Question 1

(a)

Set $\boldsymbol{y} = (y_{11}, .., y_{1T}, y_{21}, ..., y_{2T}, ..., y_{N1}, ..., y_{NT})'$, $\boldsymbol{\varepsilon} = (\varepsilon_{11}, .., \varepsilon_{1T}, \varepsilon_{21}, ..., \varepsilon_{2T}, ..., \varepsilon_{N1}, ..., \varepsilon_{NT})'$ and $\boldsymbol{X} = [\boldsymbol{x}_{11}, ..., \boldsymbol{x}_{1T}, \boldsymbol{x}_{21}, ..., \boldsymbol{x}_{2T}, ..., \boldsymbol{x}_{N1}, ..., \boldsymbol{x}_{NT}]'$, where $\boldsymbol{x}_{it} = (x_{it,1}, ..., x_{it,k})'$.

The linear regression model is given as

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it},\ i = 1, ..., N; t = 1, ..., T,$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)'$.

Then, let $\boldsymbol{g} = (\alpha_{11}, .., \alpha_{1T}, \alpha_{21}, ..., \alpha_{2T}, ..., \alpha_{N1}, ..., \alpha_{NT})'$, where $\alpha_{it} = \alpha_i$ for $t = 1, .., T$, the model can be expressed in matrix vector form as

$$\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

It follows that the vector $\boldsymbol{g}$ can be further expressed as

$$\boldsymbol{g} = \boldsymbol{D}\boldsymbol{\alpha},$$

where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$ and

$$\boldsymbol{D} = [\boldsymbol{d}_1, .., \boldsymbol{d}_N] = \begin{bmatrix} \mathbf{1}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_T \end{bmatrix}.$$

Therefore, the matrix form of the model can be expressed as

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

(b)

Suppose that $\boldsymbol{X}$ has full column rank. By the Frisch-Waugh-Lovell Theorem, the OLS estimate of coefficient $\boldsymbol{\beta}$ of the model in (a) is

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{R_d}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R_d}\boldsymbol{y},$$

where $\boldsymbol{R_d} = \boldsymbol{I}_{NT} - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'$.

Since the selection matrix is

$$\boldsymbol{D} = \begin{bmatrix} \mathbf{1}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_T \end{bmatrix},$$

then

$$\boldsymbol{D'D} = \begin{bmatrix} \mathbf{1}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_T \end{bmatrix}' \begin{bmatrix} \mathbf{1}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_T \end{bmatrix} = \begin{bmatrix} T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & T \end{bmatrix} = T\boldsymbol{I}_N.$$

It follows that the $NT \times NT$ projection matrix is

$$\boldsymbol{D}(\boldsymbol{D'D})^{-1}\boldsymbol{D'} = \begin{bmatrix} \mathbf{1}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_T \end{bmatrix} T^{-1}\boldsymbol{I}_N \begin{bmatrix} \mathbf{1}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_T \end{bmatrix}' = T^{-1} \begin{bmatrix} \boldsymbol{J}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{J}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{J}_T \end{bmatrix},$$

where $\boldsymbol{J}_T$ is a $T \times T$ unit matrix.

Hence, by applying the residual operator $\boldsymbol{R_d}$ on $\boldsymbol{X}$, we have

$$\boldsymbol{R_d X} = \left(\boldsymbol{I}_{NT} - T^{-1} \begin{bmatrix} \boldsymbol{J}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{J}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{J}_T \end{bmatrix} \right) \boldsymbol{X} = \boldsymbol{X} - \begin{bmatrix} \bar{\boldsymbol{X}}_1 \\ \bar{\boldsymbol{X}}_2 \\ \vdots \\ \bar{\boldsymbol{X}}_N \end{bmatrix},$$

where $\bar{\boldsymbol{X}}_i = [T^{-1}\sum_{t=1}^T \boldsymbol{x}_{it}, T^{-1}\sum_{t=1}^T \boldsymbol{x}_{it}, ..., T^{-1}\sum_{t=1}^T \boldsymbol{x}_{it}]'$ is a $T \times K$ matrix.

Similarly, by applying the residual operator $\boldsymbol{R_d}$ on $\boldsymbol{y}$, we have

$$\boldsymbol{R_d y} = \left(\boldsymbol{I}_{NT} - T^{-1} \begin{bmatrix} \boldsymbol{J}_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{J}_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{J}_T \end{bmatrix} \right) \boldsymbol{y} = \boldsymbol{y} - \begin{bmatrix} \bar{\boldsymbol{y}}_1 \\ \bar{\boldsymbol{y}}_2 \\ \vdots \\ \bar{\boldsymbol{y}}_N \end{bmatrix},$$

where $\bar{\boldsymbol{y}}_i = (T^{-1}\sum_{t=1}^T y_{it}, T^{-1}\sum_{t=1}^T y_{it}, ..., T^{-1}\sum_{t=1}^T y_{it})'$ is a length T vector.

Because $\boldsymbol{R_d}$ is a idempotent matrix, which means $\boldsymbol{R_d} = \boldsymbol{R_d}' = \boldsymbol{R_d}'\boldsymbol{R_d} = \boldsymbol{R_d R_d}'$, the regression coefficient $\boldsymbol{b}$ can be expressed as

$$\boldsymbol{b} = (\boldsymbol{X'R_dR_dX})^{-1}\boldsymbol{X'R_dR_dy},$$

which is the solution by regressing $\{y_{it} - \bar{y}_i\}$ on $\{\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i\}$, where $\bar{y}_i = T^{-1}\sum_{i=1}^T y_{it}$ and $\bar{\boldsymbol{x}}_i = T^{-1}\sum_{i=1}^T \boldsymbol{x}_{it}$.

(c)

Suppose that $\mathbb{E}[\boldsymbol{\varepsilon}_t|\boldsymbol{X}_t] = \mathbf{0}$ and $\mathbb{E}[\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t|\boldsymbol{X}_t] = \sigma^2\boldsymbol{I}_N$ where $\boldsymbol{X}_t = (\boldsymbol{x}_{1t}, ..., \boldsymbol{x}_{Nt})'$ and $\boldsymbol{\varepsilon}_t = (\boldsymbol{\varepsilon}_{1t}, ..., \boldsymbol{\varepsilon}_{Nt})'$.

If the pairs $(\boldsymbol{X}_t, \boldsymbol{\varepsilon}_t)$, $t = 1, ..., T$, are independently distributed, then

$$\mathbb{E}[\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}|\boldsymbol{X}] = \begin{bmatrix} \mathbb{E}[\boldsymbol{\varepsilon}_1'\boldsymbol{\varepsilon}_1|\boldsymbol{X}] & \mathbb{E}[\boldsymbol{\varepsilon}_1'\boldsymbol{\varepsilon}_2|\boldsymbol{X}] & \cdots & \mathbb{E}[\boldsymbol{\varepsilon}_1'\boldsymbol{\varepsilon}_T|\boldsymbol{X}] \\ \mathbb{E}[\boldsymbol{\varepsilon}_2'\boldsymbol{\varepsilon}_1|\boldsymbol{X}] & \mathbb{E}[\boldsymbol{\varepsilon}_2'\boldsymbol{\varepsilon}_2|\boldsymbol{X}] & \cdots & \mathbb{E}[\boldsymbol{\varepsilon}_2'\boldsymbol{\varepsilon}_T|\boldsymbol{X}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[\boldsymbol{\varepsilon}_T'\boldsymbol{\varepsilon}_1|\boldsymbol{X}] & \mathbb{E}[\boldsymbol{\varepsilon}_T'\boldsymbol{\varepsilon}_2|\boldsymbol{X}] & \cdots & \mathbb{E}[\boldsymbol{\varepsilon}_T'\boldsymbol{\varepsilon}_T|\boldsymbol{X}] \end{bmatrix} = \begin{bmatrix} \sigma^2\boldsymbol{I}_N & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma^2\boldsymbol{I}_N & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma^2\boldsymbol{I}_N \end{bmatrix} = \sigma^2\boldsymbol{I}_{NT},$$

because for $i \neq j$, $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\varepsilon}_j$ are independent given $\boldsymbol{X}$, $\mathbb{E}[\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_j|\boldsymbol{X}] = \mathbf{0}$.

(d)

Under the assumptions stated in (a), (b) and (c), there are N regressors in $\boldsymbol{alpha}$ and K regressors in $\boldsymbol{X}$. The number of observations is NT. Therefore, using the Thereon 1.4 from Lecture note I, we have the unbiased OLS estimator of σ^2 denoted by s^2

$$s^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{NT - N - K},$$

where $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{D}\boldsymbol{a} - \boldsymbol{X}\boldsymbol{b}$.

By adding $\boldsymbol{D}'$ on both side of the regression equation, we have

$$\boldsymbol{D}'\boldsymbol{y} = \boldsymbol{D}'\boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{D}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}'\boldsymbol{\varepsilon}$$

Question 2

(a)

$$\prod_{i=1}^{n} \binom{k}{r_i} p_d^{r_i} (1 - p_d)^{k - r_i}$$

$$L = \prod_{i=1}^{n} p^{\mathbb{1}(r_i = 0)} (1 - p)^{\mathbb{1}(r_i > 0)}$$