Correção da P1

1. (a) $E[\varepsilon Z_1] = 0, E[XZ_1] \neq 0$ e Z_1 afeta y somente atravé de X.
   $E[\varepsilon Z_2] = 0, E[XZ_2] \neq 0$ e Z_2 afeta y somente atravé de X.

   (b) Contrapartidas Amostrais
   (i) $1/N \sum \hat{\varepsilon}_i = 0$
   $1/N \sum y_i - \hat{b}_0 - \hat{b}_1 x_i = 0$
   $\overline{y} - \hat{b_0} - \hat{b_1}\overline{x} = 0$
   $\hat{b_0} = \overline{y} - \hat{b_1}\overline{x}$

   (ii) $1/N \sum \hat{\varepsilon}_i Z_i = 0$
   $1/N \sum (y_i - \hat{b_0} - \hat{b_1} x_i) Z_i = 0$
   $1/N \sum (y_i - (\overline{y} - \hat{b_1}\overline{x}) - \hat{b_1} x_i) Z_i = 0$
   $1/N \sum (y_i - \overline{y}) Z_i - 1/N \sum \hat{b_1}(x_i - \overline{x}) Z_i = 0$
   $\hat{b_1} = \frac{\sum (y_i - \overline{y}) Z_i}{\sum (x_i - \overline{x}) Z_i}$
   $plim \quad \hat{b_1}^{IV} \longrightarrow \frac{c\hat{o}v(y,z)}{c\hat{o}v(x,z)}$

   (c) Para estima b_1 através do uso de uma variável instrumenal utilizando MQ2E, primeiramente, devemos regredir a variável endógena X em Z_1 e a partir dos coeficientes estimados calculamos o valor de $\hat{X}$ (o valor de X livre de endogeneidade). Posteriormente realizamos um MQO de y sobre $\hat{X}$.

   Passos:
   1º : MQO X vs Z (Primeiro Estágio)
   $\hat{\alpha_0} = \overline{x} - \hat{\alpha_1}\overline{z}$
   $\hat{\alpha_1} = \frac{c\hat{o}v(x,Zz)}{v\hat{a}r(Z)} \qquad (1)$
   Calculamos $\hat{x}$: $\hat{x} = \hat{\alpha_0} + \hat{\alpha_1} Z_i$

   2º : MQO y vs $\hat{x}$ (Segundo Estágio)
   $\hat{b}^{MQ2E} = \frac{c\hat{o}v(y,\hat{x})}{v\hat{a}r(\hat{x})}$

   *Calculando a $v\hat{a}r(\hat{x})$

$\hat{var}(\hat{x}) = \hat{var}(\hat{\alpha_0} + \hat{\alpha_1} Z) = \alpha_1^2 \hat{var}(Z)$

Substituindo α_1 por (1), temos:
$\hat{var} = \frac{\hat{cov}(x,Z)}{\hat{var}(Z)^2}^2 \hat{var}(Z)$

$\hat{var} = \frac{\hat{cov}(x,Z)^2}{\hat{var}(Z)} \qquad (2)$

*Calculando a $\hat{Cov}(\hat{x}, y)$

$\hat{cov}(\hat{x}, y) = \hat{Cov}(\alpha_0 + \hat{\alpha_1} Z_1, y)$
$\hat{cov}(\hat{x}, y) = \hat{\alpha_1} cov(Z, y)$

Substituindo α_1 por (1), temos:

$\hat{cov}(\hat{x}, y) = \frac{\hat{cov}(x,Z)}{\hat{var}(Z)} cov(Z, y)$

*Sustituindo (2) no b^{MQ2E}, temos:

$\hat{b_1}^{MQ2E} = \frac{\frac{\hat{cov}(x,Z)}{\hat{var}(Z)} cov(Z,y)}{\hat{var}(\hat{x})}$
$\hat{b_1}^{MQ2E} = \frac{\hat{cov}(x,Z)\hat{cov}(Z,y)}{\hat{var}(\hat{x})\hat{var}(Z)} \qquad (3)$

*Sustituindo (2) em (3), temos:

$\hat{b_1}^{MQ2E} = \frac{\hat{cov}(x,Z)\hat{cov}(Z,y)}{\hat{var}(Z)} \frac{\hat{var}(Z)}{\hat{cov}(x,Z)^2}$
$\hat{b_1}^{MQ2E} = \frac{\hat{cov}(Z,y)}{\hat{cov}(x,Z)} = \hat{b_1}^{IV}$

(d) Seria preferível a VI que obtivesse a menor soma dos quadrados dos resíduos, pois, desta forma, teríamos um intrumento mais correlacionado com a variável endógena e a variância do estimador seria menor. Observamos isso na istribuição do estimador abaixo :

$$\hat{b}^{VI} \sim N(b, \frac{\sigma^2}{SQT_x \rho_{xz}^2}), onde :$$

$$\rho^2_{xz} = 1 - \frac{SQE_{xz}}{SQT_x}$$

Portanto, preferiríamos o estimador que usasse o instrumento Z_1.

2. Supondo $\alpha = 5\%$

   (a) $H_0 : b_1 = -3$

   $\frac{\hat{b}-b}{\sigma} = \frac{4+3}{\sqrt{16}} = 1,8 < 1,96$, portanto, não rejeita.

   (b) $H_0 : b_1 = b_0$

   $\frac{\hat{b}-b}{\sigma} = \frac{5}{\sqrt{0,5+16-4}} = 1,44 < 1,96$, portanto, não rejeita.

   (c) $H_0 : \beta = 0$

$$\begin{bmatrix} -1 & 4 & 9 \end{bmatrix} \begin{bmatrix} 4 & -0.5 & 0 \\ -0.5 & 1/8 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} -1 & 4 & 9 \end{bmatrix} \begin{bmatrix} -6 \\ 1 \\ 9 \end{bmatrix} = 6+4+81 = 91$$

   Como a distribuição segue uma $\chi^2_{(3)}$, rejeitamos a hipótese nula.

   (d) $H_0 = \begin{bmatrix} b_1 = 4 \\ b_2 = 3 \end{bmatrix}$

$$\begin{bmatrix} 4 & 3 \end{bmatrix} \begin{bmatrix} 1/16 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = 1 + 9 = 10$$

   Como a distribuição segue uma $\chi^2_{(2)}$, não- rejeitamos a hipótese nula.

3. (a) Formato da matriz de covariância de ε - Matriz bloco diagonal

$$\sum = \begin{pmatrix} \sum_1 & \cdots & 0 \\ 0 & \ddots & \vdots \\ 0 & \cdots & \sum_n \end{pmatrix} \tag{1}$$

onde cada somatório na diagonal da matrix tem a seguinte estrutura:

$$\sum_i = \begin{pmatrix} \sigma_v^2 + \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_u^2 & \cdots & \vdots \\ \vdots & & \ddots & \sigma_u^2 \\ \sigma_c^2 & & & \sigma_c^2 + \sigma_u^2 \end{pmatrix} \tag{2}$$

(b) Neste caso teriamos que testar :

$H_0 : b_2 > 1$

Considerando o teste para o caso assintótico, temos

$$\frac{\hat{b}_2 - 1}{[X'(\sum)^{-1}X]^{-1}_{3X3}} \sim Z$$

4. Modelo:

$$salrio_{it} = \alpha_0 + \alpha_1 calorias_{it} + u_{it}$$

enquanto, por hipótese, leva tempo para que a maior renda tenha impacto sobre o consumo de calorias do indivíduo, ou seja:

$$calorias_{it} = \beta_0 + \beta_1 salrio_{it-1} + \beta_2 z_{it} + \beta_3 z_{2it} + \eta_{it}$$

Suponha ainda que:

- $cov(calorias_{is}, u_{it}) = 0 \qquad \forall s < t;$
- $cov(z_{it}, u_{it}) = 0;$
- $cov(z_{2it}, u_{it}) = 0;$
- $cov(z_{it}, \eta_{it}) = 0;$

- $cov(z_{2it}, \eta_{it}) = 0$;
- $cov(\eta_{it}, u_{it}) = 0$

(a) $plim \quad \hat{\alpha_1} = \frac{cov(calorias_{it}, salario_{it})}{var(calorias_{it})} = \frac{cov(calorias_{it}, \alpha_0 + \alpha_1 calorias_{it} + u_{it})}{var(calorias_{it})}$

$= \alpha_1 + \frac{cov(calorias_{it}, u_{it})}{var(calorias_{it})}$

mas,

$cov(calorias_{it}, u_{it}) = cov(\beta_0 + \beta_1 salrio_{it-1} + \beta_2 z_{it} + \beta_3 z_{2it} + \eta_{it}, u_{it})$

$= \beta_1 cov(salrio_{it-1}, u_{it}) = \beta_1 cov(\alpha_0 + \alpha_1 calorias_{it-1} + u_{it-1}, u_{it})$

$= \beta_1 cov(u_{it-1}, u_{it}) \neq 0$

Logo, $plim \quad \hat{\alpha_1} = \alpha_1 + \frac{cov(calorias_{it}, u_{it})}{var(calorias_{it})} = \alpha_1 + \frac{\beta_1 cov(u_{it}, u_{it-1})}{var(calorias_{it})} \neq \alpha_1$,

e o estimador MQO de α_1 é inconsistente.

Podemos ter $cov(u_{it-1}, u_{it}) \neq 0$ se o salário em t for correlacionado com o salario em $t-1$, o que seria devido à, por exemplo, fatores intrínsecos do indivíduo (habilidade, inteligência, etc).

(b) Com $\beta_2 = 0$ e $\beta_3 = 0$, temos $calorias_{it} = \beta_0 + \beta_1 salrio_{it-1} + \eta_{it}$.

Deste modo, teríamos o problema clássico de equações simultâneas. VI: $salario_{it-1}$

$plim \ \hat{\alpha_1}^{VI} = \frac{cov(salario_{it}, salrio_{it-1})}{cov(salario_{it-1}, calorias_{it})}$

$plim \ \hat{\alpha_1}^{VI} = \frac{cov(\alpha_0 + \alpha_1 calorias_{it} + u_{it}, \alpha_0 + \alpha_1 calorias_{it-1} + u_{it-1})}{cov(\alpha_0 + \alpha_1 calorias_{it-1} + u_{it-1}, calorias_{it})}$

$plim \ \hat{\alpha_1}^{VI} = \frac{\alpha_1^2 cov(calorias_{it}, calorias_{it-1}) + \alpha_1 cov(calorias_{it}, u_{it-1}) + cov(u_{it}, u_{it-1})}{\alpha_1 cov(calorias_{it-1}, calorias_{it}) + cov(u_{it-1}, calorias_{it})} \neq \alpha_1$

Com $\beta_2 \neq 0$ e $\beta_3 = 0$, temos $calorias_{it} = \beta_0 + \beta_1 salrio_{it-1} + \beta_2 z_{it} + \eta_{it}$.

Assim, z_{it} é instrumento para calorias e $\hat{\alpha} = \frac{Cov(z_{it}, salrio)}{Cov(z_{it}, cal)}$
Desta forma teriamos uma variável que afeta a variável endógena, é não correlacionada com o erro u_{it}, e não afeta salário diretamente, assim, poderíamos estimar consistentemente o coeficiente α_1.

(c) Com $\beta_2 \neq 0$ e $\beta_3 \neq 0$
$calorias_{it} = \pi_0 + \pi_1 z_{it} + \pi_2 z_{jt} + v_{2t}$, estimamos por MQO e encontramos $\hat{v_{2t}}$ (z_{jt} é qualquer outra variável exógena)

$salario_{it} = \alpha_0 + \alpha_1 calorias_{it} + \alpha_2 \hat{v_{2t}} + w_{it}$

Realizamos um teste de hipótese, onde :
$H_0 : \alpha_2 = 0 \rightarrow calorias_{it}$ é exógena ($cov(v_{2t}, u_{it}) = 0$)
$H_0 : \alpha_2 \neq 0 \rightarrow calorias_{it}$ é endógena ($cov(v_{2t}, u_{it}) \neq 0$)

5. (a) Podemos observar pela tabela 1 que independente do tipo de estimação realizada o coeficiente da produção de petróleo é estatisticamente significante (vemos isso fazendo a estatística do teste), ou seja, há covariância entre a variável endógena e o instrumento proposto. Além disso vemos que a magnitude do coeficiente não é desprezível e podemos observar que a estatística F e o R^2 apresentam um valores altos, o que comprova a relevância conjunta dos coeficientes. O fato de ser considerado um "produtor de petróleo" se dar de forma aleatória nos permite considerar que o instrumento não tenha correlação com o erro na regressão.

(b) Observando a Tabela 6 e assumindo que o instrumento seja válido, podemos suspeitar de endógeneidade na variável, pois há diferença significativa nas estimativas do coeficiente do royalties do petróleo entre a estimativa por MQO e por variável instrumental.

(c) Uma das hipóteses para que uma variável possa instrumentalizar uma variável endógena, é que ela afete a variável dependente apenas através da variável endógena. Pelas hipóteses apresentadas fica evidente que essa condição não é respeitada e portanto não seria possível usar essas variáveis como instrumento.