HOMEWORK 3 SOLUTION

Problem SMSA. Consider the data at

http://pages.uoregon.edu/dlevin/DATA/smsa.txt

This file is tab-delimited, so use the following to read it:

```
> smsa = na.omit(
+   read.table("http://pages.uoregon.edu/dlevin/DATA/smsa.txt",
+                              header=T,sep="\t",row.names=1))
```

The variables are explained in

http://pages.uoregon.edu/dlevin/DATA/SMSA.html

The goal is to understand the relationship of Mortality to the other variables.

Note that `NOx` and `NOxPot` are identical, so exclude one (why?)

The variables can be divided into demographic and climate variables. Is there significant evidence that the climate variables should be included in modeling mortality? Explain any test that you perform.

Give a confidence interval for the expected Mortality in Indianapolis. What assumptions are you making to guarantee that the confidence level of this interval? Are these assumptions reasonable? Is this interval useful? What is the source of uncertainty, if any, about the mortality rate in Indianapolis?

Give a confidence interval for the coefficient of NOx. Does the length of this interval depend on which other variables you include in the model? Discuss.

Suppose all variables are included except NOxPot, and you want to test that the coefficients of NOx and Education are both zero. Estimate the power of the appropriate test of this hypothesis, when the coefficients are 1 and -10, respectively.

If the power is low, discuss why.

Solution to Problem SMSA. Since `NOx` and `NOxPot` are identical, including both will cause the design matrix $\boldsymbol{X}$ to have less than full rank, leading to difficulties computing the OLS estimates. We leave NOx out.

We want to determine if any of the climate variables should be included in the model. In other words, we test

$$H_0 : \beta_{\text{JanTemp}} = \beta_{\text{JulyTemp}} = \beta_{\text{RelHum}} = \beta_{\text{Rain}} = 0\,.$$

To test whether these coefficients are simultaneously zero, we perform an F-test, comparing the model with all variables to the sub-model omitting these three variables.

The result of that test is summarized in Table 1. Since the p-value is 0.03, there is reasonable evidence that this hypothesis is false.

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	48	68123.14				
2	44	53806.71	4	14316.43	2.93	0.0313

TABLE 1. F-test that climate coefficients are zero

Fitting a model with all variables, a 95% confidence interval for expected mortality in a city with identical covariates (the variables on the right-hand side of the model equation) as Indianapolis is

$$[941.5,\ 980.57]\,.$$

The assumptions used to assert the 95% confidence level of this interval are

- The data are observations on variables $\boldsymbol{Y}$ and $\boldsymbol{X}$ satsfying

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,. \tag{1}$$

- The error vector $\boldsymbol{\varepsilon}$ in (1) have a Normal distribution and are uncorrelated across observations,
- The errors are independent of the covariates

All of these assumptions require a leap of faith. Indeed, it is hard to imagine that there are not other variables which are important determinants of mortality besides those included here. Furthermore, having surely omitted some variables which are correlated with both the included covariates and mortality, it is unlikely the covariates are uncorrelated with the $\boldsymbol{\varepsilon}$ in (1).

Since you know what the measured mortality is in Indianapolis, what is the point of finding a confidence interval for $\mathbb{E}[Y_{\text{indianapolis}} \mid \boldsymbol{x}^{(\text{indianapolis})}]$, where $\boldsymbol{x}^{(\text{indianapolis})}$ are the observed covariates in Indianapolis? How we interpret this expectation depends on how we interpret the error term in the model

$$y_{\text{indianapolis}} = \boldsymbol{\beta}\boldsymbol{x}^{(\text{indianapolis})} + \varepsilon_{\text{indianapolis}}\,.$$

The error piece ε may include other variables which are determinants of mortality but were not included, measurement error (the mortality rate may itself be imperfectly measured, if it is determined through random sampling), and intrinsic randomness. The confidence interval is useful if you believe that the error is produced largely from measurement error and intrinsic randomness rather than from omitted covariates. Indeed, if the error comes from measurement or pure chance, then if the data is collected in Indianapolis in other years, assuming the included covariates remain constant, then the mortality may change due to this chance components.

In all circumstances, the interval is only useful if you believe the underlying model.

The 95% confidence interval for β_{NOx} when all the variables are included is

$$[-0.664, 3.021]\,,$$

while the 95% confidence interval when the climate variables are omitted is

$$[-0.371, 3.456]\,.$$

The intervals are different depending on what variables are included in the regression model. This can be seen from the formula for the the standard error for β_i, which is $\hat{\sigma^2}(\boldsymbol{X}'\boldsymbol{X})^{-1}_{i,i}$. As discussed elsewhere, unless the variables are orthogonal, including or excluding variables changes the matrix entries $(\boldsymbol{X}'\boldsymbol{X})^{-1}_{i,i}$.

Let β_{NOx} and β_{educ} be the coefficients of `NOx` and `education`. We want to test

$$H_0 : \beta_{\text{NOx}} = \beta_{\text{educ}} = 0\,.$$

using the F-test. Let V_0 be the linear span of all the variables excluding $\boldsymbol{x}_{\text{NOx}}$ and $\boldsymbol{x}_{\text{educ}}$, let $W = \mathcal{L}(\boldsymbol{x}_{\text{NOx}}, \boldsymbol{x}_{\text{educ}})$, let $\boldsymbol{x}^{\perp}_{\text{educ}} = \Pi_{V_0^{\perp}}(\boldsymbol{x}_{\text{educ}})$, and let $\boldsymbol{x}^{\perp}_{\text{NOx}} =$

$\Pi_{V_0^\perp}(\boldsymbol{x}_{\text{NOx}})$. When $\beta_{\text{educ}} = -10$ and $\beta_{\text{NOx}} = 1$,

$$\begin{aligned}
\|\Pi_W(\boldsymbol{\theta})\|^2 &= \beta_{\text{educ}}^2 \|\boldsymbol{x}_{\text{educ}}^\perp\|^2 + \beta_{\text{NOx}}^2 \|\boldsymbol{x}_{\text{NOx}}^\perp\|^2 + 2\beta_{\text{educ}}\beta_{\text{NOx}} \langle \boldsymbol{x}_{\text{educ}}^\perp, \boldsymbol{x}_{\text{NOx}}^\perp \rangle \\
&= \beta_{\text{educ}}^2 15.165 + \beta_{\text{NOx}}^2 1468.512 + 2\beta_{\text{educ}}\beta_{\text{NOx}} 9.507 \\
&= 100 \cdot 15.165 + 1 \cdot 1468.512 + 2(10)(-1)9.507 \\
&= 2794.872\,.
\end{aligned}$$

Since $\hat{\sigma^2} = 1222.9009$,

$$\begin{aligned}
\delta &= \frac{\|\Pi_W \boldsymbol{\theta}\|^2}{\sigma^2} \\
\hat{\delta} &= \frac{2794.872}{1222.9009} = 2.285
\end{aligned}$$

Since $f^\star = 3.209$ is the 95-th percentile of the F distribution with 2 and 44 degrees of freedom, the power of the test is approximately

$$\mathbb{P}(F > f^\star \mid \delta = 2.285) = 0.238\,.$$

The power is low because $\beta_{\text{NOx}} \|\boldsymbol{x}_{\text{NOx}}^\perp\|/\sigma$ and $\beta_{\text{edu}} \|\boldsymbol{x}_{\text{educ}}^\perp\|/\sigma$ are small.

By fitting the model with the given variables, it is assumed that the mortality in Indianapolis is the observed value of a random variable equal to a linear combination of the covariates (the variables on the right-hand side of the model equation) plus a random disturbance term.

□

Problem Freedman 4.2. In the OLS regression model, do the residuals always have mean 0? Discuss briefly.

Solution to Problem Freedman 4.2. The meaning of the word "mean" here is abiguous. In the sense of expectation,

$$\begin{aligned}
\mathbb{E}[\Pi_{V^\perp}(\boldsymbol{Y} \mid \boldsymbol{X})] &= \mathbb{E}[\Pi_{V^\perp}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \mid \boldsymbol{X}] \\
&= \Pi_{V^\perp} \mathbb{E}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}]
\end{aligned}$$

So under the assumptions that $\boldsymbol{\varepsilon}$ has expecation 0 and is independent of $\boldsymbol{X}$, the expectation of the residuals is zero.

What about if "mean" is interpreted as "sample mean"? Provided that an intercept is included, $\hat{\boldsymbol{Y}} - \boldsymbol{Y}$ is orthogonal to $\mathbf{1}$, whence

$$0 = \langle \mathbf{1}, \hat{\boldsymbol{Y}} - \boldsymbol{Y} \rangle = \sum_i (\hat{Y}_i - Y_i)\,,$$

in this case, the residuals sum to zero and their sample mean must also be zero.

If no intercept is included, then the residuals may not be orthogonal to $\mathbf{1}$, whence the residuals may not sum to zero. □

Problem Freedman 4.3. True or false, and explain. If, after conditioning on X, the disturbance terms in a regression equation are correlated with each other across subjects, then

(a) the OLS estimates are likely to be biased;
(b) the estimated standard errors are likely to be biased.

Solution to Problem Freedman 4.3. Note that

$$\mathbb{E}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}) \mid \boldsymbol{X}] = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbb{E}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}]$$

The last term vanishes if $\mathbb{E}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}] = 0$, which does not depend on the convariance structure.

On the other hand, let $\boldsymbol{M} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$; we have

$$\mathrm{Var}(\boldsymbol{M}\boldsymbol{X}'\boldsymbol{Y}) = \boldsymbol{M}\boldsymbol{X}'\mathrm{Cov}(\varepsilon)\boldsymbol{X}\boldsymbol{M}\,.$$

If $\mathrm{Cov}(\varepsilon) = \sigma^2 \boldsymbol{I}$, then this reduces to $\sigma^2\boldsymbol{M}$, whence the expression $\sigma\sqrt{\boldsymbol{M}_{i,i}}$ gives the standard deviation of $\hat{\beta}_i$. If, on the other hand, $\mathrm{Cov}(\varepsilon)$ is not a multiple of the identity, then $\sigma\sqrt{\boldsymbol{M}_{i,i}}$ may not be the standard deviation of $\hat{\beta}_i$. □

Problem Freedman 4.5. You are using OLS to fit a regression equation. True or false, and explain:

(a) If you exclude a variable from the equation, but the excluded variable is orthogonal to the other variables in the equation, you won't bias the estimated coefficients of the remaining variables.
(b) If you exclude a variable from the equation, and the excluded variable isn't orthogonal to the other variables, your estimates are going to be biased.
(c) If you put an extra variable into the equation, you won't bias the estimated coefficients – as long as the error term remains independent of the explanatory variables.
(d) If you put an extra variable into the equation, you are likely to bias the estimated coefficients – if the error term is dependent on that extra variable.

Solution to Problem Freedman 4.5. Let $\tilde{\boldsymbol{\beta}} = (\beta_0, \ldots, \beta_{k-1})'$ be the vector of parameters omitting the coefficient of $\boldsymbol{x}_k$, and $\hat{\tilde{\boldsymbol{\beta}}}$ for the corresponding OLS estimates. If $\tilde{\boldsymbol{X}}$ is the matrix of covariates excluding $\boldsymbol{x}_k$, then

$$\hat{\tilde{\boldsymbol{\beta}}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{Y}\,.$$

Taking conditional expectation,

$$\begin{aligned}\mathbb{E}[\hat{\tilde{\boldsymbol{\beta}}} \mid \boldsymbol{X}] &= (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'[(\tilde{\boldsymbol{X}}\tilde{\beta} + \boldsymbol{x}_k\beta_k) + \mathbb{E}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}]] \\ &= (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{\beta}} + (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{x}_k\beta_k \\ &= \tilde{\boldsymbol{\beta}} + (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{x}_k\beta_k \qquad (2)\end{aligned}$$

(a) True. From (2), if $\boldsymbol{x}_k \perp \tilde{\boldsymbol{X}}$, then

$$\mathbb{E}[\hat{\tilde{\boldsymbol{\beta}}} \mid \boldsymbol{X}] = \tilde{\boldsymbol{\beta}}\,,$$

and the OLS estimator is unbiased.
(b) True. From (2), if $\boldsymbol{x}_j'\boldsymbol{x}_k \neq 0$ for any $j \neq k$, then the second term in (2) is non-zero, and the estimator is biased.
(c) True. Adding a variable which is independent of the error will not introduce bias, because the model equation

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

still remains valid, even if $\beta_k = 0$ in reality. The assumptions needed to guarantee an unbiased estimate still hold.
(d) True, see Exercise 14 below for details.

□

Problem Freedman 4.9. True or false, and explain:

(a) Collinearity leads to bias in the OLS estimates.
(b) Collinearity leads to bias in the estimated standard errors for the OLS estimates
(c) Collinearity leads to big standard errors for some estimates.

Solution to Problem Freedman 4.9. Only the last statement is true. Collinearity will not violate any of the assumptions which make the OLS estimators unbiased. However, a near collinear relationship will make the standard errors large. To see this, partition $\boldsymbol{X} = [\tilde{\boldsymbol{X}}\ \boldsymbol{x}_k]$. Then

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} \tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} & \tilde{\boldsymbol{X}}'\boldsymbol{x}_k \\ \boldsymbol{x}_k'\tilde{\boldsymbol{X}} & \boldsymbol{x}_k'\boldsymbol{x}_k \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} - \tilde{\boldsymbol{X}}'\boldsymbol{x}_k(\boldsymbol{x}_k'\boldsymbol{x})^{-1}\boldsymbol{x}_k'\tilde{\boldsymbol{X}})^{-1} & B_{12} \\ B_{21} & (\boldsymbol{x}_k'\boldsymbol{x}_k - \boldsymbol{x}_k'\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{x}_k)^{-1} \end{bmatrix}, \tag{3}$$

where

$$B_{12} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{x}_k(\boldsymbol{x}_k'\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{x}_k - \boldsymbol{x}_k'\boldsymbol{x}_k)^{-1}$$
$$B_{21} = (\boldsymbol{x}_k'\boldsymbol{x}_k)^{-1}\boldsymbol{x}_k'\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\boldsymbol{x}_k\boldsymbol{x}_k'\boldsymbol{x}_k^{-1}\boldsymbol{x}_k'\tilde{\boldsymbol{X}} - \tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}$$

Let V_k be the span of the columns of $\tilde{\boldsymbol{X}}$. Note that if $\boldsymbol{x}_k^{\perp} = \boldsymbol{x} - \Pi_{V_k}\boldsymbol{x}$, then the lower right-hand block of the matrix above is

$$(\|\boldsymbol{x}_k\|^2 - \|\Pi_{V_k}\boldsymbol{x}_k\|^2)^{-1} = \|\boldsymbol{x}_k^{\perp}\|^{-2}.$$

If $\boldsymbol{x}_k$ is nearly a linear combination of the other columns, then $\boldsymbol{x}_k^{\perp}$ is small and thus the variance of $\hat{\beta}_k$ is large, since the variance is $\sigma^2\|\boldsymbol{x}_k\|^{-2}$.] □

Problem Freedman 4.10. Suppose $(X_i, W_i, \varepsilon_i)$ are IID as triplets across subjects $i = 1, \dots, n$, where n is large; $\mathbb{E}(X_i) = \mathbb{E}(W_i) = \mathbb{E}(\varepsilon_i) = 0$, and ε_i is independent of (X_i, W_i). Happily, X_i and W_i have positive variance; they are not perfectly correlated. The response variable Y_i is in truth this:

$$Y_i = aX_i + bW_i + \varepsilon_i\,.$$

We can recover a and b, up to random error, by running a regression of Y_i on X_i and W_i. No intercept is needed. Why not? What happens if X_i and W_i are perfectly correlated (as random variables)?

Solution to Problem Freedman 4.10. No intercept is needed in fitting the model $Y_i = aX_i + bW_i + \varepsilon_i$ since there is no intercept in the model. If W_i and X_i and perfecly correlated, then $W_i = rX_i + s$, and thus the design matrix has rank less than 2 and the inverse of $[\boldsymbol{X}\boldsymbol{W}]'[\boldsymbol{X}\boldsymbol{W}]$, required in the OLS estimates of a and b, does not exists. □

Problem Freedman 4.11. (This continues question Freedman 4.10.) Tom elects to run a regression of Y_i on X_i, omitting W_i. He will use the coefficient of X_i to estimate a.

(a) What happens to Tom if X_i and W_i are independent?
(b) What happens to Tom if X_i and W_i are dependent?

Solution to Problem Freedman 4.11. Regressing Y_i on X_i alone, the OLS estimator of a is given by:

$$\hat{a} = \frac{\langle \boldsymbol{Y}, \boldsymbol{X} \rangle}{\|\boldsymbol{X}\|^2}\,.$$

We always have

$$\begin{aligned}\mathbb{E}[\langle \boldsymbol{X}, \boldsymbol{W} \rangle \mid \boldsymbol{X}] &= \sum_{i=1}^{n} \mathbb{E}[X_i W_i \mid \boldsymbol{X}] \\ &= n X_1 \mathbb{E}[W_1 \mid X_1]\,.\end{aligned}$$

Note that

$$\operatorname{Var}(\mathbb{E}[W_1 \mid X_1]) = \operatorname{Var}(W_1) - \mathbb{E}[\operatorname{Var}(W_1 \mid X_1)]\,,$$

so that if $\operatorname{Var}(W_1 \mid X_1) < \operatorname{Var}(W_1)$ with positive probability, then $\operatorname{Var}(\mathbb{E}[W_1 \mid X_1]) > 0$. In this case, $\mathbb{E}[W_1 \mid X_1] \neq 0$ with positive probability.

Since

$$\mathbb{E}[\hat{a} \mid \boldsymbol{X}] = \frac{a\|\boldsymbol{X}\|^2 + ab\mathbb{E}\left[\langle \boldsymbol{X}, \boldsymbol{W} \rangle \mid \boldsymbol{X}\right]}{\|\boldsymbol{X}\|^2}\,,$$

if $\operatorname{Var}(W_1 \mid X_1) < \operatorname{Var}(W_1)$ with positive probability, then $\mathbb{E}[\hat{a} \mid \boldsymbol{X}] \neq a$ with positive probability. In the case of independence, $\mathbb{E}[W_1 \mid X_1] = \mathbb{E}[W_1] = 0$, whence $\mathbb{E}[\hat{a} \mid \boldsymbol{X}] = a$. □

Problem Freedman 4.12. Suppose $(X_i, \delta_i, \varepsilon_i)$ are IID as triplets across subjects $i = 1, \ldots, n$, where n is large; and X_i, δ_i, ε_i are mutually independent. Furthermore, $\mathbb{E}(X_i) = \mathbb{E}(\delta_i) = \mathbb{E}(\varepsilon_i) = 0$ while $\mathbb{E}(X_i^2) = \mathbb{E}(\delta_i^2) = 1$ and $\mathbb{E}(\varepsilon_i^2) = \sigma^2 > 0$. The response variable Y_i is in truth this:

$$Y_i = aX_i + \varepsilon_i\,.$$

We can recover a, up to random error, by running a regression of Y_i on X_i . No intercept is needed. Why not?

Solution to Problem Freedman 4.12. No intercept is needed because there is no constant term in the model equation. □

Problem Freedman 4.13. Let c, d, e be real numbers and let $W_i = cX_i + d\delta_i + e\varepsilon_i$. Dick elects to run a regression of Y_i on X_i and W_i, again without an intercept. Dick will use the coefficient of X_i in his regression to estimate a. If $e = 0$, Dick still gets a, up to random error– as long as $d \neq 0$. Why? And what's wrong with $d = 0$?

Solution to Problem Freedman 4.13. If $e = 0$, then both $\boldsymbol{X}$ and $\boldsymbol{W}$ are independent of ε, which is required in our assumptions for the OLS estimator to be unbiased. If $d = e = 0$, then $\boldsymbol{X} = c\boldsymbol{W}$, and the design matrix has less than full rank. □

Problem Freedman 4.14. (Continues questions 4.12 and 4.13.) Suppose, however, that $e \neq 0$. Then Dick has a problem. To see the problem more clearly, assume that n is large. Let $Q = [\boldsymbol{X}\boldsymbol{W}]$ be the design matrix, i.e., the first column is the X_i's and the second column is the W_i's. Show that

$$\frac{1}{n}\boldsymbol{Q}'\boldsymbol{Q} \doteq \begin{bmatrix} \mathbb{E}(X_i^2) & \mathbb{E}(X_i W_i) \\ \mathbb{E}(X_i W_i) & \mathbb{E}(W_i^2) \end{bmatrix}, \quad \frac{1}{n}\boldsymbol{Q}'\boldsymbol{Y} \doteq \begin{bmatrix} \mathbb{E}(X_i W_i) \\ \mathbb{E}(W_i Y_i) \end{bmatrix}. \tag{4}$$

(a) Suppose $a = c = d = e = 1$. What will Dick estimate for the coefficient of X_i in his regression?
(b) Suppose $a = c = d = 1$ and $e =$ ˇ1. What will Dick estimate for the coefficient of X_i in his regression?
(c) A textbook on regression advises that, when in doubt, put more explanatory variables into the equation, rather than fewer. What do you think?

Solution to Problem Freedman 4.14.

$$\boldsymbol{Q}'\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{W} \\ \boldsymbol{W}'\boldsymbol{X} & \boldsymbol{W}'\boldsymbol{W} \end{bmatrix}$$

The Law of Large numbers implies that

$$\begin{aligned} \frac{1}{n}\langle \boldsymbol{X}, \boldsymbol{X} \rangle &= \frac{1}{n}\boldsymbol{X}'\boldsymbol{X} = \frac{1}{n}\sum_i X_i^2 \to \mathbb{E}X_1^2 = 1 \\ \frac{1}{n}\langle \boldsymbol{X}, \boldsymbol{W} \rangle &= \frac{1}{n}\boldsymbol{X}'\boldsymbol{W} = \frac{1}{n}\sum_i X_i W_i \to \mathbb{E}X_1 W_1 = c \\ \frac{1}{n}\langle \boldsymbol{W}, \boldsymbol{W} \rangle &= \frac{1}{n}\boldsymbol{W}'\boldsymbol{W} = \frac{1}{n}\sum W_i^2 \to \mathbb{E}W_1^1 = c^2 + d^2 + e^2\sigma^2 \end{aligned}$$

Thus

$$\frac{1}{n}\boldsymbol{Q}'\boldsymbol{Q} \doteq \begin{bmatrix} \mathbb{E}(X_1 Y_1) \\ \mathbb{E}(W_1 Y_1) \end{bmatrix} = \begin{bmatrix} 1 & c \\ c & c^2 + d^2 + e^2\sigma^2 \end{bmatrix}$$

Similarly,

$$\frac{1}{n}\boldsymbol{Q}'\boldsymbol{Y} \doteq \begin{bmatrix} a \\ ac + \varepsilon\sigma^2 \end{bmatrix} \tag{5}$$

Thus

$$(\boldsymbol{Q}'\boldsymbol{Q})^{-1}\boldsymbol{Q}'\boldsymbol{Y} \doteq \frac{1}{d^2 + e^2\sigma^2}\begin{bmatrix} c^2 + d^2 + e^2\sigma^2 & -c \\ -c & 1 \end{bmatrix}\begin{bmatrix} a \\ ac + e\sigma^2 \end{bmatrix} = \begin{bmatrix} a + \frac{(ae-c)e\sigma^2}{d^2} \\ \frac{e\sigma^2}{d^2} \end{bmatrix}.$$

(a) If $a = c = d = e = 1$, then the coefficient of X_i will be, up to random error, 1.
(b) If $a = c = d = 1$ and $e = -1$, then the coefficient of X_i will be, up to random error, $a - 2\sigma^2$.
(c) If the variables you seek to include are correlated with the error, then adding these variables may bias your estimates of the other variables! Thus caution is advised.

□