1. Tenemos que el estimador LS satisface

$$\boldsymbol{X}^\top \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{Y},$$

luego, podemos escribir

$$\widehat{\boldsymbol{\beta}}(d) = (\boldsymbol{X}^\top \boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}^\top \boldsymbol{X}\widehat{\boldsymbol{\beta}} + d\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top \boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}^\top \boldsymbol{X} + d\boldsymbol{I})\widehat{\boldsymbol{\beta}}.$$

De este modo,

$$\mathsf{E}(\widehat{\boldsymbol{\beta}}(d)) = (\boldsymbol{X}^\top \boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}^\top \boldsymbol{X} + d\boldsymbol{I})\,\mathsf{E}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top \boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}^\top \boldsymbol{X} + d\boldsymbol{I})\boldsymbol{\beta}.$$

Es decir, para $d \in [0,1)$ sigue que $\widehat{\boldsymbol{\beta}}(d)$ es un estimador sesgado. Por otro lado,

$$\begin{aligned}
\mathsf{Cov}(\widehat{\boldsymbol{\beta}}(d)) &= \mathsf{Cov}\{(\boldsymbol{X}^\top \boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}^\top \boldsymbol{X} + d\boldsymbol{I})\widehat{\boldsymbol{\beta}}\} \\
&= (\boldsymbol{X}^\top \boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}^\top \boldsymbol{X} + d\boldsymbol{I})\,\mathsf{Cov}(\widehat{\boldsymbol{\beta}})(\boldsymbol{X}^\top \boldsymbol{X} + d\boldsymbol{I})(\boldsymbol{X}^\top \boldsymbol{X} + \boldsymbol{I})^{-1} \\
&= \sigma^2(\boldsymbol{X}^\top \boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}^\top \boldsymbol{X} + d\boldsymbol{I})(\boldsymbol{X}^\top \boldsymbol{X})^{-1}(\boldsymbol{X}^\top \boldsymbol{X} + d\boldsymbol{I})(\boldsymbol{X}^\top \boldsymbol{X} + \boldsymbol{I})^{-1}.
\end{aligned}$$

Se debe notar que esto permite calcular $\mathsf{MSE}(\widehat{\boldsymbol{\beta}}(d))$ lo que ofrece un procedimiento para seleccionar d.

2. Tenemos $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ con $\boldsymbol{\epsilon} \sim \mathsf{N}_n(\boldsymbol{0}, \sigma^2\boldsymbol{\Omega})$, donde

$$\boldsymbol{\Omega} = (1-\rho)\boldsymbol{I} + \rho\boldsymbol{1}\boldsymbol{1}^\top.$$

Recordando que σ^2 y ρ son conocidos, podemos obtener el estimador mínimos cuadrados generalizados (GLS), como:

$$\widehat{\boldsymbol{\beta}}_{\mathsf{GLS}} = (\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y}, \qquad \boldsymbol{\Omega}^{-1} = \frac{1}{1-\rho}\Big\{\boldsymbol{I} - \frac{\rho}{1+(n-1)\rho}\boldsymbol{1}\boldsymbol{1}^\top\Big\}.$$

Ahora, $\boldsymbol{X} = (\boldsymbol{1}, \boldsymbol{Z})$ y

$$\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X} = \begin{pmatrix} \boldsymbol{1}^\top \\ \boldsymbol{Z}^\top \end{pmatrix}\boldsymbol{\Omega}^{-1}(\boldsymbol{1}, \boldsymbol{Z}) = \begin{pmatrix} \boldsymbol{1}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{1} & \boldsymbol{1}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{1} & \boldsymbol{Z}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Z} \end{pmatrix}.$$

Note que

$$\boldsymbol{Z}^\top\boldsymbol{\Omega}^{-1} = \frac{1}{1-\rho}\boldsymbol{Z}^\top\Big\{\boldsymbol{I} - \frac{\rho}{1+(n-1)\rho}\boldsymbol{1}\boldsymbol{1}^\top\Big\} = \frac{1}{1-\rho}\boldsymbol{Z}^\top.$$

De ahí que $\boldsymbol{Z}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{1} = \boldsymbol{0}$. Por otro lado,

$$\begin{aligned}
\boldsymbol{1}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{1} &= \frac{1}{1-\rho}\boldsymbol{1}^\top\Big\{\boldsymbol{I} - \frac{\rho}{1+(n-1)\rho}\boldsymbol{1}\boldsymbol{1}^\top\Big\}\boldsymbol{1} = \frac{n}{1-\rho}\Big(1 - \frac{\rho n}{1+(n-1)\rho}\Big) \\
&= \frac{n}{1+(n-1)\rho}.
\end{aligned}$$

Análogamente, tenemos que

$$\begin{aligned}\mathbf{1}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y} &= \frac{1}{1-\rho}\mathbf{1}^\top\Big\{\boldsymbol{I} - \frac{\rho}{1+(n-1)\rho}\mathbf{1}\mathbf{1}^\top\Big\}\boldsymbol{Y} = \frac{n\overline{Y}}{1-\rho}\Big(1 - \frac{\rho n}{1+(n-1)\rho}\Big)\\ &= \frac{n\overline{Y}}{1+(n-1)\rho}\end{aligned}$$

$$\widehat{\boldsymbol{\beta}}_{\mathsf{GLS}} = \begin{pmatrix}\widehat{\delta}\\ \widehat{\boldsymbol{\theta}}\end{pmatrix} = (\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top = \begin{pmatrix}\mathbf{1}^\top\boldsymbol{\Omega}^{-1}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{Z}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Z}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{1}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y}\\ \boldsymbol{Z}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y}\end{pmatrix}.$$

Finalmente,

$$\widehat{\delta} = (\mathbf{1}^\top\boldsymbol{\Omega}^{-1}\mathbf{1})^{-1}\mathbf{1}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y} = \overline{Y},$$

y

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{Z}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{Y} = (1-\rho)(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\frac{1}{1-\rho}\boldsymbol{Z}^\top\boldsymbol{Y} = (\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{Y}.$$

3.a. Note que, podemos escribir

$$(n-p-1)s^2 = \sum_{j\neq i}^n (Y_j - \boldsymbol{x}_j^\top\widehat{\boldsymbol{\beta}}_{(i)})^2 = \sum_{j=1}^n (Y_j - \boldsymbol{x}_j^\top\widehat{\boldsymbol{\beta}}_{(i)})^2 - (Y_i - \boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}_{(i)})^2.$$

Usando que

$$\widehat{\boldsymbol{\beta}}_{(i)} = \widehat{\boldsymbol{\beta}} - \frac{e_i}{1-h_{ii}}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{x}_i.$$

Esto nos permite calcular,

$$Y_j - \boldsymbol{x}_j^\top\widehat{\boldsymbol{\beta}}_{(i)} = e_j + \frac{e_i}{1-h_{ii}}\boldsymbol{x}_j^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{x}_i = e_j + \frac{e_i h_{ij}}{1-h_{ii}}.$$

De este modo,

$$\begin{aligned}(n-p-1)s_{(i)}^2 &= \sum_{j=1}^n\Big(e_j + \frac{e_i h_{ij}}{1-h_{ii}}\Big)^2 - \Big(\frac{e_i}{1-h_{ii}}\Big)^2\\ &= \sum_{j=1}^n e_j^2 + \frac{2e_i}{1-h_{ii}}\sum_{j=1}^n e_j h_{ij} + \Big(\frac{e_i}{1-h_{ii}}\Big)^2\sum_{j=1}^n h_i j^2 - \Big(\frac{e_i}{1-h_{ii}}\Big)^2\\ &= (n-p)s^2 + \Big(\frac{e_i}{1-h_{ii}}\Big)^2 h_{ii} - \Big(\frac{e_i}{1-h_{ii}}\Big)^2 = (n-p)s^2 - \frac{e_i^2}{1-h_{ii}},\end{aligned}$$

pues $\boldsymbol{He} = \boldsymbol{H}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y} = \mathbf{0}$, y $\boldsymbol{H}^2 = \boldsymbol{H}$, de ahí que $\sum_{i=1}^n h_{ij}^2 = h_{ii}$. Por otro lado, sabemos que el residuo estandarizado es definido como,

$$r_i = \frac{e_i}{s\sqrt{1-h_{ii}}},$$

lo que permite escribir

$$(n-p-1)s^2 = (n-p)s^2 + r_i^2 s^2 = s^2(n-p-r_i^2),$$

y el resultado sigue.

3.b. Como $\boldsymbol{e} \sim \mathsf{N}(\boldsymbol{0}, \sigma^2(\boldsymbol{I} - \boldsymbol{H}))$, sigue que

$$e_i \sim \mathsf{N}(0, \sigma^2(1 - h_{ii})), \qquad i = 1, \dots, n,$$

de ahí que $z_i = e_i/(\sigma\sqrt{1 - h_{ii}}) \sim \mathsf{N}(0, 1)$. Por tanto,

$$z_i^2 = \frac{e_i^2}{\sigma^2(1 - h_{ii})} \sim \chi^2(1).$$

3.c. Sabemos que

$$\frac{(n-p)s^2}{\sigma^2} = \frac{\boldsymbol{Y}^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}}{\sigma^2} \sim \chi^2(n-p).$$

Sea $\boldsymbol{d}_i = (0, \dots, 0, 1, 0, \dots, 0)^\top$ un vector con ceros salvo un 1 en la i-ésima posición. Esto nos permite escribir

$$e_i = \boldsymbol{d}_i^\top \boldsymbol{e} = \boldsymbol{d}_i^\top (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y},$$

y por la parte **3.b)** tenemos,

$$\frac{e_i^2}{\sigma^2(1 - h_{ii})} = \frac{\boldsymbol{Y}^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{d}_i\boldsymbol{d}_i^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}}{\sigma^2\boldsymbol{d}_i^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{d}_i} \sim \chi^2(1).$$

Lo anterior permite escribir

$$(n-p-1)s_{(i)}^2 = \boldsymbol{Y}^\top\Big(\boldsymbol{I} - \boldsymbol{H} - \frac{(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{d}_i\boldsymbol{d}_i^\top(\boldsymbol{I} - \boldsymbol{H})}{1 - h_{ii}}\Big)\boldsymbol{Y} = \boldsymbol{Y}^\top\boldsymbol{M}\boldsymbol{Y}.$$

Es fácil notar que $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{H} - (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{d}_i\boldsymbol{d}_i^\top(\boldsymbol{I} - \boldsymbol{H})/(1 - h_{ii})$ es matriz idempotente. Ahora,

$$\begin{aligned}
\operatorname{rg}(\boldsymbol{M}) &= \operatorname{tr}(\boldsymbol{I} - \boldsymbol{H}) - \frac{1}{1 - h_{ii}} \operatorname{tr}\big((\boldsymbol{I} - \boldsymbol{H})\boldsymbol{d}_i\boldsymbol{d}_i^\top(\boldsymbol{I} - \boldsymbol{H})\big) \\
&= n - p - \frac{1}{1 - h_{ii}} \operatorname{tr}(\boldsymbol{d}_i^\top(\boldsymbol{I} - \boldsymbol{H})^2\boldsymbol{d}_i) = n - p - 1.
\end{aligned}$$

Como

$$\frac{(n-p)s^2}{\sigma^2} = \frac{(n-p-1)s_{(i)}^2}{\sigma^2} + \frac{e_i^2}{\sigma^2(1 - h_{ii})},$$

y $\operatorname{rg}(\boldsymbol{I} - \boldsymbol{H}) = \operatorname{rg}(\boldsymbol{M}) + \operatorname{rg}(\boldsymbol{d}_i^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{d}_i)$ por el Teorema de Cochran, sigue la independencia entre $s_{(i)}^2$ y e_i^2.

4.a. El estadístico F para probar que $r - p$ parámetros del modelo son cero es dado por:

$$F = \frac{(\mathsf{RSS}_H - \mathsf{RSS})/(r - p)}{s^2},$$

donde RSS_H es la suma de cuadrados residual cuando existen p parámetros en el modelo, es decir, $\mathsf{RSS}_H = \mathsf{RSS}_p$. Luego,

$$F_p = \frac{(\mathsf{RSS}_p - \mathsf{RSS})/(r - p)}{s^2},$$

de ahí que

$$\begin{aligned}(r-p)F_p &= \frac{\mathsf{RSS}_p}{s^2} - \frac{(n-r)s^2}{s^2} = \frac{\mathsf{RSS}_p}{s^2} - n + r \\ &= C_p - 2p + r = (C_p - p) + (r - p),\end{aligned}$$

lo que lleva a

$$F_p = \frac{C_p - p}{r - p} + 1.$$

4.b. Sabemos que, el estadístico F para la hipótesis $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1}$ es dado por

$$F = \frac{(\mathsf{RSS}_H - \mathsf{RSS})/(p-1)}{\mathsf{RSS}/(n-p)}.$$

Bajo H_0 tenemos el modelo reducido $\boldsymbol{Y} = \beta_0\mathbf{1} + \boldsymbol{\epsilon}$, esto permite notar que

$$\mathsf{RSS}_H = \sum_{i=1}^n (Y_i - \overline{Y})^2 = SYY.$$

Por otro lado, para modelos con intercepto tenemos:

$$R^2 = 1 - \frac{\mathsf{RSS}}{SYY},$$

es decir, $\mathsf{RSS} = (1 - R^2)SYY$. Esto permite escribir

$$F = \Big(\frac{n-p}{p-1}\Big)\frac{SYY - (1-R^2)SYY}{(1-R^2)SYY} = \frac{R^2/(p-1)}{(1-R^2)/(n-p)} \sim \mathsf{F}(p-1, n-p).$$

Para notar la relación con la distribución Beta, considere

$$\frac{p-1}{n-p}F = \frac{R^2}{1-R^2}.$$

Sea $u = p - 1$ y $v = n - p$, luego

$$\frac{u}{v}F(1-R^2) = R^2 \qquad \Rightarrow \qquad \frac{u}{v}F = R^2\Big(1\frac{u}{v}F\Big).$$

Esto lleva a escribir

$$R^2 = \frac{uF/v}{1 + uF/v} \sim \mathsf{Beta}(p-1, n-p).$$

De donde es directo que

$$\mathsf{E}(R^2) = \frac{p-1}{p-1+n-p} = \frac{p-1}{n-1}.$$