Solutions (mostly for odd-numbered exercises)

"Microeconometrics: Methods and Applications"

1. Chapter 1: Introduction

No exercises.

2. Chapter 2: Causal and Noncausal Models

No exercises.

3. Chapter 3: Microeconomic Data Structures

No exercises.

4. Chapter 4: Linear Models

4-1 (a) For the diagonal entries $i = j$ and $\mathrm{E}[u_i^2] = \sigma^2$.
For the first off-diagonal $i = j - 1$ or $i = j + 1$ so $|i - j| = 1$ and $\mathrm{E}[u_i u_j] = \rho\sigma^2$.
Otherwise $|i - j| > 1$ and $\mathrm{E}[u_i u_j] = 0$.

(b) $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ is asymptotically normal with mean $\mathbf{0}$ and asymptotic variance matrix

$$\mathrm{V}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

where

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & 0 & \cdots & 0 \\ \rho\sigma^2 & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \rho\sigma^2 \\ 0 & \cdots & 0 & \rho\sigma^2 & \sigma^2 \end{bmatrix}.$$

(c) This example is a simple departure from the simplest case of $\boldsymbol{\Omega} = \sigma^2\mathbf{I}$.

Here $\mathbf{\Omega}$ depends on just two parameters and hence can be consistently estimated as $N \to \infty$.
So we use

$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{\Omega}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

where

$$\widehat{\mathbf{\Omega}} = \begin{bmatrix} \widehat{\sigma}^2 & \widehat{\rho\sigma}^2 & 0 & \cdots & 0 \\ \widehat{\rho\sigma}^2 & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \widehat{\rho\sigma}^2 \\ 0 & \cdots & 0 & \widehat{\rho\sigma}^2 & \widehat{\sigma}^2 \end{bmatrix}$$

and $\widehat{\mathbf{\Omega}} \xrightarrow{p} \mathbf{\Omega}$ if $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$ and $\widehat{\rho\sigma^2} \xrightarrow{p} \rho\sigma^2$ or $\widehat{\rho} \xrightarrow{p} \rho$.
For $\sigma^2 =\mathrm{E}[u_i^2]$ the obvious estimate is $\widehat{\sigma}^2 = N^{-1}\sum_{i=1}^N \widehat{u}_i^2$, where $\widehat{u}_i = y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}$.
For ρ we can directly use $\rho\sigma^2 =\mathrm{E}[u_i u_{i-1}]$ consistently estimated by $\widehat{\rho\sigma^2} = N^{-1}\sum_{i=2}^N \widehat{u}_i\widehat{u}_{i-1}$.
Or use $\rho =\mathrm{E}[u_i u_{i-1}]/\sqrt{\mathrm{E}[u_i]\mathrm{E}[u_{i-1}]} =\mathrm{E}[u_i u_{i-1}]/\mathrm{E}[u_i^2]$ consistently estimated by $\widehat{\rho} = N^{-1}\sum_{i=2}^N \widehat{u}_i\widehat{u}_{i-1}/N^{-1}\sum_{i=1}^N \widehat{u}_i^2$ and hence $\widehat{\rho}\widehat{\sigma}^2 = N^{-1}\sum_{i=2}^N \widehat{u}_i\widehat{u}_{i-1}$.

(d) To answer (d) and (e) it is helpful to use summation notation:

$$\begin{aligned} \widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] &= \left[\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\widehat{\sigma}^2\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i' + 2\widehat{\rho}\widehat{\sigma}^2\sum_{i=2}^N \mathbf{x}_i\mathbf{x}_{i-1}'\right]\left[\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i'\right]^{-1} \\ &= \widehat{\sigma}^2\left[\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i'\right]^{-1} + 2\widehat{\rho}\widehat{\sigma}^2\left[\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=2}^N \mathbf{x}_i\mathbf{x}_{i-1}'\right]\left[\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i'\right]^{-1} \end{aligned}$$

(d) No. The usual OLS output estimate $\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$ is inconsistent as it ignores the off-diagonal terms and hence the second term above.

(e) No. The White heteroskedasticity-robust estimate is inconsistent as it also ignores the off-diagonal terms and hence the second term above.

4-3 (a) The error u is conditionally heteroskedastic, since $\mathrm{V}[u|x] =\mathrm{V}[x\varepsilon|x] = x^2\mathrm{V}[\varepsilon|x] = x^2\mathrm{V}[\varepsilon] = x^2 \times 1 = x^2$ which depends on the regressor x.

(b) For scalar regressor $N^{-1}\mathbf{X}'\mathbf{X} =N^{-1}\sum_i x_i^2$.
Here x_i^2 are iid with mean 1 (since $\mathrm{E}[x_i^2] =\mathrm{E}[(x_i-\mathrm{E}[x_i])^2] =\mathrm{V}[x_i] = 1$ using $\mathrm{E}[x_i] = 0$).
Applying a LLN (here Kolmogorov), $N^{-1}\mathbf{X}'\mathbf{X} =N^{-1}\sum_i x_i^2 \xrightarrow{p} \mathrm{E}[x_i^2] = 1$, so $\mathbf{M}_{\mathbf{xx}} = 1$.

(c) $\mathrm{V}[u] =\mathrm{V}[x\varepsilon] =\mathrm{E}[(x\varepsilon)^2] - (\mathrm{E}[x\varepsilon])^2 =\mathrm{E}[x^2]\mathrm{E}[\varepsilon^2] - (\mathrm{E}[x]\mathrm{E}[\varepsilon])^2 =\mathrm{V}[x]\mathrm{V}[\varepsilon] - 0 \times 0 = 1 \times 1 = 1$ where use independence of x and ε and fact that here $\mathrm{E}[x] = 0$ and $\mathrm{E}[\varepsilon] = 0$.

(d) For scalar regressor and diagonal $\mathbf{\Omega}$,

$$N^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X} = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 x_i^2 = \frac{1}{N}\sum_{i=1}^{N}{x_i}^2 x_i^2 = \frac{1}{N}\sum_{i=1}^{N}{x_i}^4$$

using $\sigma_i^2 = x_i^2$ from (a).N
Here x_i^4 are iid with mean 3 (since $\mathrm{E}[x_i^4] = \mathrm{E}[(x_i - \mathrm{E}[x_i])^4] = 3$ using $\mathrm{E}[x_i] = 0$ and the fact that fourth central moment of normal is $3\sigma^4 = 3 \times 1 = 3$).
Applying a LLN (here Kolmogorov), $N^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X} = N^{-1}\sum_i x_i^4 \xrightarrow{p} \mathrm{E}[x_i^4] = 3$, so $\mathbf{M}_{\mathbf{x}\mathbf{\Omega}\mathbf{x}} = 3$.

(e) Default OLS result

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left[\mathbf{0},\ \sigma^2\mathbf{M}_{\mathbf{xx}}^{-1}\right] = \mathcal{N}\left[0,\ 1 \times (1)^{-1}\right] = \mathcal{N}\left[0,1\right].$$

(f) White OLS result

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left[\mathbf{0},\ \mathbf{M}_{\mathbf{xx}}^{-1}\mathbf{M}_{\mathbf{x}\mathbf{\Omega}\mathbf{x}}\mathbf{M}_{\mathbf{xx}}^{-1}\right] = \mathcal{N}\left[0,\ (1)^{-1} \times 3 \times (1)^{-1}\right] = \mathcal{N}[0,3].$$

(g) Yes. Expect that failure to control for conditional heteroskedasticity when should control for it will lead to inconsistent standard errors, though a priori the direction of the inconsistency is not known. That is the case here.
What is unusual compared to many applications is that there is a big difference in this example - the true variance is three times the default estimate and the true standard errrors are $\sqrt{3}$times larger.

4-5 **(a)** Differentiate

$$\begin{aligned}
\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{\partial \mathbf{u}'\mathbf{W}\mathbf{u}}{\partial \boldsymbol{\beta}} \\
&= \frac{\partial \mathbf{u}'}{\partial \boldsymbol{\beta}} \times \frac{\partial \mathbf{u}'\mathbf{W}\mathbf{u}}{\partial \mathbf{u}} \text{ by chain rule for matrix differentiation} \\
&= \mathbf{X}' \times 2\mathbf{W}\mathbf{u} \text{ assuming } \mathbf{W} \text{ is symmetric} \\
&= 2\mathbf{X}'\mathbf{W}\mathbf{u}
\end{aligned}$$

Set to zero

$$\begin{aligned}
& 2\mathbf{X}'\mathbf{W}\mathbf{u} = \mathbf{0} \\
\Rightarrow\ & 2\mathbf{X}'\mathbf{W}(\mathbf{y} - \mathbf{X}\beta) = \mathbf{0} \\
\Rightarrow\ & \mathbf{X}'\mathbf{W}\mathbf{y} = \mathbf{X}'\mathbf{W}\mathbf{X}\beta \\
\Rightarrow\ & \widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}
\end{aligned}$$

where need to assume the inverse exists.
Here $\mathbf{W}$ is rank $r > K$ =rank($\mathbf{X}$) $\Rightarrow$ $\mathbf{X}'\mathbf{Z}$ and $\mathbf{Z}'\mathbf{Z}$ are rank $K \Rightarrow \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$ is of full rank K.

(b) For $\mathbf{W} = \mathbf{I}$ we have $\widehat{\boldsymbol{\beta}}$=$(\mathbf{X}'\mathbf{I}\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ which is OLS.
Note that $(\mathbf{X}'\mathbf{X})^{-1}$ exists if $N \times K$ matrix $\mathbf{X}$ is of full rank K.

(c) For $\mathbf{W}$ =$\boldsymbol{\Omega}^{-1}$ we have $\widehat{\boldsymbol{\beta}}$=$(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$ which is GLS (see (4.28)).

(d) For $\mathbf{W} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ we have $\widehat{\boldsymbol{\beta}}$=$(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$ which is 2SLS (see (4.53)).

4-7 Given the information, $\mathrm{E}[x] = 0$ and $\mathrm{E}[z] = 0$ and

$$\begin{array}{l}
\mathrm{V}[x] = \mathrm{E}[x^2] = \mathrm{E}[(\lambda u + \varepsilon)^2] = \lambda^2\sigma_u^2 + \sigma_\varepsilon^2 \\
\mathrm{V}[z] = \mathrm{E}[z^2] = \mathrm{E}[(\gamma\varepsilon + v)^2] = \gamma^2\sigma_\varepsilon^2 + \sigma_v^2 \\
\mathrm{Cov}[x, z] = \mathrm{E}[xz] = \mathrm{E}[(\lambda u + \varepsilon)(\gamma\varepsilon + v)] = \lambda\sigma_\varepsilon^2 \\
\mathrm{Cov}[x, u] = \mathrm{E}[xu] = \mathrm{E}[(\lambda u + \varepsilon)u] = \lambda\sigma_u^2
\end{array}$$

(a) For regression of y on x we have $\widehat{\beta}_{\mathrm{OLS}} = \left(\sum_i x_i^2\right)^{-1}\sum_i x_i y_i$ and as usual

$$\begin{array}{ll}
\mathrm{plim}(\widehat{\beta}_{\mathrm{OLS}} - \beta) & = \left(\mathrm{plim}\sum_i x_i^2\right)^{-1}\mathrm{plim}\sum_i x_i u_i \\
& = \left(\mathrm{E}[x^2]\right)^{-1}\mathrm{E}[xu] \text{ as here data are iid} \\
& = (\lambda^2\sigma_u^2 + \sigma_\varepsilon^2)^{-1}\lambda\sigma_u^2.
\end{array}$$

(b) The squared correlation coefficient is

$$\begin{array}{ll}
\rho_{XZ}^2 & = [\mathrm{Cov}[x, z]^2]/[\mathrm{V}[x]\mathrm{V}[z]] \\
& = [\gamma\sigma_\varepsilon^2]^2/[(\lambda^2\sigma_u^2 + \sigma_\varepsilon^2)(\gamma^2\sigma_\varepsilon^2 + \sigma_v^2)]
\end{array}$$

(c) For single regressor and single instrument

$$\begin{array}{ll}
\widehat{\beta}_{\mathrm{IV}} & = \left(\sum_i z_i x_i\right)^{-1}\sum_i z_i y_i \\
& = \left(\sum_i z_i x_i\right)^{-1}\sum_i z_i(x_i\beta + u_i) \\
& = \beta + \left(\sum_i z_i x_i\right)^{-1}\sum_i z_i u_i \\
& = \beta + \left(\sum_i z_i(\lambda u_i + \varepsilon_i)\right)^{-1}\sum_i z_i u_i \\
& = \beta + \left(\lambda\sum_i z_i u_i + \sum_i z_i\varepsilon_i\right)^{-1}\sum_i z_i u_i \\
& = \beta + \left(\lambda m_{zu} + m_{z\varepsilon}\right)^{-1} m_{zu} \\
\widehat{\beta}_{\mathrm{IV}} - \beta & = m_{zu}/\left(\lambda m_{zu} + m_{z\varepsilon}\right)
\end{array}$$

where $m_{zu} = N^{-1}\sum_i z_i u_i$ and $m_{z\varepsilon} = N^{-1}\sum_i z_i\varepsilon_i$.
By a LLN $m_{zu} \xrightarrow{p} \mathrm{E}[m_{zu}] = \mathrm{E}[z_i u_i] = \mathrm{E}[(\gamma\varepsilon_i+v_i)u_i] = 0$ since ε, u and v are independent with zero means.
By a LLN $m_{z\varepsilon} \xrightarrow{p} \mathrm{E}[m_{z\varepsilon}] = \mathrm{E}[z_i\varepsilon_i] = \mathrm{E}[(\gamma\varepsilon_i + v_i)\varepsilon_i] = \gamma\mathrm{E}[\varepsilon_i^2] = \gamma\sigma_\varepsilon^2$.

$$\widehat{\beta}_{\mathrm{IV}} - \beta \quad \xrightarrow{p} 0/\lambda \times 0 + \gamma\sigma_\varepsilon^2) = 0.$$

(d) If $m_{zu} = -m_{z\varepsilon}/\lambda$ then $\lambda m_{zu} = -m_{z\varepsilon}$ so $\lambda m_{zu} - m_{z\varepsilon} = 0$ and $\widehat{\beta}_{\mathrm{IV}} - \beta = m_{zu}/0$ which is not defined.

(e) First

$$\begin{aligned}\widehat{\beta}_{\mathrm{IV}} - \beta \ &= m_{zu}/\left(\lambda m_{zu} + m_{z\varepsilon}\right) \\ &= 1/(\lambda + m_{z\varepsilon}/m_{zu}).\end{aligned}$$

If m_{zu} is large relative to $m_{z\varepsilon}/\lambda$ then λ is large relative to $m_{z\varepsilon}/m_{zu}$ so $\lambda + m_{z\varepsilon}/m_{zu}$ is close to λ and $1/(\lambda + m_{z\varepsilon}/m_{zu})$ is close to $1/\lambda$.

(f) Given the definition of ρ_{XZ}^2 in part (c), ρ_{XZ}^2 is smaller the smaller is γ, the smaller is σ_ε^2, and the larger is λ. So in the weak instruments case with small correlation between x and z (ao ρ_{XZ}^2 is small), $\widehat{\beta}_{\mathrm{IV}} - \beta$ is likely to converge to $1/\lambda$ rather than 0, and there is "finite sample bias" in $\widehat{\beta}_{\mathrm{IV}}$.

4-11 (a) The true variance matrix of OLS is

$$\begin{aligned}\mathrm{V}[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2(\mathbf{I}_N+\mathbf{A}\mathbf{A}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\mathbf{A}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.\end{aligned}$$

(b) This equals or exceeds $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ since $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\mathbf{A}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ is positive semi-definite. So the default OLS variance matrix, and hence standard errors, will generally understate the true standard errors (the exception being if $\mathbf{X}'\mathbf{A}\mathbf{A}'\mathbf{X} = \mathbf{0}$).

(c) For GLS

$$\begin{aligned}\mathrm{V}[\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}] &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'[\sigma^2(\mathbf{I} + \mathbf{A}\mathbf{A}')]^{-1}\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'[\mathbf{I} + \mathbf{A}\mathbf{A}']^{-1}\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'[\mathbf{I}_N - \mathbf{A}(\mathbf{I}_m + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{A}(\mathbf{I}_m + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{X})^{-1}.\end{aligned}$$

(d) $\sigma^2(\mathbf{X}'\mathbf{X})^{-1} \leq \mathrm{V}[\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}]$ since

$$\begin{aligned} & \mathbf{X}'\mathbf{X} \geq \mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{A}(\mathbf{I}_m + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{X} \text{ in the matrix sense} \\ \Rightarrow \quad & (\mathbf{X}'\mathbf{X})^{-1} \leq (\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{A}(\mathbf{I}_m + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{X})^{-1} \text{ in the matrix sense.} \end{aligned}$$

If we ran OLS and GLS and used the incorrect default OLS standard errors we would obtain the puzzling result that OLS was more effficient than GLS. But this is just an artifact of using the wrong estimated standard errors for OLS.

(e) GLS requires $(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$ which from (c) requires $(\mathbf{I}_m + \mathbf{A}'\mathbf{A})^{-1}$ which is the inverse of an $m \times m$ matrix.
[We also need $(\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{A}(\mathbf{I}_m + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{X})^{-1}$ but this is a smaller $k \times k$ marix given $k < m < N$.]

4-13 (a) Here $\boldsymbol{\beta} = [1\ 1]$ and $\alpha = [1\ 0]$.
From bottom of page 86 the intercept will be $\beta_1 + \alpha_1 \times F_\varepsilon^{-1}(q) = 1 + 1 \times F_\varepsilon^{-1}(q) = 1 + F_\varepsilon^{-1}(q)$.
The slope will be $\beta_2 + \alpha_2 \times F_\varepsilon^{-1}(q) = 1 + 0 \times F_\varepsilon^{-1}(q) = 1$.
The slope should be 1 at all quantiles.
The intercept varies with $F_\varepsilon^{-1}(q)$. Here $F_\varepsilon^{-1}(q)$ takes values -2.56, -1.68 , -1.05, -0.51, 0.0, 0.51, 1.05, 1.68 and 2.56 for $q = 0.1, 0.2, \ldots.$, 0.9. It follows that the intercept takes values -1.56, -0.68 , -0.05, 0.49, 1.0, 1.51, 2.05, 2.68.
[For example $F_\varepsilon^{-1}(0.9)$ is ε^* such that $\Pr[\varepsilon \leq \varepsilon^*] = 0.9$ for $\varepsilon \sim \mathcal{N}[0,4]$ or equivalently ε^* such that $\Pr[z \leq \varepsilon^*/2] = 0.9$ for $z \sim \mathcal{N}[0,1]$. Then $\varepsilon^*/2 = 1.28$ so $\varepsilon^* = 2.56$.]

(b) The answers accord quite closely with theory as the slope and intercepts are quite precisely estimated with slope coefficient standard errors less than 0.01 and intercept coefficient standard errors less than 0.04.

(c) Now both the intercept and slope coefficients vary with the quantile. Both intercept and slope coefficients increase with the quantile, and for $1 = 0.5$ are within two standard errors of the true values of 1 and 1.

(d) Compared to (b) it is now the intercept that is constant and the slope that varies across quantiles.
This is predicted from theory similar to that in part (a). Now $\boldsymbol{\beta} = [1\ 1]$ and $\alpha = [0\ 1]$.
From bottom of page 86 the intercept will be $\beta_1 + \alpha_1 \times F_\varepsilon^{-1}(q) = 1 + 0 \times F_\varepsilon^{-1}(q) = 1$ and the slope will be $\beta_2 + \alpha_2 \times F_\varepsilon^{-1}(q) = 1 + 1 \times F_\varepsilon^{-1}(q) = 1 + F_\varepsilon^{-1}(q)$.

4-15 (a) The OLS slope estimate and standard error are 0.05209 and 0.00291, and the IV estimates are 0.18806 and 0.02614. The IV slope estimate is much larger and indicates a very large return to schooling. There is a lossin precision with IV standard error ten times larger, but the coefficient is still statististically significant.

(b) OLS of wage76 on an intercept and col4 gives slope coefficient 0.1559089 and OLS regression of grade76 on an intercept and col4 gives slope coefficient 0.829019. From (4.46) , $dy/dx = (dy/dz)/(dx/dz) = 0.1559089/0.829019 = 0.18806$. This is the same as the IV estimate in part (a).

(c) We obtain Wald = (1.706234 - 1.550325) / (13.52703 - 12.69801) = 0.18806. This is the same as the IV estimate in part (a).

(d) From OLS regression of grade76 on col4, $R^2 = 0.0208$ and $F = 60.37$. This does not suggest a weak instruments problem, except that precision of IV will be much lower than that of OLS due to the relatively low R^2.

(e) Including the additional regressors the OLS slope estimate and standard error are 0.03304 and 0.00311, and the IV estimates are 0.09521 and 0.04932. The IV slope estimate is again much larger and indicates a very large return to schooling. There is a loss in precision with IV standard error now eighteed ten times larger, but the coefficient is still statististically significant using a one-tail test at five percent.
Now OLS of wage76 on an intercept and col4 and other regressors gives slope coefficient 0.1559089 and OLS regression of grade76 on an intercept and col4 gives slope coefficient 0.829019. From (4.46) , $dy/dx = (dy/dz)/(dx/dz) = 0.1559089/0.829019 = 0.18806$. This is the same as the IV estimate in part (a).

4-17 (a) The average of $\widehat{\beta}_{\text{OLS}}$ over 1000 simulations was 1.502518.
This is close to the theoretical value of 1.5: $\text{plim}(\widehat{\beta}_{\text{OLS}} - \beta) = \lambda\sigma_u^2/\left(\lambda^2\sigma_u^2 + \sigma_\varepsilon^2\right) = (1 \times 1)/(1 \times 1 + 1) = 1/2$ and here $\beta = 1$.

(b) The average of $\widehat{\beta}_{\text{IV}}$ over 1000 simulations was 1.08551.
This is close to the theoretical value of 1: $\text{plim}(\widehat{\beta}_{\text{IV}} - \beta) = 0$ and here $\beta = 1$.

(c) The observed values of $\widehat{\beta}_{\text{IV}}$ over 1000 simulations were skewed to the right of $\beta = 1$, with lower quartile .964185, median 1.424028 and upper quartile 1.7802471. Exercise 4-7 part (e) suggested concentration of $\widehat{\beta}_{\text{IV}} - \beta$ around $1/\lambda = 1$ or concetration of $\widehat{\beta}_{\text{IV}}$ around $\beta + 1 = 2$ since here $\beta = 1$.

(d) The R^2 and F statistics across simulations from OLS regression (with intercept) of z on x do indicate a likely weak instruments problem.
Over 1000 simulations, the average R^2 was 0.0148093 and the average F was 1.531256.
[Aside: From Exercise 4-7 (b) $\rho_{XZ}^2 = [\gamma\sigma_\varepsilon^2]^2/[(\lambda^2\sigma_u^2 + \sigma_\varepsilon^2)(\gamma^2\sigma_\varepsilon^2 + \sigma_v^2) = [0.01]^2/(1 + 1)(0.01^2 + 1) = 0.00005$.]

5. Chapter 5: Extremum, ML, NLS

5-1 First note that

$$\begin{aligned}
\frac{\partial \widehat{\mathrm{E}}[y|x]}{\partial x} &= \frac{\partial}{\partial x} \exp(1+0.01x)[1+\exp(1+0.01x)]^{-1} \\
&= 0.01 \exp(1+0.01x)[1+\exp(1+0.01x)]^{-1} \\
&\quad - \exp(1+0.01x) \times 0.01 \exp(1+0.01x)[1+\exp(1+0.01x)]^{-2} \\
&= 0.01 \times \frac{\exp(1+0.01x)}{[1+\exp(1+0.01x)]^2} \text{ upon simplification}
\end{aligned}$$

(a) The average marginal effect over all observations.

$$\frac{\partial \widehat{\mathrm{E}}[y|x]}{\partial x} = \frac{1}{100}\sum_{i=1}^{100} 0.01 \times \frac{\exp(1+0.01i)}{1+\exp(1+0.01i)} = 0.0014928.$$

(b) The sample mean $\bar{x} = \frac{1}{100}\sum_{i=1}^{100} i = 50.5$. Then

$$\left.\frac{\partial \widehat{\mathrm{E}}[y|x]}{\partial x}\right|_{\bar{x}} = 0.01 \times \frac{\exp(1+0.01 \times 50.5)}{[1+\exp(1+0.01x \times 50.5)]^2} = 0.0014867.$$

(c) Evaluating at $x = 90$

$$\left.\frac{\partial \widehat{\mathrm{E}}[y|x]}{\partial x}\right|_{90} = 0.01 \times \frac{\exp(1+0.01 \times 90)}{[1+\exp(1+0.01x \times 90)]^2} = 0.0011318.$$

(d) Using the finite difference method

$$\left.\frac{\Delta \widehat{\mathrm{E}}[y|x]}{\Delta x}\right|_{90} = \frac{\exp(1+0.01 \times 90)}{1+\exp(1+0.01x \times 90)} - \frac{\exp(1+0.01 \times 90)}{1+\exp(1+0.01x \times 90)} = 0.0011276.$$

Comment: This example is quite linear, leading to answers in (a) and (b) being close, and similarly for (c) and (d). A more nonlinear function, with greater variation is obtained using $\widehat{\mathrm{E}}[y|x] = \exp(0+0.04x)/[1+\exp(0+0.04x)]$ for $x = 1, ..., 100$. Then the answers are 0.0026163, 0.0013895, 0.00020268, and 0.00019773.

5-2 (a) Here

$$\begin{aligned}\ln f(y) &= \ln y - 2\ln\lambda - y/\lambda \text{ with } \lambda = \exp(\mathbf{x}'\boldsymbol{\beta})/2 \text{ and } \ln\lambda = \mathbf{x}'\boldsymbol{\beta} - \ln 2 \\ &= \ln y - 2(\mathbf{x}'\boldsymbol{\beta} - \ln 2) - y/[\exp(\mathbf{x}'\boldsymbol{\beta})/2] \\ &= \ln y - 2\mathbf{x}'\boldsymbol{\beta} + 2\ln 2 - 2y\exp(-\mathbf{x}'\boldsymbol{\beta})\end{aligned}$$

so

$$Q_N(\boldsymbol{\beta})=\frac{1}{N}\sum\nolimits_i \ln f(y_i) = \frac{1}{N}\sum\nolimits_i \{\ln y_i - 2\mathbf{x}'\boldsymbol{\beta} + 2\ln 2 - 2y_i\exp(-\mathbf{x}'\boldsymbol{\beta})\}.$$

(b) Now using $\mathbf{x}$ nonstochastic so need only take expectations wrt y

$$\begin{aligned}Q_0(\boldsymbol{\beta}) &= \operatorname{plim} Q_N(\boldsymbol{\beta}) \\ &= \operatorname{plim}\frac{1}{N}\sum\nolimits_i \ln y_i - \operatorname{plim}\frac{1}{N}\sum\nolimits_i 2\mathbf{x}_i'\boldsymbol{\beta} + \operatorname{plim}\frac{1}{N}\sum\nolimits_i 2\ln 2 - \operatorname{plim}\frac{1}{N}\sum\nolimits_i 2y_i\exp(-\mathbf{x}_i'\boldsymbol{\beta}) \\ &= \lim\frac{1}{N}\sum\nolimits_i \mathrm{E}[\ln y_i] - 2\lim\frac{1}{N}\sum\nolimits_i \mathbf{x}_i'\boldsymbol{\beta} + 2\ln 2 - 2\lim\frac{1}{N}\sum\nolimits_i \mathrm{E}[y_i]\exp(-\mathbf{x}_i'\boldsymbol{\beta}) \\ &= \lim\frac{1}{N}\sum\nolimits_i \mathrm{E}[\ln y_i] - 2\lim\frac{1}{N}\sum\nolimits_i \mathbf{x}_i'\boldsymbol{\beta} + 2\ln 2 - 2\lim\frac{1}{N}\sum\nolimits_i \exp(\mathbf{x}_i'\boldsymbol{\beta}_0)\exp(-\mathbf{x}_i'\boldsymbol{\beta}),\end{aligned}$$

where the last line uses $\mathrm{E}[y_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta}_0)$ in the dgp and we do not need to evaluate $\mathrm{E}[\ln y_i]$ as the first sum does not invlove $\boldsymbol{\beta}$ and will therefore have derivative of $\mathbf{0}$ wrt $\boldsymbol{\beta}$.

(c) Differentiate wrt $\boldsymbol{\beta}$ (not $\boldsymbol{\beta}_0$)

$$\begin{aligned}\frac{\partial Q_0(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} &= -2\lim\frac{1}{N}\sum\nolimits_i \mathbf{x}_i + \lim\frac{2}{N}\sum\nolimits_i \exp(\mathbf{x}_i'\boldsymbol{\beta}_0)\exp(-\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i \\ &= \mathbf{0} \text{ when } \boldsymbol{\beta} = \boldsymbol{\beta}_0.\end{aligned}$$

[Also $\partial^2 Q_0(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = -2\lim N^{-1}\sum_i \exp(\mathbf{x}_i'\boldsymbol{\beta}_0)\exp(-\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i\mathbf{x}_i'$ is negative definite at $\boldsymbol{\beta}_0$, so local max.]
Since $\operatorname{plim} Q_N(\boldsymbol{\beta})$ attains a local maximum at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, conclude that $\widehat{\boldsymbol{\beta}} = \arg\max Q_N(\boldsymbol{\beta})$ is consistent for $\boldsymbol{\beta}_0$.

(d) Consider the last term. Since $y_i\exp(-\mathbf{x}_i'\boldsymbol{\beta})$ is not iid need to use Markov SLLN. This requires existence of second moments of y_i which we have assumed.

5-3 (a) Differentiating $Q_N(\boldsymbol{\beta})$ wrt $\boldsymbol{\beta}$

$$\begin{aligned}\frac{\partial Q_N}{\partial\boldsymbol{\beta}} &= \frac{1}{N}\sum\nolimits_i \left(-2\mathbf{x}_i + 2y_i\exp(-\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i\right) \\ &= \frac{1}{N}\sum\nolimits_i 2\times\{y_i\exp(-\mathbf{x}_i'\boldsymbol{\beta}) - 1\}\mathbf{x}_i \quad \text{rearranging} \\ &= \frac{1}{N}\sum\nolimits_i 2\times\frac{y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta})}{\exp(\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{x}_i \quad \text{multiplying by } \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\exp(\mathbf{x}_i'\boldsymbol{\beta})}\end{aligned}$$

(b) Then

$$\lim \mathrm{E}\left[\left.\frac{\partial Q_N}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}_0}\right] = \lim \frac{1}{N}\sum\nolimits_i 2 \times \frac{y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}_0)}{\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)}\mathbf{x}_i = \mathbf{0} \text{ if } \mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta}_0).$$

So essential condition is correct specification of $\mathrm{E}[y_i|\mathbf{x}_i]$.

(c) From (a)

$$\sqrt{N}\left.\frac{\partial Q_N}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}_0} = \frac{1}{\sqrt{N}}\sum\nolimits_i 2 \times \frac{y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}_0)}{\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)}\mathbf{x}_i.$$

Apply CLT to average of the term in the sum.
Now $y_i|\mathbf{x}_i$ has mean $\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)$ and variance $(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2/2$.
So $X_i \equiv 2 \times \frac{y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}_0)}{\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)}\mathbf{x}_i$ has mean $\mathbf{0}$ and variance $4 \times \frac{(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2/2}{(\exp(\mathbf{x}_i'\boldsymbol{\beta}_0))^2}\mathbf{x}_i\mathbf{x}_i' = 2\mathbf{x}_i\mathbf{x}_i'$.
Thus for $Z_N = (\mathrm{V}[\sqrt{N}\bar{X}])^{-1/2}(\sqrt{N}\bar{X} - \sqrt{N}\mathrm{E}[\bar{X}]) = \left(\frac{1}{N}\sum_i \mathrm{V}[X_i]\right)^{-1/2}\left(\frac{1}{\sqrt{N}}\sum_i X_i\right)$

$$\begin{aligned} Z_N &= \left(\frac{1}{N}\sum\nolimits_i 2\mathbf{x}_i\mathbf{x}_i'\right)^{-1/2} \times \left(\frac{1}{\sqrt{N}}\sum\nolimits_i 2 \times \frac{y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}_0)}{\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)}\mathbf{x}_i\right) \xrightarrow{d} \mathcal{N}[\mathbf{0}, \mathbf{I}] \\ &\Rightarrow \frac{1}{\sqrt{N}}\sum\nolimits_i 2 \times \frac{y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}_0)}{\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)}\mathbf{x}_i \xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \lim \frac{1}{N}\sum\nolimits_i 2\mathbf{x}_i\mathbf{x}_i'\right] \end{aligned}$$

(d) Here y_i is not iid. Use Liapounov CLT.
This will need a $(2+\delta)^{th}$ absolute moment of y_i. e.g. 4^{th} moment of y_i.

(e) Differentiating (a) wrt $\boldsymbol{\beta}'$ yields

$$\left.\frac{\partial^2 Q_N}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}_0} = \frac{1}{N}\sum\nolimits_i \left(-2\frac{\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)}{\exp(\mathbf{x}_i'\boldsymbol{\beta}_0)}\mathbf{x}_i\mathbf{x}_i'\right) \xrightarrow{p} \lim \frac{1}{N}\sum\nolimits_i -2\mathbf{x}_i\mathbf{x}_i'.$$

(f) Combining

$$\begin{aligned} &\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}[\mathbf{0}, \mathbf{A}(\boldsymbol{\beta}_0)^{-1}\mathbf{B}(\boldsymbol{\beta}_0)\mathbf{A}(\boldsymbol{\beta}_0)^{-1}] \\ &\xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \left(\lim \frac{1}{N}\sum\nolimits_i -2\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\lim \frac{1}{N}\sum\nolimits_i 2\mathbf{x}_i\mathbf{x}_i'\right)\left(\lim \frac{1}{N}\sum\nolimits_i -2\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\right] \\ &\xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \left(\lim \frac{1}{N}\sum\nolimits_i 2\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\right]. \end{aligned}$$

(g) Test $H_0 : \beta_{0j} \geq \beta_j^*$ against $H_a : \beta_{0j} < \beta_j^*$ at level .05.

$$\widehat{\boldsymbol{\beta}} \overset{a}{\sim} \mathcal{N}\left[\boldsymbol{\beta}, \left(\sum\nolimits_i 2\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\right]$$

$$\Rightarrow \quad z_j = \frac{(\widehat{\beta}_j - \beta_j)}{s_j} \overset{a}{\sim} \mathcal{N}[0,1], \text{ where } s_j \text{ is } j^{th} \text{ diag entry in } \left(\sum\nolimits_i 2\mathbf{x}_i\mathbf{x}_i'\right)^{-1}.$$

Reject H_0 at level 0.05 if $z_j < -z_{.05} = -1.645$.

5-5 (a) $t = \widehat{\theta}_1/se[\widehat{\theta}_1] = 5/2 = 2.5$. Since $|2.5| > \dot{z}_{.05} = 1.645$ we reject H_0.

(b) Rewrite as $H_0 : \theta_1 - 2\theta_2 = 0$ versus $H_0 : \theta_1 - 2\theta_2 \neq 0$.
Use (5.32). Test $H_0 : \mathbf{R}\boldsymbol{\theta} = r$ where $\mathbf{R} = [1 \ -2]$ and $r = 0$ and $\boldsymbol{\theta}' = [\theta_1 \ \theta_2]$.
Here $\widehat{\boldsymbol{\theta}} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ so $\mathbf{R}\widehat{\boldsymbol{\theta}} - \mathbf{r} = [1 \ -2]\begin{bmatrix} 5 \\ 2 \end{bmatrix} = 1$.
Also $\mathrm{V}[\widehat{\boldsymbol{\theta}}] = N^{-1}\widehat{\mathbf{C}} = \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}$ using $\mathrm{Cov}[\widehat{\theta}_1, \widehat{\theta}_2] = (\mathrm{Cor}[\widehat{\theta}_1, \widehat{\theta}_2])^2\mathrm{V}[\widehat{\theta}_1]\mathrm{V}[\widehat{\theta}_2] = 0.5^2 \times 2^2 \times 1^2 = 1$.
Then $\mathbf{R}N^{-1}\widehat{\mathbf{C}}\mathbf{R}' = [1 \ -2]\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -2 \end{bmatrix} = 4$
so $\mathrm{W} = (\mathbf{R}\widehat{\boldsymbol{\theta}} - \mathbf{r})'\left(\mathbf{R}(N^{-1}\widehat{\mathbf{C}})\mathbf{R}'\right)^{-1}(\mathbf{R}\widehat{\boldsymbol{\theta}} - \mathbf{r}) = 1 \times 4^{-1} \times 1$.
Since $\mathrm{W} = 0.25 < \chi^2_{1;.05} = 3.84$ do not reject H_0.

[Alternatively as only one restriction here, note that $\widehat{\theta}_1 - 2\widehat{\theta}_2$ has variance $\mathrm{V}[\widehat{\theta}_1] + 4\mathrm{V}[\widehat{\theta}_1] - 4\mathrm{Cov}[\widehat{\theta}_1, \widehat{\theta}_2] = 4 + 4 \times 1 - 4 \times 1 = 4$, leading to

$$t = \frac{\widehat{\theta}_1 - 2\widehat{\theta}_2}{\mathrm{se}[\widehat{\theta}_1 - 2\widehat{\theta}_2]} = \frac{5-3}{\sqrt{4}} = 0.5$$

and do not reject as $|0.5| < z_{.05} = 1.96$. Note that $t^2 = $W.]

(c) Use (5.32) Test $H_0 : \mathbf{R}\boldsymbol{\theta} = r$ where $\mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $r = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$.
Then $\mathbf{R}\widehat{\boldsymbol{\theta}} - \mathbf{r} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$
and $\mathbf{R}N^{-1}\widehat{\mathbf{C}}\mathbf{R}' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}$
so $\mathrm{W} = (\mathbf{R}\widehat{\boldsymbol{\theta}} - \mathbf{r})'\left(\mathbf{R}(N^{-1}\widehat{\mathbf{C}})\mathbf{R}'\right)^{-1}(\mathbf{R}\widehat{\boldsymbol{\theta}} - \mathbf{r}) = \begin{bmatrix} 5 & 2 \end{bmatrix}\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 5 \\ 2 \end{bmatrix} = 124$.
Since $\mathrm{W} = 124 < \chi^2_{2;.05} = 5.99$ reject H_0.

5-7 Results will vary as uses generated data. Expect $\widehat{\beta}_1 \simeq -1$ and $\widehat{\beta}_2 \simeq 1$ and standard errors similar to those below.

(a) For NLS got $\widehat{\beta}_1 = -1.1162$ and $\widehat{\beta}_2 = 1.1098$ with standard errors 0.0551 and 0.0256.

(b) Yes, will need to use sandwich errors due to heteroskedasticity as $\mathrm{V}[y|x] = \exp(\beta_1 + \beta_2 x)^2/2$. Note that standard errors given in (a) do not correct for heteroskedasticity.

(c) For MLE got $\widehat{\beta}_1 = -1.0088$ and $\widehat{\beta}_2 = 1.0262$ with standard errors 0.0224 and 0.0215.

(d) Sandwich errors can be used but are not necessary since the ML simplification that $\mathbf{A} = -\mathbf{B}$ is appropriate here.