APPENDIX E

SOLUTIONS TO PROBLEMS

E.1 This follows directly from partitioned matrix multiplication in Appendix D. Write

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}, \mathbf{X}' = (\mathbf{x}_1' \ \mathbf{x}_2' \ \ldots \ \mathbf{x}_n'), \text{ and } \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}$$

Therefore, $\mathbf{X}'\mathbf{X} = \sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t$ and $\mathbf{X}'\mathbf{y} = \sum_{t=1}^{n} \mathbf{x}_t'\mathbf{y}_t$. An equivalent expression for $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{\beta}} = \left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t \right)^{-1} \left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t' y_t \right)$$

which, when we plug in $y_t = \mathbf{x}_t\boldsymbol{\beta} + u_t$ for each t and do some algebra, can be written as

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t \right)^{-1} \left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t' u_t \right).$$

As shown in Section E.4, this expression is the basis for the asymptotic analysis of OLS using matrices.

E.3 (i) We use the placeholder feature of the OLS formulas. By definition, $\tilde{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = [(\mathbf{XA})' (\mathbf{XA})]^{-1}(\mathbf{XA})'\mathbf{y} = [\mathbf{A}'(\mathbf{X}'\mathbf{X})\mathbf{A}]^{-1}\mathbf{A}'\mathbf{X}'\mathbf{y} = \mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}')^{-1}\mathbf{A}'\mathbf{X}'\mathbf{y} = \mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}$.

(ii) By definition of the fitted values, $\hat{y}_t = \mathbf{x}_t\hat{\boldsymbol{\beta}}$ and $\tilde{y}_t = \mathbf{z}_t\tilde{\boldsymbol{\beta}}$. Plugging $\mathbf{z}_t$ and $\tilde{\boldsymbol{\beta}}$ into the second equation gives $\tilde{y}_t = (\mathbf{x}_t\mathbf{A})(\mathbf{A}^{-1}\hat{\boldsymbol{\beta}}) = \mathbf{x}_t\hat{\boldsymbol{\beta}} = \hat{y}_t$.

(iii) The estimated variance matrix from the regression of $\mathbf{y}$ and $\mathbf{Z}$ is $\tilde{\sigma}^2(\mathbf{Z}'\mathbf{Z})^{-1}$, where $\tilde{\sigma}^2$ is the error variance estimate from this regression. From part (ii), the fitted values from the two regressions are the same, which means the residuals must be the same for all t. (The dependent variable is the same in both regressions.) Therefore, $\tilde{\sigma}^2 = \hat{\sigma}^2$. Further, as we showed in part (i), $(\mathbf{Z}'\mathbf{Z})^{-1} = \mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}')^{-1}$, and so $\tilde{\sigma}^2(\mathbf{Z}'\mathbf{Z})^{-1} = \hat{\sigma}^2\mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}^{-1})'$, which is what we wanted to show.

(iv) The $\tilde{\beta}_j$ are obtained from a regression of $\mathbf{y}$ on $\mathbf{XA}$, where $\mathbf{A}$ is the $k \times k$ diagonal matrix with 1, a_2, …, a_k down the diagonal. From part (i), $\tilde{\boldsymbol{\beta}} = \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}$. But $\mathbf{A}^{-1}$ is easily seen to be the $k \times k$ diagonal matrix with 1, $a_2^{-1}, \ldots, a_k^{-1}$ down its diagonal. Straightforward multiplication shows that the first element of $\mathbf{A}^{-1}\hat{\boldsymbol{\beta}}$ is $\hat{\beta}_1$ and the j^{th} element is $\hat{\beta}_j/a_j$, $j = 2, \ldots, k$.

(v) From part (iii), the estimated variance matrix of $\tilde{\boldsymbol{\beta}}$ is $\hat{\sigma}^2\,\mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}^{-1})'$. But $\mathbf{A}^{-1}$ is a symmetric, diagonal matrix, as described above. The estimated variance of $\tilde{\beta}_j$ is the j^{th} diagonal element of $\hat{\sigma}^2\,\mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}^{-1}$, which is easily seen to be $= \hat{\sigma}^2 c_{jj}/a_j^{-2}$, where c_{jj} is the j^{th} diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. The square root of this, $\hat{\sigma}\sqrt{c_{jj}}/|a_j|$, is se($\tilde{\beta}_j$), which is simply se($\tilde{\beta}_j$)/$|a_j|$.

(vi) The t statistic for $\tilde{\beta}_j$ is, as usual,

$$\tilde{\beta}_j/\text{se}(\tilde{\beta}_j) \;=\; (\hat{\beta}_j/a_j)/[\text{se}(\hat{\beta}_j)/|a_j|],$$

and so the absolute value is $(|\hat{\beta}_j|/|a_j|)/[\text{se}(\hat{\beta}_j)/|a_j|] = |\hat{\beta}_j|/\text{se}(\hat{\beta}_j)$, which is just the absolute value of the t statistic for $\hat{\beta}_j$. If $a_j > 0$, the t statistics themselves are identical; if $a_j < 0$, the t statistics are simply opposite in sign.

E.5 (i) By plugging in for **y**, we can write

$$\tilde{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\mathbf{X}\boldsymbol{\beta}+\mathbf{u}) = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u}.$$

Now we use the fact that **Z** is a function of **X** to pull **Z** outside of the conditional expectation:

$$\text{E}(\tilde{\boldsymbol{\beta}}\,|\,\mathbf{X}) = \boldsymbol{\beta} + \text{E}[(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u}\,|\,\mathbf{X}] = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\text{E}(\mathbf{u}\,|\,\mathbf{X}) = \boldsymbol{\beta}.$$

(ii) We start from the same representation in part (i): $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u}$ and so

$$\begin{aligned}\text{Var}(\tilde{\boldsymbol{\beta}}\,|\,\mathbf{X}) &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'[\text{Var}(\mathbf{u}\,|\,\mathbf{X})]\mathbf{Z}[(\mathbf{Z}'\mathbf{X})^{-1}]' \\ &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\sigma^2\mathbf{I}_n)\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1} = \sigma^2(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1}.\end{aligned}$$

A common mistake is to forget to transpose the matrix $\mathbf{Z}'\mathbf{X}$ in the last term.

(iii) The estimator $\tilde{\boldsymbol{\beta}}$ is linear in **y** and, as shown in part (i), it is unbiased (conditional on **X**). Because the Gauss-Markov assumptions hold, the OLS estimator, $\hat{\boldsymbol{\beta}}$, is the best linear unbiased. In particular, its variance-covariance matrix is "smaller" (in the matrix sense) than $\text{Var}(\tilde{\boldsymbol{\beta}}\,|\,\mathbf{X})$. Therefore, we prefer the OLS estimator.

E.7 (i) Given that the linear model, written in matrix notation,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}$$

satisfies Assumptions E.1, E.2, and E.3.

Partitioned model is

$\mathbf{y} = \mathbf{X_1}\boldsymbol{\beta_1} + \mathbf{X_2}\boldsymbol{\beta_2} + \boldsymbol{u}$, where $\mathbf{X_1}$ **is** $n \times (k_1 + 1)$ and $\mathbf{X_2}$ is $n \times k_2$.

Now, we first regress $\mathbf{y}$ on $\mathbf{X_1}$ and obtain the residuals, say $\ddot{\mathbf{y}}$. Then, we regress $\ddot{\mathbf{y}}$ on $\mathbf{X_2}$ to get $\breve{\boldsymbol{\beta}}_2$.

$\mathbf{E}[\breve{\boldsymbol{\beta}}_2|\mathbf{X}] = \mathbf{E}[(\mathbf{X_2'M_1X_2})^{-1}\mathbf{X_2'M_1y}]$

$\mathbf{E}[\breve{\boldsymbol{\beta}}_2|\mathbf{X}] = \mathbf{E}[((\mathbf{X_2'M_1X_2})^{-1}\mathbf{X_2'M_1}(\mathbf{X_1\beta_1} + \mathbf{X_2\beta_2} + \boldsymbol{u}))|\mathbf{X}]$

$\mathbf{E}[\breve{\boldsymbol{\beta}}_2|\mathbf{X}] = \mathbf{E}[((\mathbf{X_2'M_1X_2})^{-1}\mathbf{X_2'M_1X_1}\boldsymbol{\beta_1} + (\mathbf{X_2'M_1X_2})^{-1}\mathbf{X_2'M_1X_2}\boldsymbol{\beta_2}$
$\quad + (\mathbf{X_2'M_1X_2})^{-1}\mathbf{X_2'M_1}\boldsymbol{u}))|\mathbf{X}]$

$\mathbf{E}[\breve{\boldsymbol{\beta}}_2|\mathbf{X}] = (\mathbf{X_2'M_1X_2})^{-1}\mathbf{X_2'M_1X_1}\mathbf{E}[\boldsymbol{\beta_1}|\mathbf{X}] + \mathbf{E}[\boldsymbol{\beta_2}|\mathbf{X}] + (\mathbf{X_2'M_1X_2})^{-1}\mathbf{X_2'M_1}\mathbf{E}[\boldsymbol{u}|\mathbf{X}]$

$\mathbf{E}[\breve{\boldsymbol{\beta}}_2|\mathbf{X}] = (\mathbf{X_2'M_1X_2})^{-1}\mathbf{X_2'M_1X_1}\mathbf{E}[\boldsymbol{\beta_1}|\mathbf{X}] + \boldsymbol{\beta_2}$ as we know $\mathbf{E}[\boldsymbol{u}|\mathbf{X}] = \mathbf{0}$

$\mathbf{E}[\breve{\boldsymbol{\beta}}_2|\mathbf{X}] = (\mathbf{X_2'X_2})^{-1}\mathbf{X_2'X_1}\boldsymbol{\beta_1} + \boldsymbol{\beta_2}$

Hence we show that $\breve{\boldsymbol{\beta}}_2$ is biased.

(ii) Consider

$\mathbf{y} = \mathbf{X_1}\boldsymbol{\beta_1} + \beta_k\mathbf{X}_{\boldsymbol{k}} + \boldsymbol{u}$**,** where $\mathbf{X}_{\boldsymbol{k}}$ is an $\boldsymbol{n} \times \mathbf{1}$ vector on the variable $\boldsymbol{x_{ik}}$**.**

To show $[\widetilde{\beta_k}|\mathbf{X}] = \left(\frac{\mathbf{SSR}_k}{\sum_{t=1}^{n} x_{tk}^2}\right)\beta_{\boldsymbol{k}}$, we can use part(i) proof .

$(\mathbf{X}_{\boldsymbol{k}}'\mathbf{X}_{\boldsymbol{k}})^{-1}$ results into $\sum_{t=1}^{n} \boldsymbol{x_{tk}^2}$ and we get $(\mathbf{y} - \mathbf{X_1}\boldsymbol{\beta_1} - \beta_k\mathbf{X}_{\boldsymbol{k}})'\ (\mathbf{y} - \mathbf{X_1}\boldsymbol{\beta_1} - \beta_k\mathbf{X}_{\boldsymbol{k}}) = \mathbf{SSR}_{\boldsymbol{k}}$.

The factor multiplying β_k is never greater than one because the numerator is the residual term and the denominator is a x^2 value.

(iii) Similar to part(i), for the regression $\mathbf{y} - \mathbf{X_1}\boldsymbol{\beta_1}$ on $\mathbf{X_2}$,

$\mathbf{E}[\boldsymbol{\beta_2}\mathbf{X}] = (\mathbf{X_2'X_2})^{-1}\mathbf{X_2'X_1}\boldsymbol{\beta_1} + \boldsymbol{\beta_2}$

When $\mathbf{X_2'X_1} = 0$**,** $\boldsymbol{\beta_2}$ becomes an unbiased estimator of $\boldsymbol{\beta_2}$**.**