Ec122 - HW9

Ben Juarez

11/28/2021

(2)

In the future, if I were to win a Nobel Prize for something in economics, it would perhaps be related to cryptocurrency. The future implications of cryptocurrency are seemingly endless, and there would many different angles to research and study (which might result in a Nobel Prize!). At this point, it does not seem like cryptocurrency is something that is going to go away, so in the near future, we are likely to see an even fuller scale of adoption of cryptocurrency methods. In particular, once major world governments begin to adopt certain cryptocurrency standards, then there will be an even greater demand for understanding its impact on our economy. Even though cryptocurrency is already quite prominent, there is still much unknown about its future implications, which leaves open potential for major discoveries and research!

(3)

3.1

```r
library("MASS")
mu = c(0,0)
Sigma_e = matrix(c(1,-.9,-.9,1),2,2)
Sigma_x = matrix(c(1,.9,.9,1),2,2)
epsilon = mvrnorm(4000,mu,Sigma_e)
x = mvrnorm(4000,mu,Sigma_x)

x1 = x[,1]
x2 = x[,2]
e1 = epsilon[,1]
e2 = epsilon[,2]

y1 = 2 + 2 * x1 + e1
y2 = 4 * x2 + e2
z = ifelse(y2>0,1,0)
```

3.2

```r
y1z = y1[z > 0]
x1z = x1[z > 0]
length(y1z)
```

```
## [1] 1937
```

```r
out2 = lm(y1z ~ x1z)
summary(out2)
```

```
##
## Call:
## lm(formula = y1z ~ x1z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9622 -0.6460  0.0004  0.6417  3.2455
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.65858    0.03095   53.59   <2e-16 ***
## x1z          2.23273    0.03086   72.36   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9714 on 1935 degrees of freedom
## Multiple R-squared:  0.7302, Adjusted R-squared:   0.73
## F-statistic:  5236 on 1 and 1935 DF,  p-value: < 2.2e-16
```

We can see that we have close to 2000 pairs of (y_1, x_1) such that $z = 1$. Since we know $z = 1$ only when $y_2 > 0$, and since $x_2 > 0$ is true for about half of its values due to the $N(0, 1)$ distribution, we can see that if $x_2 > 0$ then $y_2 > 0$. Thus, around half of y_2 values will also be positive.

Considering out estimates for α_1 and β_1, we can see that they are both close 2 and that they are significant at the 5% level following our summary output. However, we still are ignoring the correlation between x_1 and x_2.

3.3

```r
library("stats")
out3 = glm(z ~ x2, family = binomial(link = probit))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
summary(out3)
```

```
##
## Call:
## glm(formula = z ~ x2, family = binomial(link = probit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7787  -0.0890   0.0000   0.0679   3.6694
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.05502    0.03881  -1.418    0.156
## x2           4.11058    0.14560  28.232   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5541.2  on 3999  degrees of freedom
## Residual deviance: 1334.1  on 3998  degrees of freedom
## AIC: 1338.1
##
```

```
## Number of Fisher Scoring iterations: 8
```

```
B2 = out3$coefficients[2]
B2
```

```
##       x2
## 4.110578
```

We can see that $\hat{\beta}_2$ (displayed above) is close to 4.

3.4

```
x2z = x2[z>0]
lambda = dnorm(-B2*x2z) / (1-pnorm(-B2*x2z))
out4 = lm(y1z ~ x1z + lambda)
summary(out4)
```

```
##
## Call:
## lm(formula = y1z ~ x1z + lambda)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0506 -0.5907 -0.0013  0.6090  3.1937
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.98190    0.03815   51.95   <2e-16 ***
## x1z          2.00508    0.03403   58.91   <2e-16 ***
## lambda      -0.89805    0.06684  -13.44   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9292 on 1934 degrees of freedom
## Multiple R-squared:  0.7532, Adjusted R-squared:  0.7529
## F-statistic:  2951 on 2 and 1934 DF,  p-value: < 2.2e-16
```

It appears that the estimates of α_1 and β_1 in this case are much closer to 2 (compared to the previous part). The estimates also have similar standard errors and significance (5% level). Overall, it looks like a more accurate estimation.

3.5

Let us repeat parts 1-4 for $\rho_{\epsilon_1,\epsilon_2} = -.5, 0., .5, .9$:

3.5.i

For $\rho_{\epsilon_1,\epsilon_2} = -.5$:

```
mu = c(0,0)
Sigma_e = matrix(c(1,-.5,-.5,1),2,2)
Sigma_x = matrix(c(1,.9,.9,1),2,2)
epsilon = mvrnorm(4000,mu,Sigma_e)
x = mvrnorm(4000,mu,Sigma_x)

x1 = x[,1]
x2 = x[,2]
```

```r
e1 = epsilon[,1]
e2 = epsilon[,2]

y1 = 2 + 2 * x1 + e1
y2 = 4 * x2 + e2
z = ifelse(y2>0,1,0)

y1z = y1[z > 0]
x1z = x1[z > 0]
length(y1z)
```

```
## [1] 2006
```

```r
out5ia = lm(y1z ~ x1z)
summary(out5ia)
```

```
##
## Call:
## lm(formula = y1z ~ x1z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0027 -0.7005 -0.0298  0.6873  3.2840
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.82137    0.03140   58.01   <2e-16 ***
## x1z          2.15062    0.03006   71.54   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.993 on 2004 degrees of freedom
## Multiple R-squared:  0.7186, Adjusted R-squared:  0.7185
## F-statistic:  5118 on 1 and 2004 DF,  p-value: < 2.2e-16
```

```r
out5ib = glm(z ~ x2, family = binomial(link = probit))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
summary(out5ib)
```

```
##
## Call:
## glm(formula = z ~ x2, family = binomial(link = probit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3243  -0.0621   0.0000   0.0761   2.8036
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.01408    0.03883   0.363    0.717
## x2           4.00220    0.14400  27.794   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5545.1  on 3999  degrees of freedom
## Residual deviance: 1329.6  on 3998  degrees of freedom
## AIC: 1333.6
##
## Number of Fisher Scoring iterations: 8
```

```
B2 = out5ib$coefficients[2]
B2
```

```
##       x2
## 4.002202
```

```
x2z = x2[z>0]
lambda = dnorm(-B2*x2z) / (1-pnorm(-B2*x2z))
out5ic = lm(y1z ~ x1z + lambda)
summary(out5ic)
```

```
##
## Call:
## lm(formula = y1z ~ x1z + lambda)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0516 -0.6742 -0.0075  0.6703  3.0461
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.04277    0.04039  50.575   <2e-16 ***
## x1z          2.00139    0.03437  58.227   <2e-16 ***
## lambda      -0.59698    0.07029  -8.493   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9759 on 2003 degrees of freedom
## Multiple R-squared:  0.7284, Adjusted R-squared:  0.7281
## F-statistic:  2686 on 2 and 2003 DF,  p-value: < 2.2e-16
```

We can see that the first OLS estimates are more accurate relative to part (2). This is likely due to the lower correlation. Including $\lambda(-\hat{\beta}_2 x_2)$ similarly increases the accuracy of the estimates closer to 2.

3.5.ii

For $\rho_{\epsilon_1,\epsilon_2} = 0$:

```
mu = c(0,0)
Sigma_e = matrix(c(1,0,0,1),2,2)
Sigma_x = matrix(c(1,.9,.9,1),2,2)
epsilon = mvrnorm(4000,mu,Sigma_e)
x = mvrnorm(4000,mu,Sigma_x)

x1 = x[,1]
x2 = x[,2]
e1 = epsilon[,1]
e2 = epsilon[,2]
```

```r
y1 = 2 + 2 * x1 + e1
y2 = 4 * x2 + e2
z = ifelse(y2>0,1,0)

y1z = y1[z > 0]
x1z = x1[z > 0]
length(y1z)
```

```
## [1] 2012
```

```r
out5iia = lm(y1z ~ x1z)
summary(out5iia)
```

```
##
## Call:
## lm(formula = y1z ~ x1z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6711 -0.6745  0.0123  0.6987  3.4024
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.95857    0.03134   62.50   <2e-16 ***
## x1z          2.03470    0.03149   64.62   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.018 on 2010 degrees of freedom
## Multiple R-squared:  0.6751, Adjusted R-squared:  0.6749
## F-statistic:  4176 on 1 and 2010 DF,  p-value: < 2.2e-16
```

```r
out5iib = glm(z ~ x2, family = binomial(link = probit))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
summary(out5iib)
```

```
##
## Call:
## glm(formula = z ~ x2, family = binomial(link = probit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.0212  -0.0872   0.0000   0.0896   3.4893
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.04216    0.03739   1.127     0.26
## x2           3.92295    0.14059  27.903   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5545.0  on 3999  degrees of freedom
```

```
## Residual deviance: 1446.4  on 3998  degrees of freedom
## AIC: 1450.4
##
## Number of Fisher Scoring iterations: 8
```

```
B2 = out5iib$coefficients[2]
B2
```

```
##      x2
## 3.92295
```

```
x2z = x2[z>0]
lambda = dnorm(-B2*x2z) / (1-pnorm(-B2*x2z))
out5iic = lm(y1z ~ x1z + lambda)
summary(out5iic)
```

```
##
## Call:
## lm(formula = y1z ~ x1z + lambda)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6486 -0.6783  0.0153  0.6996  3.4018
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.96988    0.04130  47.699   <2e-16 ***
## x1z          2.02692    0.03652  55.505   <2e-16 ***
## lambda      -0.02880    0.06849  -0.421    0.674
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.019 on 2009 degrees of freedom
## Multiple R-squared:  0.6751, Adjusted R-squared:  0.6748
## F-statistic:  2087 on 2 and 2009 DF,  p-value: < 2.2e-16
```

Since the errors are now uncorrelated, the first OLS estimates are even more accurate (closer to 2). With this, including $\lambda(-\hat{\beta}_2 x_2)$ does not improve the estimates in the same way as before, and the lambda coefficient is not as significant.

3.5.iii

For $\rho_{\epsilon_1,\epsilon_2} = .5$:

```
mu = c(0,0)
Sigma_e = matrix(c(1,.5,.5,1),2,2)
Sigma_x = matrix(c(1,.9,.9,1),2,2)
epsilon = mvrnorm(4000,mu,Sigma_e)
x = mvrnorm(4000,mu,Sigma_x)

x1 = x[,1]
x2 = x[,2]
e1 = epsilon[,1]
e2 = epsilon[,2]

y1 = 2 + 2 * x1 + e1
y2 = 4 * x2 + e2
```

```r
z = ifelse(y2>0,1,0)

y1z = y1[z > 0]
x1z = x1[z > 0]
length(y1z)
```

```
## [1] 2006
```

```r
out5iiia = lm(y1z ~ x1z)
summary(out5iiia)
```

```
##
## Call:
## lm(formula = y1z ~ x1z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3568 -0.6510 -0.0085  0.6798  3.0030
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.18932    0.02998   73.02   <2e-16 ***
## x1z          1.84035    0.03034   60.67   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9861 on 2004 degrees of freedom
## Multiple R-squared:  0.6475, Adjusted R-squared:  0.6473
## F-statistic:  3680 on 1 and 2004 DF,  p-value: < 2.2e-16
```

```r
out5iiib = glm(z ~ x2, family = binomial(link = probit))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
summary(out5iiib)
```

```
##
## Call:
## glm(formula = z ~ x2, family = binomial(link = probit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.0108  -0.0904   0.0000   0.0905   3.1702
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.05040    0.03741   1.347    0.178
## x2           3.92804    0.13956  28.146   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5545.1  on 3999  degrees of freedom
## Residual deviance: 1439.2  on 3998  degrees of freedom
## AIC: 1443.2
```

```
##
## Number of Fisher Scoring iterations: 8
```

```
B2 = out5iiib$coefficients[2]
B2
```

```
##      x2
## 3.92804
```

```
x2z = x2[z>0]
lambda = dnorm(-B2*x2z) / (1-pnorm(-B2*x2z))
out5iiic = lm(y1z ~ x1z + lambda)
summary(out5iiic)
```

```
##
## Call:
## lm(formula = y1z ~ x1z + lambda)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3918 -0.6622 -0.0117  0.6672  2.9551
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.99804    0.03905  51.165  < 2e-16 ***
## x1z          1.97272    0.03474  56.786  < 2e-16 ***
## lambda       0.49178    0.06555   7.502 9.37e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9727 on 2003 degrees of freedom
## Multiple R-squared:  0.6571, Adjusted R-squared:  0.6567
## F-statistic:  1919 on 2 and 2003 DF,  p-value: < 2.2e-16
```

Similar to part (5.i), the first OLS estimates are closer to 2 relative to part (2) due to lower correlation. Again, including $\lambda(-\hat{\beta}_2 x_2)$ slightly improves the estimates.

3.5.iv

For $\rho_{\epsilon_1,\epsilon_2} = .9$:

```
mu = c(0,0)
Sigma_e = matrix(c(1,.9,.9,1),2,2)
Sigma_x = matrix(c(1,.9,.9,1),2,2)
epsilon = mvrnorm(4000,mu,Sigma_e)
x = mvrnorm(4000,mu,Sigma_x)

x1 = x[,1]
x2 = x[,2]
e1 = epsilon[,1]
e2 = epsilon[,2]

y1 = 2 + 2 * x1 + e1
y2 = 4 * x2 + e2
z = ifelse(y2>0,1,0)

y1z = y1[z > 0]
```

```r
x1z = x1[z > 0]
length(y1z)
```

```
## [1] 2015
```

```r
out5iva = lm(y1z ~ x1z)
summary(out5iva)
```

```
##
## Call:
## lm(formula = y1z ~ x1z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9146 -0.6286  0.0054  0.6398  4.5371
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.34743    0.02978   78.83   <2e-16 ***
## x1z          1.75072    0.02971   58.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9632 on 2013 degrees of freedom
## Multiple R-squared:  0.633,  Adjusted R-squared:  0.6328
## F-statistic:  3472 on 1 and 2013 DF,  p-value: < 2.2e-16
```

```r
out5ivb = glm(z ~ x2, family = binomial(link = probit))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
summary(out5ivb)
```

```
##
## Call:
## glm(formula = z ~ x2, family = binomial(link = probit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.4225  -0.0918   0.0000   0.0816   2.9701
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.07176    0.03797    1.89   0.0588 .
## x2           3.93809    0.13931   28.27   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5545.0  on 3999  degrees of freedom
## Residual deviance: 1401.9  on 3998  degrees of freedom
## AIC: 1405.9
##
## Number of Fisher Scoring iterations: 8
```

```r
B2 = out5ivb$coefficients[2]
B2
```

```
##       x2
## 3.938094
```

```r
x2z = x2[z>0]
lambda = dnorm(-B2*x2z) / (1-pnorm(-B2*x2z))
out5ivc = lm(y1z ~ x1z + lambda)
summary(out5ivc)
```

```
##
## Call:
## lm(formula = y1z ~ x1z + lambda)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5874 -0.6087 -0.0306  0.6027  4.1198
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.96378    0.03775   52.03   <2e-16 ***
## x1z          2.01883    0.03315   60.90   <2e-16 ***
## lambda       0.95912    0.06274   15.29   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9119 on 2012 degrees of freedom
## Multiple R-squared:  0.6712, Adjusted R-squared:  0.6709
## F-statistic:  2054 on 2 and 2012 DF,  p-value: < 2.2e-16
```

The effects are similar to parts (2) and (4) since the correlation is of the same magnitude.

3.6

Let us repeat parts 1-4 for $\rho_{\epsilon_1,\epsilon_2} = -.9 - .5, 0., .5, .9$ with $\rho_{x_1,x_2} = 0$:

3.6.i

For $\rho_{\epsilon_1,\epsilon_2} = -.9$:

```r
mu = c(0,0)
Sigma_e = matrix(c(1,-.9,-.9,1),2,2)
Sigma_x = matrix(c(1,0,0,1),2,2)
epsilon = mvrnorm(4000,mu,Sigma_e)
x = mvrnorm(4000,mu,Sigma_x)

x1 = x[,1]
x2 = x[,2]
e1 = epsilon[,1]
e2 = epsilon[,2]

y1 = 2 + 2 * x1 + e1
y2 = 4 * x2 + e2
z = ifelse(y2>0,1,0)
```

```r
y1z = y1[z > 0]
x1z = x1[z > 0]
length(y1z)
```

```
## [1] 1938
```

```r
out6ia = lm(y1z ~ x1z)
summary(out6ia)
```

```
##
## Call:
## lm(formula = y1z ~ x1z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5894 -0.6459  0.0296  0.6827  3.2086
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.80497    0.02287   78.92   <2e-16 ***
## x1z          1.97810    0.02256   87.69   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.007 on 1936 degrees of freedom
## Multiple R-squared:  0.7989, Adjusted R-squared:  0.7988
## F-statistic:  7689 on 1 and 1936 DF,  p-value: < 2.2e-16
```

```r
out6ib = glm(z ~ x2, family = binomial(link = probit))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
summary(out6ib)
```

```
##
## Call:
## glm(formula = z ~ x2, family = binomial(link = probit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9973  -0.1036   0.0000   0.0677   3.8923
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.02046    0.03801  -0.538     0.59
## x2           3.87465    0.13666  28.352   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5541.3  on 3999  degrees of freedom
## Residual deviance: 1394.2  on 3998  degrees of freedom
## AIC: 1398.2
##
## Number of Fisher Scoring iterations: 8
```

```r
B2 = out6ib$coefficients[2]
B2
```

```
##       x2
## 3.874654
```

```r
x2z = x2[z>0]
lambda = dnorm(-B2*x2z) / (1-pnorm(-B2*x2z))
out6ic = lm(y1z ~ x1z + lambda)
summary(out6ic)
```

```
##
## Call:
## lm(formula = y1z ~ x1z + lambda)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2128 -0.5843  0.0304  0.6433  3.0434
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.96739    0.02444   80.48   <2e-16 ***
## x1z          1.98119    0.02143   92.44   <2e-16 ***
## lambda      -0.82859    0.05714  -14.50   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9565 on 1935 degrees of freedom
## Multiple R-squared:  0.8186, Adjusted R-squared:  0.8184
## F-statistic:  4365 on 2 and 1935 DF,  p-value: < 2.2e-16
```

We see similar effects as in parts (2) and (4) with the same error covariance. However, we see that there was not much change to β_1 estimate while we see the familiar improvement for α_1.

3.6.ii

For $\rho_{\epsilon_1,\epsilon_2} = -.5$:

```r
mu = c(0,0)
Sigma_e = matrix(c(1,-.5,-.5,1),2,2)
Sigma_x = matrix(c(1,0,0,1),2,2)
epsilon = mvrnorm(4000,mu,Sigma_e)
x = mvrnorm(4000,mu,Sigma_x)

x1 = x[,1]
x2 = x[,2]
e1 = epsilon[,1]
e2 = epsilon[,2]

y1 = 2 + 2 * x1 + e1
y2 = 4 * x2 + e2
z = ifelse(y2>0,1,0)

y1z = y1[z > 0]
x1z = x1[z > 0]
length(y1z)
```

```
## [1] 1965
```

```r
out6iia = lm(y1z ~ x1z)
summary(out6iia)
```

```
##
## Call:
## lm(formula = y1z ~ x1z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0810 -0.6856 -0.0009  0.6955  3.4627
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.87956    0.02270   82.81   <2e-16 ***
## x1z          2.01352    0.02302   87.45   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.006 on 1963 degrees of freedom
## Multiple R-squared:  0.7958, Adjusted R-squared:  0.7957
## F-statistic:  7648 on 1 and 1963 DF,  p-value: < 2.2e-16
```

```r
out6iib = glm(z ~ x2, family = binomial(link = probit))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
summary(out6iib)
```

```
##
## Call:
## glm(formula = z ~ x2, family = binomial(link = probit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7457  -0.0929   0.0000   0.0785   3.2706
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.005644   0.037677   -0.15    0.881
## x2           3.946528   0.139331   28.32   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5544.0  on 3999  degrees of freedom
## Residual deviance: 1424.4  on 3998  degrees of freedom
## AIC: 1428.4
##
## Number of Fisher Scoring iterations: 8
```

```r
B2 = out6iib$coefficients[2]
B2
```

```
##       x2
```

```
## 3.946528
```

```
x2z = x2[z>0]
lambda = dnorm(-B2*x2z) / (1-pnorm(-B2*x2z))
out6iic = lm(y1z ~ x1z + lambda)
summary(out6iic)
```

```
##
## Call:
## lm(formula = y1z ~ x1z + lambda)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0961 -0.6660 -0.0024  0.7044  3.3568
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.98753    0.02546  78.069   <2e-16 ***
## x1z          2.01763    0.02260  89.285   <2e-16 ***
## lambda      -0.53868    0.06154  -8.754   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9872 on 1962 degrees of freedom
## Multiple R-squared:  0.8034, Adjusted R-squared:  0.8032
## F-statistic:  4010 on 2 and 1962 DF,  p-value: < 2.2e-16
```

The first OLS estimates are slightly better than the previous part, and including $\lambda(-\hat{\beta}_2 x_2)$ similarly helps increase the accuracy. However, this improvement is mainly in α_1.

3.6.iii

For $\rho_{\epsilon_1,\epsilon_2} = 0$:

```
mu = c(0,0)
Sigma_e = matrix(c(1,0,0,1),2,2)
Sigma_x = matrix(c(1,0,0,1),2,2)
epsilon = mvrnorm(4000,mu,Sigma_e)
x = mvrnorm(4000,mu,Sigma_x)

x1 = x[,1]
x2 = x[,2]
e1 = epsilon[,1]
e2 = epsilon[,2]

y1 = 2 + 2 * x1 + e1
y2 = 4 * x2 + e2
z = ifelse(y2>0,1,0)

y1z = y1[z > 0]
x1z = x1[z > 0]
length(y1z)
```

```
## [1] 2011
```

```
out6iiia = lm(y1z ~ x1z)
summary(out6iiia)
```

```
##
## Call:
## lm(formula = y1z ~ x1z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7094 -0.6583 -0.0054  0.6553  3.4044
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.00506    0.02164   92.66   <2e-16 ***
## x1z          1.96825    0.02139   92.00   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9704 on 2009 degrees of freedom
## Multiple R-squared:  0.8082, Adjusted R-squared:  0.8081
## F-statistic:  8464 on 1 and 2009 DF,  p-value: < 2.2e-16
```

```
out6iiib = glm(z ~ x2, family = binomial(link = probit))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(out6iiib)
```

```
##
## Call:
## glm(formula = z ~ x2, family = binomial(link = probit))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -3.15466  -0.08573   0.00000   0.10364   3.00829
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.006853   0.037416  -0.183    0.855
## x2           3.949217   0.138025  28.612   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5545.1  on 3999  degrees of freedom
## Residual deviance: 1437.6  on 3998  degrees of freedom
## AIC: 1441.6
##
## Number of Fisher Scoring iterations: 8
```

```
B2 = out6iiib$coefficients[2]
B2
```

```
##       x2
## 3.949217
```

```
x2z = x2[z>0]
lambda = dnorm(-B2*x2z) / (1-pnorm(-B2*x2z))
out6iiic = lm(y1z ~ x1z + lambda)
summary(out6iiic)
```

```
##
## Call:
## lm(formula = y1z ~ x1z + lambda)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7135 -0.6589 -0.0055  0.6515  3.4002
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.00921    0.02462  81.619   <2e-16 ***
## x1z          1.96823    0.02140  91.978   <2e-16 ***
## lambda      -0.02109    0.05950  -0.354    0.723
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9706 on 2008 degrees of freedom
## Multiple R-squared:  0.8082, Adjusted R-squared:  0.808
## F-statistic:  4230 on 2 and 2008 DF,  p-value: < 2.2e-16
```

We see that the first OLS estimates are much more improved in this case while including $\lambda(-\hat{\beta}_2 x_2)$ does not help in the same way. As before, the lambda coefficient is not as significant.

3.6.iv

For $\rho_{\epsilon_1,\epsilon_2} = .5$:

```
mu = c(0,0)
Sigma_e = matrix(c(1,.5,.5,1),2,2)
Sigma_x = matrix(c(1,0,0,1),2,2)
epsilon = mvrnorm(4000,mu,Sigma_e)
x = mvrnorm(4000,mu,Sigma_x)

x1 = x[,1]
x2 = x[,2]
e1 = epsilon[,1]
e2 = epsilon[,2]

y1 = 2 + 2 * x1 + e1
y2 = 4 * x2 + e2
z = ifelse(y2>0,1,0)

y1z = y1[z > 0]
x1z = x1[z > 0]
length(y1z)
```

```
## [1] 1995
```

```
out6iva = lm(y1z ~ x1z)
summary(out6iva)
```

```
##
## Call:
## lm(formula = y1z ~ x1z)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4439 -0.6508  0.0205  0.6699  3.4828
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.12800    0.02233   95.30   <2e-16 ***
## x1z          1.96778    0.02258   87.15   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9973 on 1993 degrees of freedom
## Multiple R-squared:  0.7921, Adjusted R-squared:  0.792
## F-statistic:  7594 on 1 and 1993 DF,  p-value: < 2.2e-16
```

```
out6ivb = glm(z ~ x2, family = binomial(link = probit))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(out6ivb)
```

```
##
## Call:
## glm(formula = z ~ x2, family = binomial(link = probit))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.94798  -0.06265   0.00000   0.07172   2.99716
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.001548   0.038663    0.04    0.968
## x2          4.082497   0.147510   27.68   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5545.2  on 3999  degrees of freedom
## Residual deviance: 1342.4  on 3998  degrees of freedom
## AIC: 1346.4
##
## Number of Fisher Scoring iterations: 8
```

```
B2 = out6ivb$coefficients[2]
B2
```

```
##       x2
## 4.082497
```

```
x2z = x2[z>0]
lambda = dnorm(-B2*x2z) / (1-pnorm(-B2*x2z))
out6ivc = lm(y1z ~ x1z + lambda)
```

```r
summary(out6ivc)
```

```
##
## Call:
## lm(formula = y1z ~ x1z + lambda)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3859 -0.6483  0.0237  0.6545  3.5549
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.04910    0.02488  82.345  < 2e-16 ***
## x1z          1.96847    0.02232  88.180  < 2e-16 ***
## lambda       0.42238    0.06149   6.869 8.63e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9859 on 1992 degrees of freedom
## Multiple R-squared:  0.7969, Adjusted R-squared:  0.7967
## F-statistic:  3909 on 2 and 1992 DF,  p-value: < 2.2e-16
```

We see the same pattern such that including $\lambda(-\hat{\beta}_2 x_2)$ slightly improves the first OLS estimates, but the estimate for β_2 doesn't change much, again.

3.6.v

For $\rho_{\epsilon_1,\epsilon_2} = .9$:

```r
mu = c(0,0)
Sigma_e = matrix(c(1,.9,.9,1),2,2)
Sigma_x = matrix(c(1,0,0,1),2,2)
epsilon = mvrnorm(4000,mu,Sigma_e)
x = mvrnorm(4000,mu,Sigma_x)

x1 = x[,1]
x2 = x[,2]
e1 = epsilon[,1]
e2 = epsilon[,2]

y1 = 2 + 2 * x1 + e1
y2 = 4 * x2 + e2
z = ifelse(y2>0,1,0)

y1z = y1[z > 0]
x1z = x1[z > 0]
length(y1z)
```

```
## [1] 2023
```

```r
out6va = lm(y1z ~ x1z)
summary(out6va)
```

```
##
## Call:
## lm(formula = y1z ~ x1z)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1885 -0.6654  0.0247  0.6568  3.4358
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.16944    0.02191   99.00   <2e-16 ***
## x1z          2.00540    0.02208   90.81   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9853 on 2021 degrees of freedom
## Multiple R-squared:  0.8031, Adjusted R-squared:  0.803
## F-statistic:  8246 on 1 and 2021 DF,  p-value: < 2.2e-16
```

```
out6vb = glm(z ~ x2, family = binomial(link = probit))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(out6vb)
```

```
##
## Call:
## glm(formula = z ~ x2, family = binomial(link = probit))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1817  -0.0925   0.0000   0.1156   2.8537
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.001508   0.036975  -0.041    0.967
## x2           3.826362   0.133527  28.656   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5544.6  on 3999  degrees of freedom
## Residual deviance: 1481.8  on 3998  degrees of freedom
## AIC: 1485.8
##
## Number of Fisher Scoring iterations: 8
```

```
B2 = out6vb$coefficients[2]
B2
```

```
##       x2
## 3.826362
```

```
x2z = x2[z>0]
lambda = dnorm(-B2*x2z) / (1-pnorm(-B2*x2z))
out6vc = lm(y1z ~ x1z + lambda)
summary(out6vc)
```

```
##
```

```
## Call:
## lm(formula = y1z ~ x1z + lambda)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9939 -0.6248 -0.0126  0.6320  3.5348
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.98104    0.02358   84.00   <2e-16 ***
## x1z          2.01454    0.02076   97.03   <2e-16 ***
## lambda       0.92713    0.05658   16.39   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.926 on 2020 degrees of freedom
## Multiple R-squared:  0.8262, Adjusted R-squared:  0.8261
## F-statistic:  4803 on 2 and 2020 DF,  p-value: < 2.2e-16
```

Again, we see the same pattern but in this case the first OLS estimates (mainly α_1) are not as good as they were in the previous part.

3.7

Considering parts (5) and (6) as a whole, we see similar patterns. It seems that for both the correlations between x_1 and x_2 and between ϵ_1 and ϵ_2, when the magnitude of the correlation decreases, the OLS estimates have a greater accuracy. For part (6), we can see that when the correlation between the x's was zero, the β_1 estimation was not changing much when we incorporated $\lambda(-\hat{\beta}_2 x_2)$ which makes sense. Furthermore, besides this case, when the correlation is nonzero, including $\lambda(-\hat{\beta}_2 x_2)$ generally improves the estimates. On a similar note, when the correlation was 0, the lambda coefficient term was not as significant each time.