ASSIGNMENT 2

1. Eitas Rimkus (u184503)
2. Nikodem Baehr (u459229)
3. Samuel Friedlaender (u848264)

Theoretical Exercises

Solution Theoretical Exercise 1

1) Consider

$$\hat{\beta}_n^{WLS} = \underset{b}{\operatorname{argmin}}(\mathbf{y} - \mathbf{X}b)^\top W(\mathbf{y} - \mathbf{X}b) = \mathbf{y}^\top W\mathbf{y} - 2b^\top\mathbf{X}^\top W\mathbf{y} + b^\top\mathbf{X}^\top W\mathbf{X}b$$

Now, determine the First Order Condition with respect to b:

$\frac{\delta\hat{\beta}_n^{WLS}}{\delta b} = -2\mathbf{X}^\top W\mathbf{y} + 2\mathbf{X}^\top W\mathbf{X}b = 0$ So, $\mathbf{X}^\top W\mathbf{y} = \mathbf{X}^\top W\mathbf{X}b \iff b = (\mathbf{X}^\top W\mathbf{X})^{-1}\mathbf{x}^\top W\mathbf{y} = \hat{\beta}^{WLS}$

Since $X^\top WX = \sum_{i=1}^n w_iX_iX_i^\top$, $X'Wy = \sum_{i=1}^n w_iX_iY_i$, we have $\hat{\beta}_n^{WLS} = (\sum_{i=1}^n w_iX_iX_i^\top)^{-1}(\sum_{i=1}^n w_iX_iY_i) = (\frac{1}{n}\sum_{i=1}^n w_iX_iX_i^\top)^{-1}(\frac{1}{n}\sum_{i=1}^n w_iX_iY_i) = (\frac{1}{n}\sum_{i=1}^n w_iX_iX_i^\top)^{-1}(\frac{1}{n}\sum_{i=1}^n w_iX_iX_i^\top\beta + \frac{1}{n}\sum_{i=1}^n w_iX_i\epsilon_i)$

According to the Law of Large Numbers, we have

$$(\frac{1}{n}\sum_{i=1}^n w_iX_iX_i^\top)^{-1}(\frac{1}{n}\sum_{i=1}^n w_iX_iX_i^\top\beta) \xrightarrow{p} = E[(w_iX_iX_i^\top)^{-1}(w_iX_iX_i^\top)\beta] = E[\beta] = \beta$$

and, using the Law of Iterated Expectations and assumption A2*, we have

$$\frac{1}{n}\sum_{i=1}^n w_iX_i\epsilon_i \xrightarrow{p} E[w_iX_i\epsilon_i] = 0$$

Hence, $\hat{\beta}_n^{WLS} \to \beta$ in probability.

Now, we have

$$\sqrt{n}(\hat{\beta}_n^{WLS} - \beta) = (\frac{1}{n}\sum_{i=1}^n w_iX_iX_i^\top)^{-1}(\frac{1}{n}\sum_{i=1}^n w_iX_i\epsilon_i)$$

So, by assumption A2*, $E[X_1\epsilon_1] = 0$, and since $E[X_1X_1^\top w(X_1)]$ as well as $Var(X_1w(X_1)\epsilon_1)$ are finite and positive definite, using the Central Limit Theorem and Slutsky, we can show

$$\sqrt{n}(\hat{\beta}_n^{WLS}-\beta) \xrightarrow{d} (E[w_1X_1X_1^\top])^{-1}\times N(0,E[\epsilon_1^2w_1^2X_1X_1^\top]) = N(0,(E(w_1X_1X_1^\top))^{-1}E[\epsilon_1^2w_1^2X_1X_1^\top](E(w_1X_1X_1^\top))^{-1})$$

Hence, $\sqrt{n}(\hat{\beta}_n^{WLS}-\beta) \xrightarrow{d} N(0,V)$

2) Let $w(X_i) = (v(X_i))^{-1}$. Then, we have

$$E[X_1X_1^\top w_1^2\epsilon_1^2] = E[E[X_1X_1^\top w_1^2\epsilon_1^2|X_1]] = E[X_1X_1^\top w_1^2[\epsilon_1^2|X_1]]$$
$$= E[X_1X_1^\top w_1^2 v(X_1)] = E[X_1X_1^\top v(X_1)^{-2}v(X_1)] = E[E[X_1X_1^\top v(X_1)^{-1}]$$

Hence, the asymptotic variance is

$$V^* = (E[X_1X_1^\top v(X_1)^{-1}])^{-1}(E[X_1X_1^\top v(X_1)^{-1}])(E[X_1X_1^\top v(X_1)^{-1}])^{-1} = (E[X_1X_1^\top v(X_1)^{-1}])^{-1}$$

3)
$$V^{*-1} - V^{-1} = E[X_1X_1^\top v(X_1)^{-1}] - E[X_1X_1^\top w_1](E[X_1X_1^\top \epsilon^2 w_1^2])^{-1}E[X_1X_1^\top w_1]$$

Let $W = (E[X_1X_1^\top \epsilon_1^2 w_1^2])^{-1}E[X_1X_1^\top w_1]$

$$E[X_1X_1^\top v(X_1)^{-1}] - E[X_1X_1^\top w_1](E[X_1X_1^\top \epsilon_1^2 w_1^2])^{-1}E[X_1X_1^\top w_1]$$
$$= E[X_1X_1^\top v(X_1)^{-1}] - W^\top E[X_1X_1^\top w_1] - E[X_1X_1^\top w_1]W + W^\top E[X_1X_1^\top \epsilon_1^2 w_1^2]W$$
$$= \begin{bmatrix} I & -W^\top \end{bmatrix} \begin{bmatrix} E[X_1X_1^\top v(X_1)^{-1}] & E[X_1X_1^\top w_1] \\ E(X_1X_1^\top w_1) & E[X_1X_1^\top \epsilon_1^2 w_1^2] \end{bmatrix} \begin{bmatrix} I \\ -W \end{bmatrix} = \begin{bmatrix} I & -W^\top \end{bmatrix} E[Z_1Z_1^\top] \begin{bmatrix} I \\ -W \end{bmatrix}$$

Hence, $H = \begin{bmatrix} I \\ -W \end{bmatrix}$

4) By assumption V and V^* are positive definite and of the same size. We have $V^* - V = H^T E[Z_1Z_1^\top]H$ has a quadratic form, such that $x^\top Ax \geq 0$ for any $x \in \mathbb{R}^n$ and any symmetric $A \in \mathbb{R}^{n\times n}$. Since $EZ_1Z_1^\top$ is symmetric and, by definition, positive semidefinite, we have that $H^T E[Z_1Z_1^\top]H \succcurlyeq 0$, i.e. $H^T E[Z_1Z_1^\top]H$ is positive semidefinite. Hence, using the hint, provided, $V - V^*$ is positive semidefinite as well.

Solution Theoretical Exercise 2

1) Let $\theta = \begin{bmatrix} \beta^\top \\ \gamma \end{bmatrix}$.

Estimating $W_i = Z_i^\top \pi + v_i$ by OLS: $\pi_n^{OLS} = (\sum_{i=1}^n Z_iZ_i^\top)^{-1}(\sum_{i=1}^n Z_iW_i) = (Z^\top Z)^{-1}Z^\top W$

Computing the residuals $\hat{v}$: $\hat{v} = W - Z^\top \pi_n^{OLS} = W - Z^\top(Z^\top Z)^{-1}Z^\top W = (I - Z^\top(Z^\top Z)^{-1}Z^\top)W = M_Z W$

So, $\hat{v}_i = W_i - Z_i^\top(Z_i^\top Z_i)^{-1}Z_i^\top W_i$

2) Let $Y_i = X_i^\top \beta + \gamma W_i + \rho \hat{v}_i + u_i$, and $X' = \begin{bmatrix} X & W \end{bmatrix}$. Hence, we have $Y_i = \theta X' + \rho \hat{v}_i + u_i$.

Also, $X^\top \hat{v} = 0 \iff X^\top M_Z W = 0 \iff X^T(I - P_Z)W = 0 \iff X^\top W = X^\top P_Z W$. Hence, $X^\top = X^\top P_Z$ or $X = P_Z X$ *(Property 1)*.

Then, using the partitioned regression, we have:

$$\hat{\theta}_n^{OLS} = (X'^\top M_{\hat{v}} X')^{-1} X'^\top M_{\hat{v}} Y = (\begin{bmatrix} X^\top \\ W^\top \end{bmatrix} M_{\hat{v}} \begin{bmatrix} X & W \end{bmatrix})^{-1} \begin{bmatrix} X^\top \\ W \end{bmatrix} M_{\hat{v}} Y = \begin{bmatrix} X^\top M_{\hat{v}} X & X^\top M_{\hat{v}} W \\ W^\top M_{\hat{v}} X & W^\top M_{\hat{v}} W \end{bmatrix}^{-1} \begin{bmatrix} X^\top M_{\hat{v}} \\ W^\top M_{\hat{v}} \end{bmatrix} Y$$

Since $M_{\hat{v}} = I - P_{\hat{v}}$ and $X^\top \hat{v}_i = \hat{v}_i^\top X = 0$, we have

$$\hat{\theta}_n^{OLS} = \begin{bmatrix} X^\top X & X^\top W \\ W^\top X & W^\top (I - P_{\hat{v}}) W \end{bmatrix}^{-1} \begin{bmatrix} X^\top \\ W^\top - W^\top P_{\hat{v}} \end{bmatrix} Y$$

Now, consider $W^\top (I - P_{\hat{v}}) W = W^\top W - W^\top P_{\hat{v}} W = W^\top W - W^\top (\hat{v}(\hat{v}^\top \hat{v})^{-1} \hat{v}^\top) W = W^\top W - W^\top M_Z W (W^\top M_Z W)^{-1} W^\top M_Z W = W^\top W - W^\top M_Z W = W^\top W - W^\top (I - M_Z) W = W^\top W - W^\top (I - (I - P_Z)) W = W^\top P_Z W$.

Hence, together with *Property 1*, we have

$$\hat{\theta}_n^{OLS} = \begin{bmatrix} X^\top P_Z X & X^\top P_Z W \\ W^\top P_Z X & W^\top P_Z W \end{bmatrix}^{-1} \begin{bmatrix} X^\top P_Z \\ W^\top P_Z \end{bmatrix} Y = (\begin{bmatrix} X^\top \\ W^\top \end{bmatrix} P_Z \begin{bmatrix} X & W \end{bmatrix})^{-1} \begin{bmatrix} X^\top \\ W^\top \end{bmatrix} P_Z Y = (X'^\top P_Z X') X'^\top P_Z Y = \hat{\theta}_n^{2SLS}$$

Therefore, it is equal to the 2SLS estimator.

Solution Empirical Exercise 1

a)

```
#OLS estimation
X = cbind(iota, exper, exper2, black, educ)
B_hat = solve(t(X)%*%X)%*%t(X)%*%lwage
B_hat[5,]
```

```
##       educ
## 0.07436848
```

```
n = length(exper)
e = lwage - X%*%B_hat
s2_OLS = (1/(n-6))*drop(t(e)%*%e)
Q_OLS = (1/n)*t(X)%*%X
W_OLS = (1/n)*t(X)%*%X
VCE_OLS = drop(s2_OLS)*solve(t(Q_OLS)%*%solve(W_OLS)%*%Q_OLS)
SEE_OLS = sqrt(diag(VCE_OLS))
```

On average, a person that studies for a year longer compared to another person, is likely to have a higher wage by approximately 7.43%, ceteris paribus.

b)

```r
#IV estimation
Z_IV = cbind(iota, exper, exper2, black, meduc)
BIV_hat = solve(t(Z_IV)%*%X)%*%t(Z_IV)%*%lwage

epsiv = lwage - X%*%BIV_hat
s2_IV = (1/(n-6))*t(epsiv)%*%epsiv

omega_iv = drop(s2_IV)*t(Z_IV)%*%Z_IV
Q_iv = (1/n)*t(Z_IV)%*%X
W_iv = (1/n)*t(Z_IV)%*%Z_IV
VCE_IV = drop(s2_IV)*solve(t(Q_iv)%*%solve(W_iv)%*%Q_iv)
SEE_IV = sqrt(diag(VCE_IV))

report_IV = cbind(BIV_hat, SEE_IV)
colnames(report_IV) = cbind("Coefficient IV", "Standard Error IV")
report_IV
```

```
##        Coefficient IV Standard Error IV
## iota      4.683734581       10.43243181
## exper     0.007145986        0.42836012
## exper2    0.001270971        0.01935524
## black    -0.124633806        1.57085778
## educ      0.136147189        0.66628520
```

```r
comparison_1 = cbind(B_hat[5,], SEE_OLS[5], BIV_hat[5,], SEE_IV[5])
colnames(comparison_1) = cbind("OLS estimator", "OLS standard errors", "IV estimator", "IV standard err
comparison_1
```

```
##      OLS estimator OLS standard errors IV estimator IV standard errors
## educ    0.07436848           0.1959752    0.1361472          0.6662852
```

Estimated return to education using an IV regression compared to OLS is, on average, approximately 6 percentage points higher.

c)

```r
Z_2sls = cbind(iota, exper, exper2, black, meduc, feduc, sibs)
PZ_2sls = Z_2sls%*%solve(t(Z_2sls)%*%Z_2sls)%*%t(Z_2sls)
B2SLS_hat = solve(t(X)%*%PZ_2sls%*%X)%*%t(X)%*%PZ_2sls%*%lwage
eps2sls = lwage - X%*%B2SLS_hat
s2_2sls = (1/(n-6))*t(eps2sls)%*%eps2sls

Q = (1/n)*t(Z_2sls)%*%X
W = (1/n)*t(Z_2sls)%*%Z_2sls
VCE_2sls = drop(s2_2sls)*solve(t(Q)%*%solve(W)%*%Q)
SEE_2sls = sqrt(diag(VCE_2sls))

report_2sls = cbind(B2SLS_hat, SEE_2sls)
colnames(report_2sls) = cbind("2SLS coefficient", "2SLS standard error")
report_2sls
```

```
##        2SLS coefficient 2SLS standard error
## iota        4.789871091          8.37077393
## exper       0.007120449          0.42418871
## exper2      0.001202063          0.01875947
## black      -0.131712725          1.50224145
## educ        0.129185851          0.52691655
```

```r
comparison_2 = cbind(BIV_hat[5,], SEE_IV[5],B2SLS_hat[5,], SEE_2sls[5])
colnames(comparison_2) = cbind("IV estimator", "IV standard error", "2SLS estimator", "2SLS standard er
comparison_2
```

```
##      IV estimator IV standard error 2SLS estimator 2SLS standard error
## educ    0.1361472         0.6662852      0.1291859           0.5269165
```

A coefficient for instruments for education in IV regression is, on average, approximately 0.7 percentage points higher than compared to an 2SLS regression.

```r
omega_2sls = drop(s2_2sls)*t(Z_2sls)%*%Z_2sls
S_n = (1/n)*t(Z_2sls)%*%(lwage-(X%*%B2SLS_hat))
J_n = n*t(S_n)%*%solve(omega_2sls)%*%S_n
chi_val = qchisq(0.95,2)
reject_hypo = J_n > chi_val
reject_hypo
```

```
##       [,1]
## [1,] FALSE
```

For checking validity of suggested instrumental variables, we are using the overidentifying restrictions test. $H_0 : E[Z_i(Y_i - X_i^\top \beta)] = 0$ vs. $H_1 : E[Z_i(Y_i - X_i^\top \beta)] \neq 0$

The null hypothesis for the test states that the instrument variables are valid. Given that our J_n variable is smaller than $\chi^2_{0.95,2}$, we fail to reject the null hypothesis. Hence, we are 95% confident that the instruments are valid.

d) At first, we regress $W_i = Z_i^\top \pi + v_i$ Then, we use the first-stage residuals to estimate:

$$Y_i = X_i^\top \beta + \gamma W_i + \kappa \hat{v}_i + \epsilon_i$$

Then, we test $H_0 : \kappa = 0$ vs. $H_1 : \kappa \neq 0$

```
#First stage
pi = solve(t(Z_IV)%*%Z_IV)%*%t(Z_IV)%*%educ
v_error = as.vector(educ - Z_IV%*%pi)

#Second stage
C = cbind(X, v_error)
C_hat = solve(t(C)%*%C)%*%t(C)%*%lwage

#Hausman test
Y_hat = C%*%C_hat
u_error = lwage-Y_hat
N_obs = length(u_error)

s2=(1/(N_obs-6))*drop(t(u_error)%*%u_error)
var_cov=s2*solve(t(C)%*%C)
SSE=diag(var_cov)^0.5
zstat=C_hat[6,]/SSE[6]

#Two-sided z-test H_0: kappa=0 vs. H_1: kappa=!0
critical_val = qnorm(0.975, 0, 1)
Reject = abs(zstat)>critical_val
Reject
```

```
## v_error
##    TRUE
```

Since the sample size is large enough, we use a two-sided z-test at 95% confidence level. Because $|Z| = |\frac{\hat{\kappa}-0}{SEE_\kappa}| > z_{score}$, we reject the null hypothesis. Hence, *educ* is an endogenous variable, and therefore the IV estimation is required.