_j^2$ in the

$j$th slice for $j = 1, ..., m_s$. Then $\hat{\tau}_i^2 = n v_j^2/(n-p)$ if case $i$ is in the $j$th slice. If the $\boldsymbol{x}_i$ are bounded, the maximum slice width $\to 0$, if $V(Y|\boldsymbol{x}^T\boldsymbol{\beta})$ is smooth, and the number of cases in each slice $\to \infty$ as $n \to \infty$, then $\hat{\tau}_i^2$ is a consistent estimator of $\tau_i^2$. This method acts as if the variance $\tau_j^2$ is constant within each slice $j$, and replaces $\hat{\boldsymbol{D}}_W = n \ diag(r_1^2, ..., r_n^2)/(n-p)$ by $diag(\hat{\tau}_1^2, ..., \hat{\tau}_n^2)$, a smoothed version of $\hat{\boldsymbol{D}}_W$. Another option would use a scatterplot smoother in a plot of $\hat{Y}_i$ vs. $r_i^2$.

The *parametric bootstrap* **does not assume** that the $e_i$ are normal, but uses

$$\boldsymbol{Y}^* = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}^*$$

where the $e_i^* \sim N(0, \hat{\tau}_i^2)$ are independent. Hence $\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}^* \sim$

$$N_p[\hat{\boldsymbol{\beta}}, (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \ diag(\hat{\tau}_1^2, ..., \hat{\tau}_n^2) \ \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}].$$

## 8.3 Simulations

Next, we describe a small simulation study that was done using $B = \max(200, 50p)$ and 5000 runs. The simulation is similar to that for the full OLS model done by Pelawa Watagoda and Olive (2021). The simulation used $p = 4, 6, 7, 8$, and 10; $n = 25p$ and $50p$; $\psi = 0, 1/\sqrt{p}$, and 0.9; and $k = 1$ and $p-2$ where $k$ and $\psi$ are defined in the following paragraph.

Let $\boldsymbol{x} = (1 \ \boldsymbol{u}^T)^T$ where $\boldsymbol{u}$ is the $(p-1) \times 1$ vector of nontrivial predictors. In the simulations, for $i = 1, ..., n$, we generated $\boldsymbol{w}_i \sim N_{p-1}(\boldsymbol{0}, \boldsymbol{I})$ where the $m = p-1$ elements of the vector $\boldsymbol{w}_i$ are iid N(0,1). Let the $m \times m$ matrix $\boldsymbol{A} = (a_{ij})$ with $a_{ii} = 1$ and $a_{ij} = \psi$ where $0 \le \psi < 1$ for $i \ne j$. Then the vector $\boldsymbol{u}_i = \boldsymbol{A}\boldsymbol{w}_i$ so that $Cov(\boldsymbol{u}_i) = \boldsymbol{\Sigma}_{\boldsymbol{u}} = \boldsymbol{A}\boldsymbol{A}^T = (\sigma_{ij})$ where the diagonal entries $\sigma_{ii} = [1+(m-1)\psi^2]$ and the off diagonal entries $\sigma_{ij} = [2\psi+(m-2)\psi^2]$. Hence the correlations are $cor(x_i, x_j) = \rho = (2\psi + (m-2)\psi^2)/(1 + (m-1)\psi^2)$ for $i \ne j$ where $x_i$ and $x_j$ are nontrivial predictors. If $\psi = 1/\sqrt{cp}$, then $\rho \to 1/(c+1)$ as $p \to \infty$ where $c > 0$. As $\psi$ gets close to 1, the predictor vectors cluster about the line in the direction of $(1, ..., 1)^T$. Let $Y_i = 1 + 1x_{i,2} + \cdots + 1x_{i,k+1} + e_i$ for $i = 1, ..., n$. Hence $\boldsymbol{\beta} = (1, .., 1, 0, ..., 0)^T$ with $k+1$ ones and $p-k-1$ zeros.

The zero mean iid errors $\epsilon_i$ were iid from five distributions: i) N(0,1), ii) $t_3$, iii) EXP(1) - 1, iv) uniform$(-1, 1)$, and v) 0.9 N(0,1) + 0.1 N(0,100). Only distribution iii) is not symmetric. Then wtype = 1 if $e_i = \epsilon_i$ (the WLS model is the OLS model), 2 if $e_i = |\boldsymbol{x}_i^T\boldsymbol{\beta} - 5|\epsilon_i$, 3 if $e_i = \sqrt{(1 + 0.5x_{i2}^2)}\epsilon_i$, 4 if $e_i = \exp[1 + \log(|x_{i2}|) + ... + \log(|x_{ip}|)]\epsilon_i$, 5 if $e_i = [1 + \log(|x_{i2}|) + ... + \log(|x_{ip}|)]\epsilon_i$, 6 if $e_i = [\exp([\log(|x_{i2}|) + ... + \log(|x_{ip}|)]/(p-1))]\epsilon_i$, 7 if $e_i = [[\log(|x_{i2}|) + ... + \log(|x_{ip}|)]/(p-1)]\epsilon_i$, The last four types were special cases of types suggested by Romano and Wolf (2017). For type 6, the weighting function is the geometric mean of $|x_{i2}|, ..., |x_{ip}|$.

When $\psi = 0$ and wtype = 1, the full model least squares confidence intervals for $\beta_i$ should have length near $2t_{96,0.975}\sigma/\sqrt{n} \approx 2(1.96)\sigma/10 = 0.392\sigma$ when $n = 100$ and the iid zero mean errors have variance $\sigma^2$. The simulation computed the Frey shorth($c$) interval for each $\beta_i$ and used bootstrap confidence regions to test $H_0 : \boldsymbol{\beta}_S = \mathbf{1}$ (whether first $k+1$ $\beta_i = 1$) and $H_0 : \boldsymbol{\beta}_E = \mathbf{0}$ (whether the last $p-k-1$ $\beta_i = 0$). The nominal coverage was 0.95 with $\delta = 0.05$. Observed coverage between 0.94 and 0.96 suggests coverage is close to the nominal value.

Table 8.1 shows two rows for each model giving the observed confidence interval coverages and average lengths of the confidence intervals. The terms "npar", "wild", and "par" are for the nonparametric, wild and parametric bootstrap. The last six columns give results for the tests. The terms pr, hyb, and br are for the prediction region method, hybrid region, and Bickel and Ren region. The 0 indicates the test was $H_0 : \boldsymbol{\beta}_E = \mathbf{0}$, while the 1 indicates that the test was $H_0 : \boldsymbol{\beta}_S = \mathbf{1}$. The length and coverage = P(fail to reject $H_0$) for the interval $[0, D_{(U_B)}]$ or $[0, D_{(U_B,T)}]$ where $D_{(U_B)}$ or $D_{(U_B,T)}$ is the cutoff for the confidence region. The cutoff will often be near $\sqrt{\chi^2_{g,0.95}}$ if the statistic $T$ is asymptotically normal. Note that $\sqrt{\chi^2_{2,0.95}} = 2.448$ is close to 2.45 for the full model regression bootstrap tests.

**Table 8.1** Bootstrapping WLS, wtype = 1, etype= $N(0,1)$

| $\psi$ | $\beta_1$ | $\beta_2$ | $\beta_{p-1}$ | $\beta_p$ | pr0 | hyb0 | br0 | pr1 | hyb1 | br1 |
|---|---|---|---|---|---|---|---|---|---|---|
| npar,0 | 0.946 | 0.950 | 0.947 | 0.948 | 0.940 | 0.941 | 0.941 | 0.937 | 0.936 | 0.937 |
| len | 0.396 | 0.399 | 0.399 | 0.398 | 2.451 | 2.451 | 2.452 | 2.450 | 2.450 | 2.451 |
| wild,0 | 0.948 | 0.950 | 0.997 | 0.996 | 0.991 | 0.979 | 0.991 | 0.938 | 0.939 | 0.940 |
| len | 0.395 | 0.398 | 0.323 | 0.323 | 2.699 | 2.699 | 3.002 | 2.450 | 2.450 | 2.457 |
| par,0 | 0.946 | 0.944 | 0.946 | 0.945 | 0.938 | 0.938 | 0.938 | 0.934 | 0.936 | 0.936 |
| len | 0.396 | 0.661 | 0.661 | 0.661 | 2.451 | 2.451 | 2.452 | 2.451 | 2.451 | 2.452 |
| npar,0.5 | 0.947 | 0.968 | 0.997 | 0.998 | 0.993 | 0.984 | 0.993 | 0.955 | 0.955 | 0.963 |
| len | 0.395 | 0.658 | 0.537 | 0.539 | 2.703 | 2.703 | 2.994 | 2.461 | 2.461 | 2.577 |
| wild,0.9 | 0.946 | 0.941 | 0.944 | 0.950 | 0.940 | 0.940 | 0.940 | 0.935 | 0.935 | 0.935 |
| len | 0.396 | 3.257 | 3.253 | 3.259 | 2.451 | 2.451 | 2.452 | 2.451 | 2.451 | 2.452 |
| par,0.9 | 0.947 | 0.968 | 0.994 | 0.996 | 0.992 | 0.981 | 0.992 | 0.962 | 0.959 | 0.970 |
| len | 0.395 | 2.751 | 2.725 | 2.735 | 2.716 | 2.716 | 2.971 | 2.497 | 2.497 | 2.599 |

Simulations in Rajapaksha (2021) suggest that the nonparametric bootstrap works better than the other methods used in Section 8.3.

## 8.4 OPLS in Low and High Dimensions

Under iid cases, OPLS theory does not depend on whether the error variance is constant or not. Hence the Olive and Zhang (2023) OPLS theory still applies. See Olive (2023f).

## 8.5 Summary

## 8.6 Complements

There is a large literature on regression with heterogeneity and sandwich estimators. See, for example, Buja et al. (2019), Eicker (1963, 1967), Hinkley (1977), Huber (1967), Long and Ervin (2000), MacKinnon and White (1985), Pötscher and Preinerstorfer (2022), White (1980), and Wu (1986). For more on the wild bootstrap, see Mammen (1992, 1993) and Wu (1986). Flachaire (2005) compares the wild and nonparametric bootstrap. Feasible weighted least squares estimates $\sigma_i^2$ or $v(\boldsymbol{x}_i)$, and is a competitor for OLS. See Romano and Wolf (2017).

## 8.7 Problems

**PROBLEMS WITH AN ASTERISK * ARE ESPECIALLY USEFUL.**

**8.1.**