Solutions for Econometrics I Homework No.2

due 2006-03-08

Feldkircher, Forstner, Pichler, Reiss, Yan, Zeugner

March 13, 2006

Exercise 2.1

Exercise 2.1 is concerned with MATLAB programming in groups of two.

Exercise 2.2

σ-Algebra $\mathcal{F}$:

- collection of subsets of S with the following properties:
  - $\phi \in \mathcal{F}$, $\Omega \in \mathcal{F}$,
  - if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ (closed under complementation)
  - if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} Ai \in \mathcal{F}$ (closed under countable unions)
- same as a field, with the difference that a field may not be closed under countably infinite unions

(Probability) Measure:

- A measure is a σ-additive content which is defined on a σ-algebra.
- content: set function μ defined on a field $\mathcal{F}$ such that
  - $\mu(A) \in [0, \infty)$ whenever $A \in \mathcal{F}$
  - $\mu(\phi) = 0$

- $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$ whenever $A_1, A_2 \in \mathcal{F}$ and $A_1 \cap A_2 = \phi$

- σ-additivity: $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_i^{\infty} \mu(A_i)$ whenever all $A_i \in \mathcal{F}$ and pairwise disjoint

- A measure P is called a probability measure if $P(\Omega) = 1$.

Probability Space:

- if $P|\mathcal{F}$ is a probability measure then $(\Omega, \mathcal{F}, P)$ is called a probability space

- Ω

Measurable Function:

A function $f : (\Omega, \mathcal{A}, \mu) \mapsto (Y, \mathcal{B})$ is called $(\mathcal{A}, \mathcal{B})$-measurable if $f^{-1}(B) \in \mathcal{A} \; \forall \; B \in \mathcal{B}$.

If $f : (\Omega, \mathcal{A}, \mu) \mapsto (Y, \mathcal{B})$ is $(\mathcal{A}, \mathcal{B})$-measurable, then we may define:

$$\mu^f(B) := \mu(f \in B) = \mu(f^{-1}(B))$$

with $B \in \mathcal{B}$. This is the image of μ under f or the distribution of f under μ. **Random Variable:**

- function from a probability space $(\Omega, \mathcal{F}, P)$ to $\mathbf{R}$.

Exercise 2.3

(i) β is the solution to (1), so $X'X\beta = X'y$. And we know that $\gamma = N\beta$,

$$W'W\gamma = W'WN\beta = M'X'X\beta = M'X'y = W'y$$

which means γ is the solution to (2).
(ii) Analogously γ is the solution to (2), so $W'W\gamma = W'y$.
And $\beta = M\gamma$, $X'X\beta = N'W'XM\gamma = N'W'W\gamma = N'W'y = X'y$, which means β is the solution to (1).

(iii) Directly from (i) and (ii), if β is the solution to (1) and γ is the solution to (2), then $\beta = M\gamma$.
Thus $X\beta = XM\gamma = W\gamma$.

Exercise 2.4

Exercise 2.4 is the same as exercise 1.8 in Econometrics problem set 1

Exercise 2.5

Show that the matrix (where we use the notation introduced in class)

$$I_T - \frac{\boldsymbol{11'}}{T}$$

is a projector, where $I_T \in \mathbb{R}^{T\times T}$ is the T-dimensional identity matrix. What is the space that this operator projects on? Consequently, is the matrix $\frac{\boldsymbol{11'}}{T}$ also a projector, and if so, what is it projecting on?

For any $y \in \mathbb{R}^T$ we get

$$\left[I_T - \frac{\mathbf{11'}}{T}\right] y = \begin{pmatrix} 1-\frac{1}{T} & -\frac{1}{T} & \cdots & \cdots & -\frac{1}{T} \\ -\frac{1}{T} & 1-\frac{1}{T} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & -\frac{1}{T} \\ -\frac{1}{T} & \cdots & & -\frac{1}{T} & 1-\frac{1}{T} \end{pmatrix} y = \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ \vdots \\ y_T - \bar{y} \end{pmatrix}$$

where $\bar{y} = \sum_{t=1}^{T} y_t$.

(i) We show that $I_T - \frac{\mathbf{11'}}{T}$ is idempotent:

$$\left(\frac{\mathbf{11'}}{T}\right)\left(\frac{\mathbf{11'}}{T}\right) = \begin{pmatrix} \frac{1}{T} & \cdots & \frac{1}{T} \\ \vdots & \ddots & \vdots \\ \frac{1}{T} & \cdots & \frac{1}{T} \end{pmatrix} \begin{pmatrix} \frac{1}{T} & \cdots & \frac{1}{T} \\ \vdots & \ddots & \vdots \\ \frac{1}{T} & \cdots & \frac{1}{T} \end{pmatrix} = \begin{pmatrix} \frac{T}{T^2} & \cdots & \frac{T}{T^2} \\ \vdots & \ddots & \vdots \\ \frac{T}{T^2} & \cdots & \frac{T}{T^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{T} & \cdots & \frac{1}{T} \\ \vdots & \ddots & \vdots \\ \frac{1}{T} & \cdots & \frac{1}{T} \end{pmatrix} = \frac{\mathbf{11'}}{T}$$

Therefore

$$\left(I_T - \frac{\mathbf{11'}}{T}\right)\left(I_T - \frac{\mathbf{11'}}{T}\right) = I_T - 2\frac{\mathbf{11'}}{T} + \left(\frac{\mathbf{11'}}{T}\right)\left(\frac{\mathbf{11'}}{T}\right) = I_T - 2\frac{\mathbf{11'}}{T} + \frac{\mathbf{11'}}{T} = I_T - \frac{\mathbf{11'}}{T}$$

(ii) We show that $I_T - \frac{\mathbf{11'}}{T}$ is symmetric:

$$\left\langle a, \left(I_T - \frac{\mathbf{11'}}{T}\right) b \right\rangle = \sum_{t=1}^{T} a_t \left(b_t - \bar{b}\right) = \sum_{t=1}^{T} a_t \left(b_t - \frac{1}{T}\sum_{s=1}^{T} b_s\right) = \sum_{t=1}^{T} a_t b_t - \sum_{t=1}^{T} \left(a_t \frac{1}{T} \sum_{s=1}^{T} b_s\right) =$$

$$= \sum_{t=1}^{T} a_t b_t - \left(\sum_{s=1}^{T} b_s\right)\left(\frac{1}{T}\sum_{t=1}^{T} a_t\right) = \sum_{t=1}^{T} a_t b_t - \bar{a}\sum_{s=1}^{T} b_s = \sum_{s=1}^{T} b_s \left(a_s - \bar{a}\right) = \left\langle \left(I_T - \frac{\mathbf{11'}}{T}\right) a, b \right\rangle$$

(iii) We show that $I_T - \frac{\mathbf{11'}}{T}$ projects onto the orthocomplement of $\mathbf{1} = (1, 1, ...1)'$:

(a)

$$\left(I_T - \frac{\mathbf{11'}}{T}\right) y = 0 \Leftrightarrow \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_T - \bar{y} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \Leftrightarrow y_1 = y_2 = \cdots = y_T = \bar{y} \Leftrightarrow y = \bar{y} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

Thus, only vectors that are multiples of $(1, 1, ...1)'$ are orthogonal to the projection space of $I_T - \frac{\mathbf{11'}}{T}$.

(b) For any $y \in \mathbb{R}^T$ we have

$$\left\langle \left(I_T - \frac{\mathbf{11'}}{T}\right) y, \mathbf{1} \right\rangle = \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_T - \bar{y} \end{pmatrix}' \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{t=1}^{T} y_t \ - \ T\bar{y} = \sum_{t=1}^{T} y_t - \frac{T}{T}\sum_{t=1}^{T} y_t = 0$$

Thus the projection of any $y \in \mathbb{R}^T$ by $I_T - \frac{\mathbf{11'}}{T}$ is orthogonal to $\mathbf{1} = (1, 1, ...1)'$.

(iv) We show that for any projector P projecting on the space spanned by the colums of some matrix A, $I - P$ is a projector projecting onto the orthocomplement of the column space of A:

(a)

$$(I - P)(I - P) = I - PI - IP + PP = I - P$$

since P is idempotent. So $I - P$ is idempotent.

(b)

$$\langle a, (I - P) b \rangle = \langle a, b \rangle - \langle a, Pb \rangle = \langle a, b \rangle - \langle Pa, b \rangle = \langle (I - P) a, b \rangle$$

since P is symmetric. So $I - P$ is symmetric.

(c)For any two vectors $y, z \in \mathbb{R}^T$ we obtain

$$\langle Py, (I - P) y \rangle = \langle Py, z - Pz \rangle = \langle Py, z \rangle - \langle Py, Pz \rangle = y'P'z - y'P'Pz = y'P'y - y'P'y = 0$$

since P is idempotent and symmetric. Thus, P and $I - P$ project onto orthogonal spaces.

For our specific application, taking all this together we obtain that $\frac{\mathbf{11'}}{T}$ is a projector onto the space spanned by $\mathbf{1} = (1, 1, ...1)'$.

Exercise 2.6

Consider the estimator under linear restrictions as discussed in class.

(i) Show that the estimator fulfills the restrictions

We hat the following set of restrictions: $R_{m \times k} \beta_{k \times 1} = r_{m \times 1}, rk(R) = m$ and $\hat{\beta}$ the ordinary least squares estimator in the standard regression model.

The restricted least squares estimator derived in class is given by:

$$\beta^* = \hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)$$

Just for convenience and later purposes define $Q := (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}$.

We see that if $\hat{\beta}$ already fulfills the restrictions, the last term will be zero and we end up with our $\hat{\beta}$ estimator. Rewriting the equation yields

$$\beta^* = [I_k - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R]\hat{\beta} + [(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}]r$$

$$\beta^* = \hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R\hat{\beta} + [(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}]r$$

So if we premultiply this equation from the left with R we get:

$$R\beta^* = R\hat{\beta} - R(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R\hat{\beta} + R[(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}]r$$

Hence it remains to show that the part on the right hand side equals r. Consider them by parts, if we start with the last one, we see that

$$R(X'X)^{-1}R'[\underbrace{R_{m\times k}(X'X)^{-1}_{k\times k}R'_{k\times m}}_{m\times m \text{ with full rank m}}]^{-1}]r = r$$

In order to fulfill the equation, the first and the second terms must cancel each other, hence we desire $R(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R\hat{\beta}$ to equal $R\hat{\beta}$.

$$\underbrace{R(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}}_{I_m}R\hat{\beta} = R\hat{\beta}$$

Finally we get

$$R\beta^* = R\hat{\beta} - R\hat{\beta} + r = r.$$

(ii) Compute the variance covariance matrix $\Sigma_{\beta^*\beta^*}$. First compute the expectation of β^*.

$$\begin{aligned}
\mathbf{E}(\beta^*) &= \mathbf{E}(\hat{\beta}) - \mathbf{E}(QR\hat{\beta}) + \mathbf{E}(Qr) \\
&= \beta - QR\beta + Qr \\
&= \beta - Q(R\beta - r) \\
&= \beta + \underbrace{(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\beta)}_{\text{bias if true } \beta \text{ does not fulfill the restrictions}}
\end{aligned}$$

We will later need the expression$(\beta^* - \mathbf{E}\beta^*)$. By plugging in, what we just calculated above we get

$$
\begin{aligned}
(\beta^* - \mathbf{E}\beta^*) &= \hat{\beta} - QR\hat{\beta} + Qr - (\beta - QR\beta + Qr) \\
&= \hat{\beta} - \beta - QR\hat{\beta} + QR\beta \\
&= \hat{\beta} - \beta - QR(\hat{\beta} - \beta)
\end{aligned}
$$

Now we can proceed and calculate the variance-covariance matrix $\Sigma_{\beta^*\beta^*}$.

$$
\begin{aligned}
\Sigma_{\beta^*\beta^*} &= \mathbf{E}[(\beta^* - \mathbf{E}\beta^*)(\beta^* - \mathbf{E}\beta^*)'] \\
&= \mathbf{E}[(\hat{\beta} - QR(\hat{\beta} - \beta))(\hat{\beta} - QR(\hat{\beta} - \beta))'] \\
&= (I_k - QR)\mathbf{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'](I_k - QR)' \\
&= \underbrace{(I_k - QR)\sigma^2(X'X)^{-1}(\hat{\beta} - \beta)'](I_k - QR)'}_{A} \\
&= \sigma^2[(X'X)^{-1} - 2(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1} \\
&+ (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}R' \\
&\quad [R(X'X)^{-1}R']^{-1}R(X'X)^{-1} \\
&= \sigma^2[(X'X)^{-1} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}
\end{aligned}
$$

Martin Wagners Comment: Go directly to equation line "A" by recognizing that the VCV of $\hat{\beta}$ is $\sigma^2(X'X)^{-1}$ and the restricted estimator is a transformation of $\hat{\beta}$ (i.e. $\beta^* = (I_k - QR)\hat{\beta} + Qr$).

(iii) Show that the variance covarinace matrix of the restricetd estimator is smaller or equal than the variance covariance matrix of the unrestricted OLS estimator, when the restrictions are correct. How do you interpret this? Intuitively it is clear that the vcv matrix of the restricted estimator is smaller or equal to that of the ols estimator because we minimize over a smaller set of β's (namely the ones that fulfill the restriction). Also from the derived result in (ii) it can be seen that

$$
\Sigma_{\beta^*\beta^*} = \underbrace{\sigma^2(X'X)^{-1}}_{Var(\hat{\beta})} - \sigma^2(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}.
$$

It remains to show that $\sigma^2(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}$ is a non neg. def. matrix. Multiply the expression by $\alpha \in \mathbf{R}^k$ and its transpose

$$\underbrace{\alpha(X'X)^{-1}R'}_{\gamma'}[R(X'X)^{-1}R']^{-1}\underbrace{R(X'X)^{-1}\alpha}_{\gamma} \overbrace{\geq}^{?} 0 \qquad \forall\, \alpha \;\in \mathbf{R}^k$$

It suffices to show that $\gamma'[R(X'X)^{-1}R']^{-1}\gamma \geq 0$ But we know that X is pos. def., hence also the inverse $(X'X)^{-1}$ is pos.def., consequently $R(X'X)^{-1}R'$ is pos. def, its inverse and finally it holds that

$$\gamma'[R(X'X)^{-1}R']^{-1}\gamma \geq 0.$$

So we have shown that $\Sigma_{\beta^*\beta^*} \leq \Sigma_{\hat{\beta}\hat{\beta}}$. This is always true, even if the restriction is wrong and thus β^* is biased.

Exercise 2.7

Show the following lemma. Under assumptions (A1), (B1) and (B2) it holds that:

(i) $\mathbb{E}(s_{yy}) = \beta' S_{xx}\beta + \frac{T-1}{T}\sigma^2$

(ii) $\mathbb{E}(s_{\widehat{y}\widehat{y}}) = \beta' S_{xx}\beta + \frac{k-1}{T}\sigma^2$

(i) *Show that* $\mathbb{E}(s_{yy}) = \beta' S_{xx}\beta + \frac{T-1}{T}\sigma^2$ *where* $S_{xx} = \frac{1}{T}X'\left(I - \frac{1}{T}\mathbf{1}\mathbf{1}'\right)X$

$$\mathbb{E}(s_{yy}) = \tfrac{1}{T}\mathbb{E}\left((\mathbf{y} - \bar{y}\mathbf{1})'(\mathbf{y} - \bar{y}\mathbf{1})\right) = \tfrac{1}{T}\mathbb{E}\left((\mathbf{y} - \tfrac{1}{T}\mathbf{1}\mathbf{y}\mathbf{1})'(\mathbf{y} - \tfrac{1}{T}\mathbf{1}\mathbf{y}\mathbf{1})\right) =$$

$$= \tfrac{1}{T}\left[\mathbb{E}(\mathbf{y}'\mathbf{y}) - \mathbb{E}(\mathbf{y}'\mathbf{1}\mathbf{1}'\mathbf{y}\tfrac{1}{T}) - \mathbb{E}(\mathbf{y}'\mathbf{1}\mathbf{1}'\mathbf{y}\tfrac{1}{T}) + \mathbb{E}(\mathbf{1}'\mathbf{y}\underbrace{\mathbf{1}'\mathbf{1}}_{T}\mathbf{y}'\mathbf{1}\tfrac{1}{T^2})\right]$$

$$= \tfrac{1}{T}\left[\mathbb{E}(\mathbf{y}'\mathbf{y}) - \tfrac{1}{T}\mathbb{E}(\mathbf{y}'\mathbf{1}\mathbf{1}'\mathbf{y})\right] = \tfrac{1}{T}\mathbb{E}(\mathbf{y}'(I - \tfrac{1}{T}\mathbf{1}\mathbf{1}')\mathbf{y}) =$$

$$= \tfrac{1}{T}\mathbb{E}\left((\beta'X' + \mathbf{u}')(I - \tfrac{1}{T}\mathbf{1}\mathbf{1}')(X\beta + \mathbf{u})\right) =$$

$$= \tfrac{1}{T}\mathbb{E}\left(\underbrace{\beta'X'(I - \tfrac{1}{T}\mathbf{1}\mathbf{1}')X\beta}_{\beta'S_{xx}\beta} + 2\beta'X'(I - \tfrac{1}{T}\mathbf{1}\mathbf{1}')\mathbf{u} + \mathbf{u}'(I - \tfrac{1}{T}\mathbf{1}\mathbf{1}')\mathbf{u}\right) =$$

$$= \beta'S_{xx}\beta + \tfrac{1}{T}\mathbb{E}\left(\mathbf{u}'(I - \tfrac{1}{T}\mathbf{1}\mathbf{1}')\mathbf{u}\right) = \beta'S_{xx}\beta + \underbrace{\tfrac{1}{T}\mathbb{E}(\mathbf{u}'\mathbf{u})}_{\sigma^2} - \tfrac{1}{T^2}\mathbb{E}(\underbrace{\mathbf{u}'\mathbf{1}\mathbf{1}'\mathbf{u}}_{=\mathbf{1}'\mathbf{u}\mathbf{u}'\mathbf{1}=\mathbf{1}'\sigma^2 I\mathbf{1}}) =$$

$$= \beta'S_{xx}\beta + \sigma^2 - \tfrac{1}{T^2}\sigma^2\underbrace{\mathbf{1}'I\mathbf{1}}_{T} =$$

$$= \beta'S_{xx}\beta + \frac{T-1}{T}\sigma^2$$

(ii) *Show that* $\mathbb{E}(s_{\widehat{y}\widehat{y}}) = \beta' S_{xx}\beta + \frac{k-1}{T}\sigma^2$ *where* $S_{xx} = \frac{1}{T}X'\left(I - \frac{1}{T}\mathbf{1}\mathbf{1}'\right)X$

If β includes a constant, i.e.X includes the one-vector $\mathbf{1}$, we know from the fact $\widehat{\mathbf{u}}'X = \mathbf{y}'X - \widehat{\mathbf{y}}'X = \mathbf{y}'X - y'XX^+X = \mathbf{0}$ that the sum of residuals $\widehat{\mathbf{u}}'\mathbf{1} = 0$ and thus $\widehat{\bar{\mathbf{u}}} = 0$. This implies $\mathbf{1}'\mathbf{y} = \mathbf{1}'\mathbf{y} - \mathbf{1}'\widehat{\mathbf{u}} = \mathbf{1}'\widehat{\mathbf{y}} \;\Rightarrow\; \bar{\mathbf{y}} = \widehat{\bar{\mathbf{y}}}$. (Moreover note the following: $\widehat{\mathbf{y}} = \mathbf{y} - \widehat{\mathbf{u}}$, hence $\widehat{\mathbf{u}} = (I - XX^+)\mathbf{u}$, and thus $\mathbb{E}(\widehat{\mathbf{u}}) = \mathbf{0}$.)

$$\mathbb{E}(s_{\widehat{y}\widehat{y}}) = \tfrac{1}{T}\mathbb{E}((\widehat{\mathbf{y}} - \bar{\mathbf{y}}\mathbf{1})'(\widehat{\mathbf{y}} - m\bar{x}y\mathbf{1})) = \mathbb{E}\left(((\mathbf{y} - \bar{\mathbf{y}}\mathbf{1}) - \widehat{\mathbf{u}})'((\mathbf{y} - \bar{\mathbf{y}}\mathbf{1}) - \widehat{\mathbf{u}})\right) =$$

$$= \tfrac{1}{T}\mathbb{E}((\mathbf{y} - \bar{\mathbf{y}}\mathbf{1})'(\mathbf{y} - \bar{\mathbf{y}}\mathbf{1})) - \underbrace{\tfrac{1}{T}2\mathbb{E}(\widehat{\mathbf{u}}'(\mathbf{y} - \bar{\mathbf{y}}\mathbf{1}))}_{=\frac{2}{T}\mathbb{E}(\widehat{\mathbf{u}}'(\widehat{\mathbf{y}}+\widehat{\mathbf{u}}-\bar{\mathbf{y}}\mathbf{1}))=\frac{2}{T}\mathbb{E}(\widehat{\mathbf{u}}'\widehat{\mathbf{u}})\text{ since }\widehat{\mathbf{u}}'\widehat{\mathbf{y}}=0} + \tfrac{1}{T}\mathbb{E}(\widehat{\mathbf{u}}'\widehat{\mathbf{u}})$$

$$= \mathbb{E}(s_{yy}) - \tfrac{1}{T}\mathbb{E}(\widehat{\mathbf{u}}'\widehat{\mathbf{u}})$$

And now, what is $\mathbb{E}\left(\widehat{\mathbf{u}}'\widehat{\mathbf{u}}\right)$? (Note that $\widehat{\mathbf{u}} = (I - XX^+)\mathbf{u}$.)

$$\begin{aligned}\mathbb{E}\left(\widehat{\mathbf{u}}'\widehat{\mathbf{u}}\right) &= \mathbb{E}\left(\mathbf{u}'(I - XX^+)(I - XX^+)\mathbf{u}\right) = \mathbb{E}\left(\mathbf{u}'(I - XX^+)\mathbf{u}\right) \\ &= \mathbb{E}\left(\operatorname{tr}(\mathbf{u}'(I - XX^+)\mathbf{u})\right) = \operatorname{tr}(\sigma^2 I_T(I_T - XX^+)) = \sigma^2(\operatorname{tr}(I_T) - \operatorname{tr}(XX^+))\end{aligned}$$

By the properties of XX^+ as a projector, we know that its trace equals the sum of its eigenvalues. Furthermore its eigenvalues can only be either 1 or 0, where there are as many ones within the eigenvalues as the rank of XX^+. This implies $\operatorname{tr}(XX^+) = \operatorname{rank}(XX^+)$. Let's define the rank of XX^+ to be k. Hence $(\operatorname{tr}(I) - \operatorname{tr}(XX^+)) = T - k$. Therefore:

$$\mathbb{E}\left(\widehat{\mathbf{u}}'\widehat{\mathbf{u}}\right) = \sigma^2(T - k)$$

Now plugging in this expression into the one for $\mathbb{E}(s_{\widehat{y}\widehat{y}})$ leads to:

$$\begin{aligned}\mathbb{E}(s_{\widehat{y}\widehat{y}}) = \mathbb{E}(s_{yy}) - \tfrac{1}{T}\mathbb{E}(\widehat{\mathbf{u}}'\widehat{\mathbf{u}}) = \beta' S_x x \beta + \frac{T-1}{T}\sigma^2 - \frac{T-k}{T}\sigma^2 = \\ = \beta' S_x x \beta + \frac{k-1}{T}\sigma^2\end{aligned}$$

Exercise 2.8

(i+ii) Corrected version:

Regressing the matrix X on regressors corresponding to the fixed effects α_i, where this large regressor-matrix takes the following form:

$$X_\alpha = \begin{pmatrix} \mathbf{1} & & & \\ & \mathbf{1} & & \\ & & \ddots & \\ & & & \mathbf{1} \end{pmatrix} = I_N \otimes \mathbf{1}$$

We hence receive the estimator for this regression by

$$\begin{aligned}
\hat{\beta}_X &= (X'_\alpha X_\alpha)^{-1} X'_\alpha X \\
&= ((I_N \otimes \mathbf{1})'(I_N \otimes \mathbf{1}))^{-1} (I_N \otimes \mathbf{1})' X \\
&= ((I_N \otimes \mathbf{1}')(I_N \otimes \mathbf{1}))^{-1} (I_N \otimes \mathbf{1}') X \\
&= (I_N \otimes \mathbf{1}'\mathbf{1})^{-1} (I_N \otimes \mathbf{1}') X \\
&= (I_N \otimes T)^{-1} (I_N \otimes \mathbf{1}') X \\
&= (I_N \otimes \frac{1}{T})(I_N \otimes \mathbf{1}') X \\
&= (I_N \otimes \frac{1}{T}\mathbf{1}') X
\end{aligned}$$

By the same computations, we receive the estimator for the regression of Y on X_α as

$$\hat{\beta}_Y = (I_N \otimes \frac{1}{T}\mathbf{1}')Y$$

For the residuals related to these two regressions, we receive

$$\begin{aligned}
\tilde{X} &= X - \hat{X} \\
&= X - X_\alpha \hat{\beta}_X \\
&= X - (I_N \otimes \mathbf{1})(I_N \otimes \frac{1}{T}\mathbf{1}') X \\
&= X - (I_N \otimes \frac{1}{T}\mathbf{1}\mathbf{1}') X \\
&= (I_{NT} - I_N \otimes \frac{1}{T}\mathbf{1}\mathbf{1}') X \\
&= (I_N \otimes I_T - I_N \otimes \frac{1}{T}\mathbf{1}\mathbf{1}') X \\
&= (I_N \otimes (I_T - \frac{1}{T}\mathbf{1}\mathbf{1}')) X
\end{aligned}$$

$$\tilde{X} = \begin{pmatrix} \left(I_T - \begin{pmatrix} \frac{1}{T} \cdots \frac{1}{T} \\ \vdots \\ \frac{1}{T} \cdots \frac{1}{T} \end{pmatrix}\right) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \left(I_T - \begin{pmatrix} \frac{1}{T} \cdots \frac{1}{T} \\ \vdots \\ \frac{1}{T} \cdots \frac{1}{T} \end{pmatrix}\right) \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ \vdots \\ \vdots \\ X_N \end{pmatrix}$$

$$= \begin{pmatrix} X_1 - \bar{X}_1 \\ \vdots \\ X_N - \bar{X}_N \end{pmatrix}$$

And by the same computations:

$$\tilde{Y} = \begin{pmatrix} Y_1 - \bar{Y}_1 \\ \vdots \\ Y_N - \bar{Y}_N \end{pmatrix}$$

So, we see that the residuals are just the demeaned observations, with demeaning occurring for each i separately.

The next step is to make use of the Frisch-Waugh-Theorem and use these residuals to compute $\hat{\beta}$:

$\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$

(iii) The projectors corresponding to the fixed effects α_i are the projectors onto the ortho-complement of span(1)

$$\begin{aligned} M_1 &= (I_T - P_1) \\ &= I_T - 1_T 1_T^+ \\ &= I_T - 1_T(1_T' 1_T)^{-1} 1_T' \\ &= I_T - \frac{1}{T} 1_T 1_T' \\ &= I_T - \begin{pmatrix} \frac{1}{T} \cdots \frac{1}{T} \\ \vdots \\ \frac{1}{T} \cdots \frac{1}{T} \end{pmatrix} \end{aligned}$$

So, again, from this notation, one can immediately see, that the regression on α_i performs a de-meaning of the projected matrix of suitable size, $T \times M$ (M arbitrary). Now, one could either proceed and apply this regressor on all y_i and X_i and afterwards stack together the N residual matrices to compute a pooled estimate of β, or create a stacked-projecotr of suitable size and apply it on the stacked $y = (y_1 \ldots y_N)'$ and $X = (X_1 \ldots X_N)'$ matrices. Both methods are of course equivalent. Choosing the second option, the suitable projecotr is obtained by $(1_N \otimes M_1) \in \mathbb{R}^{T \times NT}$.

$$(1_N \otimes M_1)y = \tilde{y} = (\tilde{y_1}, \ldots, \tilde{y_1})' \in \mathbb{R}^{NT \times 1} \tag{1}$$

$$(1_N \otimes M_1)X = \tilde{X} = (\tilde{X_1}, \ldots, \tilde{X_1})' \in \mathbb{R}^{NT \times k} \tag{2}$$

From these computations, we can see, that the procedures in sub-points (i) and (iii) yield exactly the same results.

Again applying the Frisch-Waugh-Theorem we receive exactly the same estimators as before:

$\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$

(iv) Let's consider a regression with constant for arbitrary dimension $T \times k$.

$$X_i = \begin{bmatrix} 1 & x_{11} & \dots & x_{1(k-1)} \\ \vdots & \vdots & & \vdots \\ 1 & x_{T1} & \dots & x_{t(k-1)} \end{bmatrix}$$

$$X_i'X_i = \begin{bmatrix} T & \sum_{t=1}^T x_{1t} & \dots & \dots & \sum_{t=1}^T x_{(k-1)t} \\ \sum_{t=1}^T x_{1t} & \sum_{t=1}^T x_{1t}^2 & \sum_{t=1}^T x_{1t}x_{2t} & \dots & \sum_{t=1}^T x_{1t}x_{(k-1)t} \\ \vdots & \sum_{t=1}^T x_{2t}x_{1t} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \sum_{t=1}^T x_{(k-1)t} & \sum_{t=1}^T x_{(k-1)t}x_{1t} & \dots & & \sum_{t=1}^T x_{(k-1)t}^2 \end{bmatrix}$$

Now consider the normal equations $X_i'X_i\beta = X_i'y_i$. This yields the first row

$$T\alpha + \sum_{t=1}^{T} x_{1t}\beta_1 \dots + \sum_{t=1}^{T} x_{(k-1)t}\beta_{k-1} = \sum_{t=1}^{T} y_t \ / : T$$

$$\hat{\alpha} + \bar{\hat{x}}_1\beta_1 + \dots + x_{\bar{\hat{k}}-1}\beta_{k-1} = \bar{\hat{y}}$$

$$\hat{\alpha} = \bar{\hat{y}} - \sum_{i=1}^{k-1} \bar{\hat{x}}_i\beta_i$$

$$\hat{\alpha} = \bar{\hat{y}} - \bar{\hat{x}}\hat{\beta}$$

as for any inhomogeneous regression $\bar{\hat{y}} = \bar{y}$ and $\bar{\hat{x}} = \bar{x}$, as shown in class. Hence, the regression line passes through the empirical means of the dependent and explanatory variables. In our case, we can thus take the values already derived for the $\bar{x}_i$ and $\bar{y}_i$ and calculate:

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i\hat{\beta} = P_1 y_i - P_1 x_i\hat{\beta} = P_1(y_i - x_i\hat{\beta})$$
$$\Rightarrow \hat{\alpha}_i = \overline{(y_i - x_i\hat{\beta})}$$

(v) As we estimated k parameters and computed N means, the correct degree of freedom for the estimation of σ^2 is $NT - k - N$. So $\hat{\sigma}^2 = \frac{\tilde{u}'\tilde{u}}{NT-k-N}$.

(v) As we estimated k parameters and computed N means, the correct degree of freedom for the estimation of σ^2 is $NT - k - N$. So $\hat{\sigma}^2 = \frac{\tilde{u}'\tilde{u}}{NT-k-N}$.

(iv) Let's consider a regression with constant for arbitrary dimension $T \times k$.

$$X_i = \begin{bmatrix} 1 & x_{11} & \dots & x_{1(k-1)} \\ \vdots & \vdots & & \vdots \\ 1 & x_{T1} & \dots & x_{t(k-1)} \end{bmatrix}$$

$$X_i'X_i = \begin{bmatrix} T & \sum_{t=1}^T x_{1t} & \dots & \dots & \sum_{t=1}^T x_{(k-1)t} \\ \sum_{t=1}^T x_{1t} & \sum_{t=1}^T x_{1t}^2 & \sum_{t=1}^T x_{1t}x_{2t} & \dots & \sum_{t=1}^T x_{1t}x_{(k-1)t} \\ \vdots & \sum_{t=1}^T x_{2t}x_{1t} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \sum_{t=1}^T x_{(k-1)t} & \sum_{t=1}^T x_{(k-1)t}x_{1t} & \dots & & \sum_{t=1}^T x_{(k-1)t}^2 \end{bmatrix}$$

Now consider the normal equations $X_i'X_i\beta = X_i'y_i$. This yields the first row

$$T\alpha + \sum_{t=1}^T x_{1t}\beta_1 \ldots + \sum_{t=1}^T x_{(k-1)t}\beta_{k-1} = \sum_{t=1}^T y_t \ / : T$$

$$\hat{\alpha} + \bar{\hat{x}}_1\beta_1 + \ldots + x\bar{\hat{_{k-1}}}\beta_{k-1} = \bar{\hat{y}}$$
$$\hat{\alpha} = \bar{\hat{y}} - \sum_{i=1}^{k-1} \bar{\hat{x}}_i\beta_i$$
$$\hat{\alpha} = \bar{\hat{y}} - \bar{\hat{x}}\hat{\beta}$$

as for any inhomogeneous regression $\bar{\hat{y}} = \bar{y}$ and $\bar{\hat{x}} = \bar{x}$, as shown in class. Hence, the regression line passes through the empirical means of the dependent and explanatory variables. In our case, we can thus take the values already derived for the $\bar{x}_i$ and $\bar{y}_i$ and calculate:

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i\hat{\beta} = P_1 y_i - P_1 x_i\hat{\beta} = P_1(y_i - x_i\hat{\beta})$$
$$\Rightarrow \hat{\alpha}_i = \overline{(y_i - x_i\hat{\beta})}$$