Solutions to Problem Set #2

Q1. Exercise 4.1

(a) The method of moment estimator for μ_k is

$$\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} Y_i^k.$$

(b) By taking the expectation and using the iid assumption

$$E[\widehat{\mu}_k] = E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^k\right] = \frac{1}{n}\sum_{i=1}^{n} E[Y_i^k] = \frac{1}{n} n\mu_k = \mu_k.$$

(c) Observe that

$$var[\widehat{\mu}_k] = E[(\widehat{\mu}_k - E[\widehat{\mu}_k])^2] = E[(\widehat{\mu}_k - \mu_k)^2].$$

Since

$$(\widehat{\mu}_k - \mu_k)^2 = \left(\frac{1}{n}\sum_{i=1}^{n} Y_i^k - \mu_k\right)^2 = \frac{1}{n^2}\left(\sum_{i=1}^{n}(Y_i^k - \mu_k)\right)^2$$

by the iid assumption and $E(Y_i^k - \mu_k)(Y_j^k - \mu_k) = E(Y_i^k - \mu_k)E(Y_j^k - \mu_k)$ for $i \neq j$,

$$E[(\widehat{\mu}_k - \mu_k)^2] = \frac{1}{n^2}E\left(\sum_{i=1}^{n}(Y_i^k - \mu_k)\right)^2 = \frac{1}{n^2}\sum_{i=1}^{n} E(Y_i^k - \mu_k)^2 = \frac{1}{n}var(Y_i^k).$$

Since $var(Y_i^k) = E[Y_i^{2k}] - (E[Y_i^k])^2$, we have $var[\widehat{\mu}_k] < \infty$ if $E[Y_i^{2k}] < \infty$.

(d) The method of moment estimator of $var[\widehat{\mu}_k]$ is

$$\widehat{var[\widehat{\mu}_k]} = \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^k - \frac{1}{n}\sum_{i=1}^{n} Y_i^k\right)^2\right) = \frac{1}{n^2}\sum_{i=1}^{n}\left(Y_i^k - \frac{1}{n}\sum_{i=1}^{n} Y_i^k\right)^2.$$

Q2. Exercise 4.5

Under the assumption (4.13), the model (4.12) implies that $E[\mathbf{Y}|\mathbf{X}] = \mathbf{X}\beta$. The least squares estimator $\widehat{\beta}$ is given by

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

By taking the conditional expectation, we have

$$E[\widehat{\beta}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{Y}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta = \beta.$$

Next, by taking the conditional variance operator, we have

$$var[\widehat{\beta}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'var[\mathbf{e}|\mathbf{X}]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

Q3. Exercise 4.12
Since

$$\overline{Y} - \mu = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu),$$

we can write

$$E[(\overline{Y} - \mu)^3] = \frac{1}{n^3}E\left[\left(\sum_{i=1}^{n}(Y_i - \mu)\right)^3\right].$$

Observe that

$$\left(\sum_{i=1}^{n}(Y_i - \mu)\right)^3 = \sum_{i=1}^{n}(Y_i - \mu)^3 + 3\sum_{i=1}^{n}\sum_{j\neq i}(Y_i - \mu)^2(Y_j - \mu) + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}(Y_i - \mu)(Y_j - \mu)(Y_k - \mu).$$

By the iid assumption, Y_i and Y_j are independent for $i \neq j$ and thus $E[Y_iY_j] = E[Y_i]E[Y_j]$. Using this,

$$E\left[\left(\overline{Y} - \mu\right)^3\right] = \frac{1}{n^3}\sum_{i=1}^{n}E(Y_i - \mu)^3 = \frac{\mu_3}{n^2}.$$

Q4. Exercise 4.23
The conditional expectation of the ridge estimator $\widehat{\beta}$ is

$$E[\widehat{\beta}|\mathbf{X}] = (\mathbf{X}'\mathbf{X} + \mathbf{I}_k\lambda)^{-1}\mathbf{X}'E[\mathbf{Y}|\mathbf{X}] = (\mathbf{X}'\mathbf{X} + \mathbf{I}_k\lambda)^{-1}\mathbf{X}'\mathbf{X}\beta \neq \beta.$$

Therefore, $\widehat{\beta}$ is biased for β. Note that the identity matrix is non-random so it can be taken out of the conditional expectation operator.

Q5. Exercise 4.26
We regress the standardized test score on the suggested covariates and calculate the heteroskedasticity-robust standard error (using HC1 formula) and the cluster (and heteroskedasticity) robust standard error at the school level. The number of observations is 5,269 and the number of clusters (schools) is 111. The estimation results are summaries below.

variable	$\widehat{\beta}$	conventional s.e.	cluster s.e.
tracking	0.174	0.024	0.076
agetest	-0.041	0.009	0.013
girl	0.082	0.024	0.028
etpteacher	0.181	0.024	0.037
percentile	0.017	0.0004	0.0007
constant	-0.740	0.081	0.129

(a) Comparing the two sets of standard errors, we find that the s.e. of *tracking* is the most affected and the s.e. of *girl* is the least affected.
(b) In equation (4.55), the coefficient estimate of *tracking* is 0.138. With the included control variables (i.e., covariates), the point estimate has increased to 0.174. This suggests an even stronger effect of tracking on students' test scores.

Q6. The answer should include a well defined research question and an equation of interest to answer the research question. The estimation results should be provided and how the standard errors are calculated should be discussed.