Homework 5

Rafael Espinosa

Problem 1

1)

$$T = \underbrace{X}_{n \times p} \underbrace{Q'}_{p \times p} \quad , \quad \gamma = \underbrace{Q}_{p \times p} \underbrace{\beta}_{p \times 1}$$

$$T\gamma = \underbrace{X}_{n \times p} \underbrace{Q'}_{p \times p} \underbrace{Q}_{p \times p} \underbrace{\beta}_{p \times 1}$$

$$T\gamma = X\beta$$

where we used that $Q'Q = \mathbb{1}_{p \times p}$.

2) y regressing on T is given by

$$y = T\gamma + \sigma\varepsilon$$

Using that $T\gamma = X\beta$ (found in 1)

$$y = X\beta + \sigma\varepsilon$$

Using $z = Py$,

$$\begin{aligned} z &= PX\beta + P\sigma\varepsilon \\ &= \underbrace{PXQ'}_{D}\underbrace{Q\beta}_{\gamma} + P\sigma\varepsilon \\ &= D\gamma + P\sigma\varepsilon \end{aligned}$$

So, both regressions are equivalent, we can obtain one into the other by a linear transformation.

3) Using R and $\gamma_j = \frac{z_j}{x_j}$

$\tilde{\gamma} = (-1.0936012, -0.3767857, -0.6785582, 2.1572303, -1.2684318)$

Using R, using all components we obtain

$$\hat{\beta} = \begin{pmatrix} 0.6717704 \\ 1.3122409 \\ -0.2701956 \\ 1.9925564 \\ -1.3580854 \end{pmatrix}$$

4) With the highest 4 eigenvalues we have

$\tilde{\gamma} = (-1.0936012, -0.3767857, -0.6785582, 2.1572303)$

5) We find that for $m = 2$

$$\frac{\sum_{j=1}^{m} \lambda_j^2}{\sum_{j=1}^{p} \lambda_j^2} = 0.9721144 > 0.9$$

So, we use the 2 components with the 2 smallest eigenvalues

Using R, we obtain

$$\hat{\beta}_{pca} = \begin{pmatrix} 0.3812321 \\ 0.6976637 \\ 0.2992885 \\ 0.5230571 \\ 0.5854025 \end{pmatrix}$$

Problem 2

1) $$p(y_1,\dots,y_n \mid z=0) = \int \frac{\prod_{i=1}^{N} N(x_i\beta,\sigma^2)\,\pi(\beta\mid z=0)\,\cancel{p(z=0)}}{\cancel{p(z=0)}}\,d\beta$$

$$= \int \prod_{i=1}^{N} N(x_i\beta,\sigma^2)\,\pi(\beta\mid z=0)\,d\beta$$

where

$$\prod_{i=1}^{N} N(x_i\beta,\sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - x_i\beta)^2\right)$$

$$= \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i\beta)^2\right)$$

and $$\pi(\beta\mid z=0) = \delta_0(\beta)$$

So

$$p(y_1,\dots,y_n\mid z=0) = \int \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i\beta)^2\right)\delta_0(\beta)\,d\beta$$

The delta function sets the integral as 0 for any value $\beta \neq 0$. So, the integration makes to set $\underline{\beta = 0}$. Hence.

$$\boxed{p(y_1,\dots,y_n\mid z=0) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i^2\right)}$$

2)

$$p(y_1,\dots,y_n|z=1) = \int \frac{\prod_{i=1}^{N} N(x_i\beta,\sigma^2)\,\pi(\beta|z=1)\,\cancel{\pi(z=1)}}{\cancel{\pi(z=1)}}\,d\beta$$

$$= \prod_{i=1}^{N} N(x_i\beta,\sigma^2)\,\pi(\beta|z=1)$$

where $\prod_{i=1}^{N} N(x_i\beta,\sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - x_i\beta)^2\right)$

$$= \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i\beta)^2\right)$$

and $\pi(\beta|z=1) = \frac{1}{\sqrt{2\pi}\,\tau} \exp\left(-\frac{1}{2\tau^2}\beta^2\right)$

Hence,

$$p(y_1,\dots,y_n|z=1) = \int \frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i\beta)^2\right)\exp\left(-\frac{\beta^2}{2\tau^2}\right)}{(2\pi)^{n/2}\sigma^n\sqrt{2\pi}\,\tau}\,d\beta$$

$$= \int \frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i\beta)^2 - \beta^2/2\tau^2\right)}{(2\pi)^{n/2}\sigma^n\sqrt{2\pi}\,\tau}\,d\beta$$

Now, working the exponential terms

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i\beta)^2 - \frac{\beta^2}{2\tau^2} =$$

$$= -\frac{1}{2\sigma^2}\left[\sum_{i=1}^{N} y_i^2 - 2\beta x_i y_i + (x_i\beta)^2\right] - \frac{\beta^2}{2\tau^2}$$

$$= -\frac{1}{2\sigma^2}\left[\sum_{i=1}^{N} y_i^2 - 2\beta\sum_{i=1}^{N} x_i y_i + \beta^2\sum_{i=1}^{N} x_i^2\right] - \frac{\beta^2}{2\tau^2}$$

$$= -\frac{1}{2\sigma^2}\left[\sum_{i=1}^{N} y_i^2 - 2\beta\gamma + \beta^2 n\right] - \frac{\beta^2}{2\tau^2}$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i^2 + \frac{\beta}{\sigma^2}\gamma - \frac{\beta^2 n}{2\sigma^2} - \frac{\beta^2}{2\tau^2}$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i^2 + \frac{\beta}{\sigma^2}\gamma - \frac{\beta^2}{2}\left(\frac{n}{\sigma^2} - \frac{1}{\tau^2}\right)$$

For the moment call $c_1 = \frac{\gamma}{\sigma^2}$, $c_2 = \frac{n}{\sigma^2} - \frac{1}{\tau^2}$

So,

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i\beta)^2 - \frac{\beta^2}{2\tau^2} = -\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i^2 + c_1\beta - \frac{c_2}{2}\beta^2$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i^2 - \frac{c_2}{2}\left(\beta^2 - 2\frac{c_1}{c_2}\beta\right)$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i^2 - \frac{c_2}{2}\left(\beta^2 - 2\frac{c_1}{c_2}\beta + \left(\frac{c_1}{c_2}\right)^2 - \left(\frac{c_1}{c_2}\right)^2\right)$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i^2 - \frac{c_2}{2}\left(\left(\beta - \frac{c_1}{c_2}\right)^2 - \left(\frac{c_1}{c_2}\right)^2\right)$$

Therefore

$$p(y_1,\dots,y_n \mid z=1) = \int \frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i^2 - \frac{c_2}{2}\left(\left(\beta - \frac{c_1}{c_2}\right)^2 - \left(\frac{c_1}{c_2}\right)^2\right)\right)}{(2\pi)^{n/2}\sigma^n\sqrt{2\pi}\,\tau}\,d\beta$$

We integrate with respect β

$$\frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i^2 + \frac{c_1^2}{2c_2}\right)}{(2\pi)^{n/2}\sigma^n\sqrt{2\pi}\,\tau}\underbrace{\int_{-\infty}^{\infty}\exp\left(-\frac{c_2}{2}\left(\beta_1 - \frac{c_1}{c_2}\right)^2\right)d\beta}_{\frac{\sqrt{2\pi}}{\sqrt{c_2}}}$$

$$= \frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i + \frac{c_1^2}{2c_2}\right)}{(2\pi)^{n/2}\sigma^n\sqrt{c_2}\,\tau}$$

$$= \frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i + \frac{\gamma^2}{2n}\right)}{(2\pi)^{n/2}\sigma^n\sqrt{\frac{n}{\sigma^2}-\frac{1}{\tau^2}}\,\tau}$$

where we used that $c_2 = \frac{n}{\sigma^2} - \frac{1}{\tau^2}$

$c_1 = \frac{\gamma}{\sigma^2}$

4)

$$p(z=1 \mid y_1, y_2, \dots, y_n) = \frac{p(y_1, \dots, y_n \mid z=1)\, p(z=1)}{p(y_1, \dots, y_n \mid z=0)\, p(z=0) + p(y_1, \dots, y_n \mid z=1)\, p(z=1)}$$

$$= \frac{\dfrac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i + \frac{\gamma^2}{2n}\right)}{\cancel{(2\pi)^{n/2}\sigma^n}\sqrt{\frac{n}{\sigma^2}-\frac{1}{\tau^2}}\,\tau}\, q}{\cancel{\dfrac{1}{(2\pi)^{n/2}\sigma^n}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i^2\right)(1-q) + \dfrac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i + \frac{\gamma^2}{2n}\right)}{\cancel{(2\pi)^{n/2}\sigma^n}\sqrt{\frac{n}{\sigma^2}-\frac{1}{\tau^2}}\,\tau}\, q}$$

$$= \frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i + \frac{\gamma^2}{2n}\right) q}{\sqrt{\frac{n}{\sigma^2}-\frac{1}{\tau^2}}\,\tau \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i^2\right)(1-q) + \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i + \frac{\gamma^2}{2n}\right) q}$$

If $\gamma = 0$ and n is large

$$p(z=1|y_1,\dots,y_n) \approx \frac{\cancel{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i\right)}\, q}{\sqrt{\frac{n}{\sigma^2}}\,\tau\, \cancel{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N} y_i^2\right)}\,(1-q)}$$

where we considered that in the denominator the first term dominates when $n \to \infty$ and $\frac{n}{\sigma^2} - \frac{1}{\tau^2} \approx \frac{n}{\sigma^2}$.

$$\boxed{p(z=1|y_1,\dots,y_n) \approx \frac{\sigma}{\sqrt{n}}\,\frac{q}{1-q}}$$

5) From the assumption $\sum x_i = 0$, we have that $\bar{x} = 0$,

And the least square estimator is

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\gamma}{n}$$

So, setting $\gamma = 0$ is the same to have the estimator as $\hat{\beta} = 0$. So, in such case the dummy variable must be 0, not 1. Hence, we must expect that the probability

$$p(z=1 \mid y_1, \dots, y_n) \longrightarrow 0$$

as actually happens if $n \to \infty$ with what we obtained in 4)

$$p(z=1 \mid y_1, \dots, y_n) \simeq \lim_{n \to \infty} \frac{\tau}{\sqrt{n}} \frac{q}{1-q} = 0$$