Empirical Analysis I - Problem Set 4

Timothy Schwieg
Paulo Henrique Ramos
Samuel Barker
Rafeh Qureshi

Question 1

First, we need to show that:

$$\rho^2 = \frac{(\mathbf{Cov}[X,Y])^2}{\mathbf{Var}[X]\mathbf{Var}[Y]} = 1 - \frac{\mathbf{Var}[U]}{\mathbf{Var}[Y]}.$$

Further, given $Y = \beta_0 + \beta_1 X + U$, and using the best linear prediction interpretation (which gives us $\mathbf{Cov}[X,U] = 0$ in this case):

$$\begin{aligned}\mathbf{Var}[Y] =& \mathbf{Var}[\beta_0 + \beta_1 X + U] \\ =& \beta_1^2 \mathbf{Var}[X] + \mathbf{Var}[U] + 2\mathbf{Cov}[X,U] \\ =& \left(\frac{\mathbf{Cov}[X,Y]}{\mathbf{Var}[X]}\right)^2 \mathbf{Var}[X] + \mathbf{Var}[U] \\ =& \frac{(\mathbf{Cov}[X,Y])^2}{\mathbf{Var}[X]} + \mathbf{Var}[U].\end{aligned}$$

This implies that:

$$\mathbf{Var}[U] = \mathbf{Var}[Y] - \frac{(\mathbf{Cov}[X,Y])^2}{\mathbf{Var}[X]}$$

Since we are interested in ρ^2 and want to show $\rho^2 = 1 - \frac{\mathbf{Var}[U]}{\mathbf{Var}[Y]}$, we need to divide by $\mathbf{Var}[Y]$.

$$\begin{aligned}\frac{\mathbf{Var}[U]}{\mathbf{Var}[Y]} =& 1 - \frac{(\mathbf{Cov}[X,Y])^2}{\mathbf{Var}[X]\mathbf{Var}[Y]} \\ =& 1 - \rho^2 \\ \rho^2 =& 1 - \frac{\mathbf{Var}[U]}{\mathbf{Var}[Y]}.\end{aligned}$$

Next, we need to actually determine $\mathbf{Var}[U]$ and $\mathbf{Var}[Y]$. We have shown earlier in lectures that $\beta = \mathbf{E}[XX']^{-1}\mathbf{E}[XY]$. Given the vector $(1, X)$, and $Y = \gamma X + X^2$, we can say:

$$\begin{aligned}\beta &= \mathbf{E}\left[\begin{pmatrix}1\\X\end{pmatrix}\begin{pmatrix}1 & X\end{pmatrix}\right]^{-1}\mathbf{E}\left[\begin{pmatrix}Y\\YX\end{pmatrix}\right]\\ &= \mathbf{E}\left[\begin{pmatrix}1 & X\\X & X^2\end{pmatrix}\right]^{-1}\mathbf{E}\left[\begin{pmatrix}\gamma X + X^2\\ \gamma X^2 + X^3\end{pmatrix}\right].\end{aligned}$$

Since we know $\mathbf{E}[1] = 1$, $\mathbf{E}[X] = 0$, $\mathbf{E}[X^2] = 1$, and $\mathbf{E}[X^3] = 0$, because $X \sim N(0, 1)$, the above is equal to:

$$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\begin{pmatrix}1\\\gamma\end{pmatrix} = \begin{pmatrix}1\\\gamma\end{pmatrix} = \beta.$$

It is given that $U = Y - \beta_0 - \beta_1 X$. Using the β above, we can plug in for $U = \gamma X + X^2 - 1 - \gamma X = X^2 - 1$. Thus, we have

$$\mathbf{Var}[U] = \mathbf{Var}[X^2 - 1] = 2.$$

Next, consider $Y = \gamma X + X^2$.

$$\mathbf{Var}[Y] = \gamma\mathbf{Var}[X] + \mathbf{Var}[X^2] + \gamma\mathbf{Cov}[X, X^2].$$

$\mathbf{Cov}[X, X^2] = 0$, since $\mathbf{Cov}[X, X^2] = \mathbf{E}[X^3] - \mathbf{E}[X]\mathbf{E}[X^2]$, and we know that both right hand side terms are zero. Further, since $\mathbf{Var}[X] = 1$, and $\mathbf{Var}[X^2] = 2$, we have

$$\mathbf{Var}[Y] = \gamma^2 + 2.$$

Using our previous result that $\rho^2 = 1 - \frac{\mathbf{Var}[U]}{\mathbf{Var}[Y]}$ and substituting in from above,

$$\rho^2 = 1 - \frac{\mathbf{Var}[U]}{\mathbf{Var}[Y]} = 1 - \frac{2}{\gamma^2 + 2} = \frac{\gamma^2}{\gamma^2 + 2}.$$

Question 2

a

In class we proved that there is no perfect colinearity in a matrix W if and only if $\mathbb{E}[WW']$ is invertible. So, if we show that $\mathbb{E}[WW']$ is invertible, then we have shown that there is no perfect colinearity in W.

Assume, to contrary, that $\mathbb{E}[WW']$ is not invertible. This means that there exists $c \neq 0$ such that $\mathbb{E}[WW']\,c = 0$. But then we have:

$$\begin{aligned}0 &= c'\mathbb{E}[WW']\,c\\ &= c'\mathbb{E}[AXX'A']\,c = \mathbb{E}[c'AXX'A'c]\\ &= \mathbb{E}[d'XX'd] = \mathbb{E}\left[(d'X)^2\right],\end{aligned} \tag{1}$$

where $d := A'c \neq 0$, because $c \neq 0$ and A is invertible. But because $(d'X)^2 \geq 0$ always, $\mathbb{E}\left[(d'X)^2\right] = 0$ implies $\Pr(d'X = 0) = 1$, contradicting the assumption that there is no perfect colinearity in X. Thus we cannot have $\mathbb{E}\left[WW'\right]c = 0$ for $c \neq 0$, making $\mathbb{E}\left[WW'\right]$ invertible, and implying the result we wanted.

b

Due to the first-order condition, $-2\mathbb{E}\left[X(Y - X'\beta)\right] = 0$ (and the assumptions that $\mathbb{E}\left[XX'\right]$ and $\mathbb{E}\left[XY\right]$ exist, and that there is no perfect colinearity in X), we have that

$$\beta = \mathbb{E}\left[XX'\right]^{-1}\mathbb{E}\left[XY\right]. \tag{2}$$

Similarly for $BLP(Y|W)$, the first order condition is $-2\mathbb{E}\left[W(Y - W'\gamma)\right] = 0$, which - using the no perfect colinearity of W and if $\mathbb{E}\left[WW'\right] = A\mathbb{E}\left[XX'\right]A'$ and $\mathbb{E}\left[WY\right] = A\mathbb{E}\left[XY\right]$ both exist (which is the case if A has only finite real values, since $\mathbb{E}\left[XX'\right]$ and $\mathbb{E}\left[XY\right]$ exist) - gives us that:

$$\begin{aligned}\gamma &= \mathbb{E}\left[WW'\right]^{-1}\mathbb{E}\left[WY\right] = (A\mathbb{E}\left[XX'\right]A')^{-1}A\mathbb{E}\left[XY\right] \\ &= A'^{-1}\mathbb{E}\left[XX'\right]^{-1}A^{-1}A\mathbb{E}\left[XY\right] = A'^{-1}\beta\end{aligned} \tag{3}$$

c

Define $\mathbf{W} := [W_i']$, a matrix with vectors W_i' as rows ($1 \leq i \leq n$), where $W_i' = X_i'A'$, so that $\mathbf{W} = \mathbf{XA'}$. Similarly, $\mathbf{X} := [X_i']$ and $\mathbf{Y} = [Y_i']$. Then our estimates of β and γ using OLS are:

$$\begin{aligned}\hat{\beta}_n &= (\mathbf{X'X})^{-1}\mathbf{X'Y} \\ \hat{\gamma}_n &= (\mathbf{W'W})^{-1}\mathbf{W'Y} = (\mathbf{AX'XA'})^{-1}\mathbf{AX'Y} \\ &= \mathbf{A'}^{-1}(\mathbf{X'X})^{-1}\mathbf{A}^{-1}\mathbf{AX'Y} = \mathbf{A'}^{-1}\hat{\beta}_n\end{aligned} \tag{4}$$

The conditionsl variance of the $\hat{\beta}_n$:

$$\begin{aligned}Var(\hat{\beta}_n|\mathbf{X}) &= Var((\mathbf{X'X})^{-1}\mathbf{X'Y}|\mathbf{X}) \\ &= Var((\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{X}\beta + \mathbf{u})|\mathbf{X}) \\ &= Var((\mathbf{X'X})^{-1}\mathbf{X'u}|\mathbf{X}) \\ &= (\mathbf{X'X})^{-1}\mathbf{X'}Var(\mathbf{u}|\mathbf{X})\mathbf{X}(\mathbf{X'X})^{-1}\end{aligned} \tag{5}$$

First notice that, because $\mathbf{A}$ is invertible, there is a one-to-one relation between $\mathbf{W}$ and $\mathbf{X}$. That is, given $\mathbf{X}$, we know $\mathbf{W} = \mathbf{XA'}$, and given $\mathbf{W}$, we know $\mathbf{X} = \mathbf{W}(\mathbf{A'}^{-1})$. They both have the same information.

Therefore, since any function of $\mathbf{W}$ can be written as a function of $\mathbf{X}$ and vice-versa, the space of functions of $\mathbf{W}$ is the same as the space of functions of $\mathbf{X}$. Then, if a function of $\mathbf{W}$ is the conditional expectation $\mathbb{E}\left[\hat{\gamma}_n^2|\mathbf{W}\right]$, then the same function is also $\mathbb{E}\left[\hat{\gamma}_n^2|\mathbf{X}\right]$. Similarly we have $(\mathbb{E}\left[\hat{\gamma}_n|\mathbf{W}\right])^2 = (\mathbb{E}\left[\hat{\gamma}_n|\mathbf{X}\right])^2$. Thus, because $Var(\hat{\gamma}_n|\mathbf{W}) = \mathbb{E}\left[\hat{\gamma}_n^2|\mathbf{W}\right] - (\mathbb{E}\left[\hat{\gamma}_n|\mathbf{W}\right])^2$, we have $Var(\hat{\gamma}_n|\mathbf{W}) = Var(\hat{\gamma}_n|\mathbf{X})$.

And since $\hat{\gamma}_n = \mathbf{A'}^{-1}\hat{\beta}_n$, we have:

$$\begin{aligned} Var(\hat{\gamma}_n|\mathbf{W}) = Var(\hat{\gamma}_n|\mathbf{X}) &= Var(\mathbf{A'}^{-1}\hat{\beta}_n|\mathbf{X}) \\ &= \mathbf{A'}^{-1}Var(\hat{\beta}_n|\mathbf{X})\mathbf{A}^{-1} \qquad (6) \\ &= \mathbf{A'}^{-1}(\mathbf{X'X})^{-1}\mathbf{X'}Var(\mathbf{u}|\mathbf{X})\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{A}^{-1} \end{aligned}$$

Question 3

a

$\Rightarrow$
Take $\mathbb{A}'\mathbb{X} = \mathbb{I}$

$$E(\tilde{\beta}|X_1, ..., X_n) = E(\mathbb{A}'\mathbb{Y}|X_1, ..., X_n) \qquad (7)$$
$$= \mathbb{A}'E(\mathbb{Y}|X_1, ..., X_n) \qquad (8)$$

Note, this follows as $E(f(x)y|x) = f(x)E(y|x)$. Now, we have that this becomes

$$\begin{aligned} &= \mathbb{A}'(E(Y_1|X_1, ..., X_n) \dots E(Y_n|X_1, ..., X_n))' \\ &= \mathbb{A}'(E(Y_1|X_1) \dots E(Y_n|X_n))' \\ &= \mathbb{A}'(X_1\beta \dots X_n\beta)' \\ &= \mathbb{A}'\mathbb{X}\beta \\ &= \mathbb{I}\beta = \beta \end{aligned}$$

The second equality above follows from the fact that (X_i, Y_i) is iid so Y_i is independent of all of the X_j for $j \neq i$.
$\Leftarrow$
Suppose $E(\tilde{\beta}|X_1, ..., X_n) = \beta$. Then, we have, from the equalities above, that

$$E(\tilde{\beta}|X_1, ..., X_n) = \mathbb{A}'\mathbb{X}\beta$$

Thus,

$$\begin{aligned} &\mathbb{A}'\mathbb{X}\beta - \beta = 0 \Rightarrow \\ &(\mathbb{A}'\mathbb{X} - \mathbb{I})\beta = 0 \end{aligned}$$

As this must hold true for all β, we must have that $\mathbb{A}\mathbb{X} = \mathbb{I}$ (i.e the only eigenvalue is $\lambda = 1$ for all $\beta \in \mathbb{R}^{k+1}$)

b

$$\begin{aligned} Var(\tilde{\beta}|X_1, ..., X_n) &= Var(\mathbb{A}'\mathbb{Y}|X_1, ..., X_n) \\ &= \mathbb{A}'(Var(\mathbb{Y}|X_1, ..., X_n))\mathbb{A} \end{aligned}$$

Note, again, since (X_i, Y_i) is iid so Y_i is independent of all of the X_j for $j \neq i$ and Y_i is independent of all of the Y_j for $j \neq i$, we can write the expression above as:

$$\begin{aligned}
&= \mathbb{A}'(diag(Var(Y_1|X_1)...Var(Y_n|X_n))\mathbb{A} \\
&= \mathbb{A}'diag(\sigma^2(X_1)...\sigma^2(X_n))\mathbb{A} \\
&= \mathbb{A}'\mathbb{D}\mathbb{A}
\end{aligned}$$

c

We take

$$\begin{aligned}
\mathbb{X}'\mathbb{D}^{-1}\mathbb{X} &= (X_1...X_n)diag(\frac{1}{\sigma^2(X_1)}, ..., \frac{1}{\sigma^2(X_n)})(X_1...X_n)' \\
&= \frac{X_1'X_1}{\sigma^2(X_1)} + ... + \frac{X_n'X_n}{\sigma^2(X_n)}
\end{aligned}$$

Now, note that as $\mathbb{X}$ has all its columns linearly independent, $\mathbb{X}a \neq 0 \Leftrightarrow a \neq 0$. Take such an a$\neq$ 0:

$$a'\mathbb{X}'\mathbb{D}^{-1}\mathbb{X}a = \frac{a_1^2X_1'X_1}{\sigma^2(X_1)} + ... + \frac{a_n^2X_n'X_n}{\sigma^2(X_n)}$$

We have that $a_i^2 \geq 0$ (and, by definition, $a_i^2 > 0$ for some i). Thus, the above sum is strictly positive. This shows that $\mathbb{X}'\mathbb{D}^{-1}\mathbb{X}$ is positive definite, which in turn establishes that it is invertible.

d

Take[1]

$$\begin{aligned}
Var(\tilde{\beta}|X_1, ..., X_n] &= \mathbb{A}'\mathbb{D}\mathbb{A} \\
&= (\mathbb{X}'\mathbb{D}^{-1}\mathbb{X})^{-1}\mathbb{X}'\mathbb{D}^{-1}\mathbb{D}\mathbb{D}^{-1}\mathbb{X}(\mathbb{X}\mathbb{D}'^{-1}\mathbb{X})^{-1} \\
&= (\mathbb{X}'\mathbb{D}^{-1}\mathbb{X})^{-1}\mathbb{X}'\mathbb{D}^{-1}\mathbb{X}(\mathbb{X}\mathbb{D}'^{-1}\mathbb{X})^{-1} \\
&= (\mathbb{X}'\mathbb{D}^{-1}\mathbb{X})^{-1}
\end{aligned}$$

Also, clearly:

$$((\mathbb{X}'\mathbb{D}^{-1}\mathbb{X})^{-1}\mathbb{X}'\mathbb{D}^{-1})\mathbb{X} = (\mathbb{X}'\mathbb{D}^{-1}\mathbb{X})^{-1}(\mathbb{X}'\mathbb{D}^{-1}\mathbb{X}) = \mathbb{I}$$

By (a), the estimator $\tilde{\beta}_n$ is then unbiased

[1] for a diagonal matrix D, D' = D

e

Take $\tilde{\mathbb{A}}'\mathbb{Y}$ as another unbiased estimator of β wherein $\gamma_n \equiv \tilde{\mathbb{A}}'\mathbb{Y}$. As $E(\gamma_n|X_1,...,X_n) = \beta$, we have that $\tilde{\mathbb{A}}'\mathbb{X} = \mathbb{I}$. Now, we run through the argument used in the Gauss-Markov theorem: Take $C = \tilde{\mathbb{A}} - \mathbb{A}$, then:

$$\begin{aligned}
Var(\gamma_n|X_1,...,X_n) - Var(\tilde{\beta}_n|X_1,...,X_n) &= (C+\mathbb{A})'\mathbb{D}(C+\mathbb{A}) - \mathbb{A}'\mathbb{D}\mathbb{A} \\
&= C'\mathbb{D}C + \mathbb{A}'\mathbb{D}C + C'\mathbb{D}\mathbb{A} \\
&= C'\mathbb{D}C + (\mathbb{X}'\mathbb{D}^{-1}\mathbb{X})^{-1}\mathbb{X}'\mathbb{D}^{-1}\mathbb{D}C + C'\mathbb{D}\mathbb{D}^{-1}\mathbb{X}(\mathbb{X}'\mathbb{D}^{-1}\mathbb{X}) \\
&= C'\mathbb{D}C + (\mathbb{X}'\mathbb{D}^{-1}\mathbb{X})^{-1}\mathbb{X}'C + C'\mathbb{X}(\mathbb{X}'\mathbb{D}^{-1}\mathbb{X}) \\
&= C'\mathbb{D}C + (\mathbb{X}'\mathbb{D}^{-1}\mathbb{X})^{-1}((\tilde{\mathbb{A}} - \mathbb{A})'\mathbb{X})' + (\tilde{\mathbb{A}} - \mathbb{A})'\mathbb{X}(\mathbb{X}'\mathbb{D}^{-1}\mathbb{X}) \\
&= C'\mathbb{D}C
\end{aligned}$$

where we used the unbiasedness of both estimators. Again, using the above argument in (c), we see that the differences in the variances is a positive semidefinite matrix s.t

$$a'C'\mathbb{D}Ca = \sum_{i=1}^{n} \frac{a_i^2}{\sigma^2(X_i)} \geq 0$$

This establishes that the estimator in the previous part is *best* in the Gauss-Markov sense.

Question 4

Let $\{(Y_i, X_i)\}_{i=1}^n$ be an i.i.d. sequence of random vectors. Suppose that $\mathbb{E}[X_iX_i']$ and $\mathbb{E}[X_iY_i]$ exists. Suppose further that there is no perfect colinearity in X_i, Hence $\mathbb{E}[X_iX_i']$ is invertible.

a

Does it also follow that

$$\frac{1}{n}\sum_{i=1}^{n} X_iX_i'$$

is invertible?

No. As a trivial case, consider when $n = 1, k = 2$ and $X_2 \sim \mathcal{N}(1,1)$. Let a be any realization of X_2.

$$\frac{1}{n}\sum_{i=1}^{n} X_iX_i' = (1,a)'(1,a) = \begin{pmatrix} 1 & a \\ a & a^2 \end{pmatrix}$$

We can see that the second column is a times the first column, and the matrix is not invertible. This occurs because for any vector $x \in \mathbb{R}^k$, xx' always has rank 1.

b

For any $\lambda_n > 0$ show that

$$\frac{1}{n}\sum_{i=1}^{n}(X_iX_i' + \lambda_n\mathbb{I})$$

is invertible.

Note that this can be rewritten as

$$\left[\frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right] + \lambda_n\mathbb{I}$$

For any given i, X_iX_i' is positive semi-definite. The sum of positive semi-definite matrices is also positive semi-definite. This tells us that the first matrix is always positive semi-definite.

$$\frac{1}{n}\sum_{i=1}^{n}X_iX_i' \succeq 0$$

It is obvious that $\lambda_n\mathbb{I}$ is a positive definite matrix. The sum of a positive definite matrix and a positive semi-definite matrix is positive definite.

Proof: Let A be a positive semi-definite matrix, and B be a positive definite matrix. Then $\forall x \in \mathbb{R}^k, x'Bx > 0$ and $x'Ax \geq 0$. Consider two cases:

Case 1: $x \in \mathbb{R}^k, x'Ax > 0, x'Bx > 0$. Then:

$$\begin{aligned}(x'A + x'B)\,X &> 0\\ x'(A+B)x &> 0\end{aligned}$$

Case 2: $x \in \mathbb{R}^k, x'Ax = 0, x'Bx > 0$ Then:

$$\begin{aligned}x'Ax + x'Bx &> 0\\ (x'A + x'B)\,X &> 0\\ x'(A+B)x &> 0\end{aligned}$$

This tells us that:

$$\left[\frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right] + \lambda_n\mathbb{I} \succ 0$$

Any positive definite matrix has strictly positive eigenvalues, and therefore has a strictly positive determinant. This implies that the matrix is invertible.

c

Suppose that $\lambda_n \to 0$ as $n \to \infty$. Find the limit in probability of

$$\tilde{\beta}_n = \left(\frac{1}{n}\sum_{i=1}^{n}(X_iX_i' + \lambda_n\mathbb{I})\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_iY_i\right)$$

From the weak law of large numbers, we know that $\frac{1}{n}\sum_{i=1}^{n}X_iX_i' \xrightarrow{p} \mathbb{E}[XX']$ and $\frac{1}{n}\sum_{i=1}^{n}X_iY_i \xrightarrow{p} \mathbb{E}[XY]$; and also $\lambda_n \xrightarrow{p} 0$ by assumption. Thus, using CMT, we can get the final result that $\tilde{\beta}_n \xrightarrow{p} \mathbb{E}[XX']^{-1}\mathbb{E}[XY]$.

We can also the same result through a different route. First, we wish to show that

$$\frac{1}{n}\sum_{i=1}^{n}X_iX_i' + \lambda_n\mathbb{I} \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n}X_iX_i'$$

Applying the definition of convergence in probability.

$$\lim_{n\to\infty}\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_iX_i' + \lambda_n\mathbb{I} - \frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right| < \epsilon\right) = \lim_{n\to\infty}\Pr(|\lambda_n\mathbb{I}| < \epsilon)$$

We will consider this on an element-wise basis. Note that if we are not on a diagonal, $(\lambda_n\mathbb{I})_{ij} = 0$. So we may restrict ourselves to the diagonal elements of this matrix. However all the diagonal elements are the same, so this question amounts to the convergence of $|\lambda_n|$. Since λ_n is non-random:

$$\lim_{n\to\infty}\Pr(|\lambda_n| < \epsilon) = 1$$

As we have assumed that $\lambda_n \to 0$ above.

Thus

$$\frac{1}{n}\sum_{i=1}^{n}X_iX_i' + \lambda_n\mathbb{I} \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n}X_iX_i' \xrightarrow{p} \mathbb{E}[XX']$$

As multiplication and inverting a matrix are continuous functions, we may apply the continuous mapping theorem to get that

$$\tilde{\beta}_n = \left(\frac{1}{n}\sum_{i=1}^{n}(X_iX_i' + \lambda_n\mathbb{I})\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_iY_i\right) \xrightarrow{p} \mathbb{E}[XX']^{-1}\mathbb{E}[XY] = \beta$$

Question 5

a)

Because we have $f(.)$ continuously differentiable at β with nonzero derivative, we can simply use the Delta Method to get:

$$n^{\frac{1}{2}}(f(\hat{\beta}_n) - f(\beta)) \xrightarrow{d} N(0, D_\beta f(\beta)\Omega D_\beta f(\beta)').$$

b)

With $\hat{\beta}_n$ and $\hat{\Omega}_n$ being consistent estimators of β and Ω respectively, using the CMT, we have $D_{\hat{\beta}_n} \xrightarrow{p} D_\beta f(\beta)$. Also, due to CMT, Ω being positive definite and the derivative being non zero, we have

$$\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\Omega D_{\hat{\beta}_n} f(\hat{\beta}_n)'} \xrightarrow{p} \sqrt{D_\beta f(\beta)\Omega D_\beta f(\beta)'}$$

.

Therefore, by slutsky, we have:

$$\frac{\sqrt{n}(f(\hat{\beta}_n) - f(\beta))}{\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\Omega D_{\hat{\beta}_n} f(\hat{\beta}_n)'}} \xrightarrow{d} N(0,1).$$

Since our test is one-sided we only want to reject the null (H_0: $f(\beta) \leq 0$) in one direction. The critical value is based on the standard normal distribution:

$$c_n := \Phi^{-1}(1-\alpha) := z_{1-\alpha},$$

where Φ is the CDF of $N(0,1)$. We want our c_n to be such that the probability of z (to which our test statistic T_n is converging in distribution) being less than c_n is $1-\alpha$. Thus, out test is:

$$\Phi_n = \mathbf{1}_{\{T_n > c_n\}}.$$

where $T_n = \frac{\sqrt{n} f(\hat{\beta}_n)}{\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\hat{\Omega}_n D_{\hat{\beta}_n} f(\hat{\beta}_n)'}}$. To show that this test is consistent in level, we have to show that, under the null:

$$\lim_{n\to\infty} \sup \mathbf{E}_P[\Phi_n] \leq \alpha$$

Consider,

$$\mathbf{E}_P[\Phi_n] = \mathbf{Pr}(T_n > c_n) = \mathbf{Pr}\left(\frac{\sqrt{n} f(\hat{\beta}_n)}{\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\hat{\Omega}_n D_{\hat{\beta}_n} f(\hat{\beta}_n)'}} > z_{1-\alpha}\right).$$

Add and subtract $f(\beta)$,

$$\mathbf{Pr}\left(\frac{\sqrt{n}(f(\hat{\beta}_n) - f(\beta))}{\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\hat{\Omega}_n D_{\hat{\beta}_n} f(\hat{\beta}_n)'}} + \frac{\sqrt{n} f(\beta)}{\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\hat{\Omega}_n D_{\hat{\beta}_n} f(\hat{\beta}_n)'}} > z_{1-\alpha}\right).$$

Under the null, we have that $f(\hat{\beta}_n) \leq 0$, and so

$$\mathbf{E}[\Phi_n] \leq \mathbf{Pr}\left(\frac{\sqrt{n}(f(\hat{\beta}_n) - f(\beta))}{\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\hat{\Omega}_n D_{\hat{\beta}_n} f(\hat{\beta}_n)'}} > z_{1-\alpha}\right) \leq \mathbf{Pr}\left(\frac{\sqrt{n}(f(\hat{\beta}_n) - f(\beta))}{\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\hat{\Omega}_n D_{\hat{\beta}_n} f(\hat{\beta}_n)'}} \geq z_{1-\alpha}\right),$$

where the weak inequality is so that we can apply the Portmanteau Lemma. Thus, taking lim sup of both sides,

$$\lim_{n\to\infty} \sup \mathbf{E}[\Phi_n] \leq \lim_{n\to\infty} \sup \mathbf{Pr}\left(\frac{\sqrt{n}(f(\hat{\beta}_n) - f(\beta))}{\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\hat{\Omega}_n D_{\hat{\beta}_n} f(\hat{\beta}_n)'}} \geq z_{1-\alpha}\right)$$

We already know that the inside the probability on the RHS converges in distribution to a standard normal. Thus,

$$\begin{aligned}
\lim_{n\to\infty} \sup \mathbf{E}[\Phi_n] &\leq \mathbf{Pr}\left(Z \geq z_{1-\alpha}\right) \\
&= 1 - \mathbf{Pr}(Z < z_{1-\alpha}) \\
&= 1 - \Phi(z_{1-\alpha}) \\
&= 1 - (1 - \alpha) \\
&= \alpha.
\end{aligned}$$

Our test is consistent in level.

c)

We can easily construct a confidence region with the result from **B)**.

$$\begin{aligned}
C_n &:= \left\{ x \in \mathbb{R} | \left(\frac{\sqrt{n}|f(\hat{\beta}_n) - x|}{\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\hat{\Omega}_n D_{\hat{\beta}_n} f(\hat{\beta}_n)'}} \leq z_{1-\frac{\alpha}{2}}\right)\right\} \\
C_n &:= \left[f(\hat{\beta}_n) - z_{1-\frac{\alpha}{2}} \frac{\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\hat{\Omega}_n D_{\hat{\beta}_n} f(\hat{\beta}_n)'}}{\sqrt{n}}, f(\hat{\beta}_n) + z_{1-\frac{\alpha}{2}} \frac{\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\hat{\Omega}_n D_{\hat{\beta}_n} f(\hat{\beta}_n)'}}{\sqrt{n}}\right]
\end{aligned} \tag{9}$$

Now can show the result we want, using the convergence in distribution established in the last item:

$$\begin{aligned}
\Pr(f(\beta) \in C_n) &= 1 - \Pr\left(\frac{\sqrt{n}|f(\hat{\beta}_n) - f(\beta)|}{\sqrt{D_{\hat{\beta}_n} f(\hat{\beta}_n)\hat{\Omega}_n D_{\hat{\beta}_n} f(\hat{\beta}_n)'}} \geq z_{1-\frac{\alpha}{2}}\right) \\
&\to 1 - \Pr(|z| \geq z_{1-\frac{\alpha}{2}}) = 1 - \alpha
\end{aligned} \tag{10}$$

Question 6

a

Due to the first-order condition of $BLP(Y_i|W_i)$, we have $\mathbb{E}[W_iU_i] = 0$, which is equivalent to $\mathbb{E}[U_i] = 0$, $\mathbb{E}[X_iU_i] = 0$ and $\mathbb{E}[Z_iU_i] = 0$. Therefore, we have that $Cov(U_i, W_i) = \mathbb{E}[U_iW_i] - \mathbb{E}[U_i]\mathbb{E}[W_i] = 0$ and thus W_i and U_i are uncorrelated.

In this case, they are also mean independent. Since $\mathbb{E}[Y_i|W_i] = W_i'\beta$ is the best predictor of Y_i and it is also linear, then $BLP(Y_i|W_i) = W_i'\beta$, and we have $Y_i = W_i'\beta + U_i$. Taking conditional expectations we get:

$$\mathbb{E}[Y_i|W_i] = W_i'\beta + \mathbb{E}[U_i|W_i] \Longrightarrow \mathbb{E}[U_i|W_i] = 0. \tag{11}$$

Thus U_i is mean independent of W_i. This is due to, in this case, the best linear predictor being actually equal to the conditional expectation.

b

Since, as seen in letter (a) above, $Y_i = W_i'\beta + U_i$, we have that:

$$Var(U_i|W_i) = Var(Y_i - W_i'\beta|W_i) = Var(Y_i|W_i) \tag{12}$$

Also, we have that $Var(U_i|W_i) = \mathbb{E}[U_i^2|W_i] - (\mathbb{E}[U_i|W_i])^2 = \mathbb{E}[U_i^2|W_i]$, since we have shown $\mathbb{E}[U_i|W_i] = 0$ above. Homoskedasticity would mean both $\mathbb{E}[U_i|W_i] = 0$ and $\mathbb{E}[U_i^2|W_i]$ not depending on W_i.

Because $\mathbb{E}[Y_i|W_i] = W_i'\beta$, and Y_i takes values in $\{0, 1\}$, we then have $Y_i|W_i$ distributed as bernoulli with $p = W_i'\beta$. This implies that $Var(U_i|W_i) = Var(Y_i|W_i) = W_i'\beta(1 - W_i'\beta)$. Thus $Var(U_i|W_i) = \mathbb{E}[U_i^2|W_i]$ depends on W_i, unless $\beta = 0$, making it unreasonable to assume homoskedasticity, since this would imply W_i is not useful in predictiong Y_i, and therefore our model is flawed from the start.

c

Define $W := [W_i']$, that is, a matrix with W_i' as its rows ($1 \le i \le n$). Similarly define $Y := [Y_i]$. With that, the OLS estimator of β is:

$$\hat{\beta}_n = (W'W)^{-1}W'Y \tag{13}$$

Now take the conditional expectation of $\hat{\beta}_n$:

$$\mathbb{E}\left[\hat{\beta}_n|W\right] = \mathbb{E}\left[(W'W)^{-1}W'(W\beta + U)|W\right] = \beta + (W'W)^{-1}W'\mathbb{E}[U|W] = \beta \tag{14}$$

the last equality being due to $\mathbb{E}[U|W] = 0$, since $\mathbb{E}[U_i|W_i] = 0$ and the observations are i.i.d.. Thus we have conditional unbiasedness of $\hat{\beta}_n$. Applying the Law of Iteraded expectations we obtain unconditional as well: $\mathbb{E}\left[\hat{\beta}_n\right] = \mathbb{E}\left[\mathbb{E}\left[\hat{\beta}_n|W\right]\right] = \mathbb{E}[\beta] = \beta$.

d

We can use CLT, WLLN and CMT to show:

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_n - \beta) &= \sqrt{n}\Big((\frac{1}{n}\sum_{i=1}^{n} W_i W_i')^{-1} (\frac{1}{n}\sum_{i=1}^{n} W_i Y_i) - \beta \Big) \\
&= \sqrt{n}\Big((\frac{1}{n}\sum_{i=1}^{n} W_i W_i')^{-1} (\frac{1}{n}\sum_{i=1}^{n} W_i (W_i'\beta + U_i)) - \beta \Big) \\
&= \sqrt{n}\Big(\beta + (\frac{1}{n}\sum_{i=1}^{n} W_i W_i')^{-1} (\frac{1}{n}\sum_{i=1}^{n} W_i U_i) - \beta \Big) \\
&= \Big(\frac{1}{n}\sum_{i=1}^{n} W_i W_i'\Big)^{-1} \Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i U_i\Big) \\
&\xrightarrow{d} N(0, \mathbb{E}\left[W_i W_i'\right]^{-1} Var(W_i U_i) \mathbb{E}\left[W_i W_i'\right]^{-1}) = N(0, \Omega),
\end{aligned} \tag{15}$$

where we define Ω accordingly. The last result is due to: (i) $\Big(\frac{1}{n}\sum_{i=1}^{n} W_i W_i'\Big)^{-1} \xrightarrow{p} \mathbb{E}\left[W_i W_i'\right]^{-1}$ due to WLLN and CMT; (ii) $\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i U_i\Big) \xrightarrow{d} N(0, Var(W_i U_i))$ due to CLT and $\mathbb{E}\left[W_i U_i\right] = 0$, as we have shown above; (iii) using slutsky and the fact that $\mathbb{E}\left[W_i W_i'\right]^{-1}$ is symmetric we get the final result.

Because it is not reasonable to assume homoskedasticity, a consistent estimate of Ω would be:

$$\hat{\Omega}_n = \Big(\frac{1}{n}\sum_{i=1}^{n} W_i W_i'\Big)^{-1} \Big(\frac{1}{n}\sum_{i=1}^{n} W_i W_i' \hat{U}_i^2\Big) \Big(\frac{1}{n}\sum_{i=1}^{n} W_i W_i'\Big)^{-1}, \tag{16}$$

where $\hat{U}$ are the residuals, since we do not know the true errors. In words, we are substituting the terms in Ω by sample analogs. In class we proved this leads to a consistent estimator. We will use this fact below.

Now, using the CMT and $r := \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}'$ we have:

$$\begin{aligned}
&\sqrt{n}(r'\hat{\beta}_n - r'\beta) \xrightarrow{d} N(0, r'\Omega r) \\
&\sqrt{n}(\hat{\beta}_{n,2} - \beta_2) \xrightarrow{d} N(0, \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \Omega \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}) = N(0, \Omega_{3,3}),
\end{aligned} \tag{17}$$

where $\Omega_{3,3}$ is the $(3,3)$ element of the matrix Ω.

Because the function that maps from a vector to a coordinate is continuous, we can use the CMT to conclude that $\hat{\Omega}_{3,3} \xrightarrow{p} \Omega_{3,3}$, where $\hat{\Omega}_{3,3}$ is the $(3,3)$ element of the matrix $\hat{\Omega}_n$. Again by the CMT we have $\sqrt{\hat{\Omega}_{3,3}} \xrightarrow{p} \sqrt{\Omega_{3,3}}$.

Using Slutsky (and CMT when applying the absolute value function), we then have:

$$\frac{\sqrt{n}(|\hat{\beta}_{n,2} - \beta_2|)}{\sqrt{\hat{\Omega}_{3,3}}} \xrightarrow{d} |N(0,1)|, \tag{18}$$

Then, to test the null $H_0 : \beta_2 = 0$, we could use the test statistic $T_n := \frac{\sqrt{n}(|\hat{\beta}_{n,2}|)}{\sqrt{\hat{\Omega}_{3,3}}}$, and reject the null if $T_n > z_{1-\frac{\alpha}{2}}$. Then we have, using Portmanteau:

$$\begin{aligned} \limsup \Pr(T_n > z_{1-\frac{\alpha}{2}}) &\leq \limsup \Pr(Tn \geq z_{1-\frac{\alpha}{2}}) \\ &\leq \Pr(|z| \geq z_{1-\frac{\alpha}{2}}) = \alpha, \end{aligned} \tag{19}$$

where $|z|$ is standard normal, and $z_{1-\frac{\alpha}{2}}$ the $1-\frac{\alpha}{2}$ quantile. Thus, we have a test consistent at level α.

e

We know that, because the regression without Z_i still has a constant, the estimate for β_1 would be (where variables with overlines are sample means):

$$\begin{aligned} \hat{\beta}_{n,1} &= \Big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n)\Big) \\ &= \Big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(\beta_0 + X_i\beta_1 + Z_i\beta_2 + U_i - \beta_0 - \overline{X}_n\beta_1 - \overline{Z}_n\beta_2 - \overline{U}_n)\Big) \\ &= \Big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)((X_i - \overline{X}_n)\beta_1 + (Z_i - \overline{Z}_n)\beta_2 + (U_i - \overline{U}_n))\Big) \\ &= \beta_1 + \beta_2\frac{\big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(Z_i - \overline{Z}_n)\big)}{\big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\big)} + \frac{\big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(U_i - \overline{U}_n)\big)}{\big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\big)} \end{aligned} \tag{20}$$

We know the last term in the last line converges in probabiliy to $\frac{Cov(X_i,U_i)}{Var(X_i)} = \frac{\mathbb{E}[(X_i-\mathbb{E}[X_i])(U_i-\mathbb{E}[U_i])]}{\mathbb{E}[(X_i-\mathbb{E}[X_i])^2]}$ due to WLLN and CMT, since observations are i.i.d.. But we also know, from item (a) above, that $Cov(X_i, U_i) = 0$; thus, the last term is converging in probability to zero. Also, again using WLLN and CMT, since observations are i.i.d., we have that the second fraction in the last line in converging in probability to $\beta_2\frac{Cov(X_i,Z_i)}{Var(X_i)} = \beta_2\frac{\mathbb{E}[(X_i-\mathbb{E}[X_i])(Z_i-\mathbb{E}[Z_i])]}{\mathbb{E}[(X_i-\mathbb{E}[X_i])^2]}$.

Therefore, for $\hat{\beta}_{n,1} \xrightarrow{p} \beta_1$ to hold, we need either $\beta_2 = 0$ or $Cov(X_i, Z_i) = 0$. That is, for the estimate of β_1 omitting Z_i to be consistent, we need X_i and Z_i to be uncorrelated.

f

Using $\hat{\beta}_n = (W'W)^{-1}W'Y$, we have that:

$$Var(\hat{\beta}_n|W) = (W'W)^{-1}W'Var(U|W)W(W'W)^{-1} \tag{21}$$

We know $Var(U|W)$ is a matrix with diagonal elements equal to $W_i'\beta(1 - W_i'\beta)$, and off-diagonal elements zero, because the observations are i.i.d.. We do not know β, but we can proceed by first obtaining an OLS estimate $\hat{\beta}_n = (W'W)^{-1}W'Y$. Then we use this OLS estimate to estimate $Var(U|W)$ by the matrix $\hat{\Omega}_n$ that has $W_i'\hat{\beta}_n(1 - W_i'\hat{\beta}_n)$ in its diagonals and zero off-diagonals. Because OLS is consistent, using the CMT we obtain that $\hat{\Omega}_n \xrightarrow{p} Var(U|W)$.

Now we reestimate β using $\hat{\beta}_n^* = (W'\hat{\Omega}_n^{-1}W)^{-1}W'\hat{\Omega}_n^{-1}Y$. This gives us that:

$$\begin{aligned} Var(\hat{\beta}_n^*|W) &= (W'\hat{\Omega}_n^{-1}W)^{-1}W'\hat{\Omega}_n^{-1}Var(U|W)\hat{\Omega}_n^{-1}W(W'\hat{\Omega}_n^{-1}W)^{-1} \\ &\xrightarrow{p} (W'\Omega^{-1}W)^{-1}W'\Omega^{-1}\Omega\Omega^{-1}W(W\Omega^{-1}W)^{-1} = (W\Omega^{-1}W)^{-1} \end{aligned} \tag{22}$$

Thus, by the results of question 2, using Ω as our D, we obtain an estimator whose variance converges to the best variance possible among unbiased estimators, in the gauss-markov sense.

Question 7

a

Note that this follows simply from the random assignment. Because individuals are not aware of their assignment before the experiment and have equal likelihood of being assigned to treatment or control groups, their probability of being assigned to the treatment is independent of their α_i and β_i. Thus, D_i is independent of (a_i, β_i)

b

Note, that we can write down the β (i.e from class) as $Var(D_i)^{-1}Cov(D_i, Y_i)$ for the special case of a bivariate regression with constant.

$$\begin{aligned} \beta &= Var(D_i)^{-1}Cov(D_i, Y_i) \\ &= Var(D_i)^{-1}(Cov(D_i, \alpha_i) + Cov(\beta_i D_i, D_i)) \end{aligned}$$

Since (α_i, β_i) are independent of D_i, the first term is just 0, and the second term is $Cov(\beta_i D_i, D_i) = \mathbb{E}[\beta_i]\mathbb{E}[D_i^2] - \mathbb{E}[\beta_i]\mathbb{E}[D_i]^2 = E(\beta_i)Var(D_i)$. Thus, we have:

$$\begin{aligned} \beta &= Var(D_i')^{-1}E(\beta_i)Var(D_i') \\ &= E(\beta_i) \end{aligned}$$

Using the above, note we can also solve for α:

$$\alpha = E(y - \beta D_i)$$

$$\begin{aligned} \alpha &= E(Y_i) - \beta E(D_i) \\ &= E(\alpha_i + \beta_i D_i) - \beta E(D_i) \\ &= E(\alpha_i) + E(\beta_i D_i) - \beta E(D_i) \end{aligned}$$

Again, by the independence of β_i and D_i, we get that this equals

$$\begin{aligned} &= E(\alpha_i) + E(\beta_i)E(D_i) - \beta E(D_i) \\ &= E(\alpha_i) + \beta E(D_i) - \beta E(D_i) \\ &= E(\alpha_i) \end{aligned}$$

c

Using the results above, we can construct the error term as: $\epsilon_i = \alpha_i - \mathbb{E}[\alpha_i] + (\beta_i - \mathbb{E}[\beta_i])D_i$. Also, since we aren't given homoskedasticity[2], we can use robust standard errors:

$$\begin{aligned} \hat{\Omega} &= (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \hat{\epsilon}_i)(\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1} \\ &\xrightarrow{p} \Omega \end{aligned}$$

From the result in class we know that: $\sqrt{n}((\hat{\alpha}, \hat{\beta})' - (\alpha, \beta)') \xrightarrow{d} N(0, \Omega)$.
Thus, using $r = (0,1)'$, we get by CMT:

$$\begin{aligned} \sqrt{n}(r'(\hat{\alpha}, \hat{\beta})' - r'(\alpha, \beta)') &\xrightarrow{d} N(0, r'\Omega r) \\ \sqrt{n}(\hat{\beta} - \beta) &\xrightarrow{d} N(0, \Omega_{2,2}) \end{aligned} \tag{23}$$

Where (2,2) indicates de element of the matrix. As $\hat{\Omega}$ is a consistent estimator of Ω (from class and delineated above), we have, by CMT and Slutsky, that: $\frac{\sqrt{n}(\hat{\beta}-\beta)}{\sqrt{\hat{\Omega}_{2,2}}} \xrightarrow{d} N(0,1)$. Thus, we can construct the confidence region:

$$C_n = [\hat{\beta}_n - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\hat{\Omega}_{2,2}}{n}}, \hat{\beta}_n + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\hat{\Omega}_{2,2}}{n}}]$$

And, because $\frac{\sqrt{n}|\hat{\beta}-\beta|}{\sqrt{\hat{\Omega}_{2,2}}} \xrightarrow{d} |N(0,1)|$, by CMT, this confidence interval has level $1-\alpha$ as we wanted.

Question 8

a

```
data <- read.csv( "ps4.csv" )

k <- ncol(data)
```

[2] In fact, in this setup it is probable that $Var(\epsilon_i)$ is not constant as α_i and β_i depend on i, and indeed, $Var(\epsilon_i|D_i) = (1-D_i)Var(\alpha_i) + D_i Var(\alpha_i + \beta_i)$

```r
N <- nrow(data)

## Since we are not calling lm, we want to do matrix algebra, we need
## R to not store this stuff as a data frame. What a terrible language.

Y <-  as.matrix(data$y)
X <- as.matrix(cbind( rep(1,N), data[,2:3] ))

## Remember that matrix multiplication uses the %*%
mat <-  t(X)%*%X

## Rather than using inverses, let's be numerically stable and use the
## Cholesky decomp and forward/back substitution for legitimate answers
F <- chol(mat)

## We now have $X'X\beta = X'Y$
## This is equivalent to $F'F\beta = X'Y$
## Thus $\beta = F^{-1}F'^{-1}X'Y$

## Note that F' is lower triangular so we use forward substitution.
beta <-  backsolve( F, forwardsolve( t(F), t(X)%*%Y ) )
```

Our estimated values of β are: $(0.1680066, 1.0843565, 0.9203671)'$.

b

```r
## Now lets build our variance estimates.

outerproduct <- function( row ){
    row%*%t(row)
}

## We are interested in estimating $\left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1}$


## The inner apply() forms the outer product matrices, the outer
## averages over them The matrix() reforms them as a matrix since
## apply flattens them.  This is equivalent to just doing
## $\frac{1}{n}X'X$, I just wanted some R practice.
outerProductGradient <- matrix( apply( apply( X, 1, outerproduct ), 1,
                                      mean ), nrow = k, ncol = k )

## Mama told me to never invert a matrix on a computer
varF <- chol( outerProductGradient )
```

```r
informationEstimate <- backsolve( varF, forwardsolve( t(varF), diag(k) ) )

## Now lets get the heteroskedasticity-robust version of this bad boy.
## We multiply the matrix of $X_i X_i'$ by $\hat{u}_i^2$ component wise, hence no %
monstronsity <- matrix( apply(
    matrix( rep( (Y - X%*%beta)^2, k*k ), nrow=k*k, ncol = N, byrow = TRUE )
     * apply( X, 1, outerproduct ), 1, mean ), nrow = k, ncol = k )

## This is what are interested in: $\mathbb{V}\left(\widehat{\beta}_N|X\right)$
condVarHetero <- informationEstimate%*%monstronsity%*%informationEstimate


## Note that it's  possible to just use matrix operations to get there
## I just chose this way for practice and to have it look like the notes.
## One could always do $(X'X)^{-1}X'\widehat{\Sigma}_N X(X'X)^{-1}$
```

Our estimated Variance-Covariance Matrix of $\widehat{\beta}_N$ is:

$$\mathbb{V}\left(\widehat{\beta}_N|X\right) = \begin{pmatrix} 4.8905355 & 0.4493318 & -1.6478739 \\ 0.4493318 & 0.4517238 & -0.3702895 \\ -1.6478739 & -0.3702895 & 0.7567006 \end{pmatrix}$$

c

```r
## Now we face multiple linear restrictions in the form of $R\beta = r$

## We don't really know anything about the nature of $R\mathbb{V}\left(\widehat{\beta}_N\right)$
## So we can't rely on any decompositions, and we'll let solve() work here
multipleLinearTest <- function( R, r, N, beta, Var ){
    N*t(R%*%beta - r )%*%solve(R%*%Var%*%t(R))%*%(R%*%beta -r  )
}


R <-  matrix( c( 0, 0, 1, 0 ,0,1 ), nrow = 2, ncol = 3 )
r <- c(  1, 1 )

## This is free to be changed.
alpha <- .05

## This c is the critical value used in a hypothesis test
c <- qchisq( alpha, df = 2, lower.tail = FALSE )


```

```r
testStat <- multipleLinearTest( R, r, N, beta, condVarHetero )
pValue <-  pchisq( testStat, df = 2, lower.tail = FALSE )
```

Our test statistic value is 1.599558 and our p-value is: 0.4494283

d

```r
## Testing: $f(\beta) = (\beta_1 - \beta_2)^2 = 0$
## However we need the rows of the total derivative to be linearly independent.
## $\nabla f(\beta) = (0, 2(\beta_1 - \beta_2), -2(\beta_1 - \beta_2))'$
## The rows are not linearly independent - The standard nonlinear test will not work.

## Worse yet, if we attempt to simply take the square root of both
## sides we lose the reliability as this is a Wald-Test. Wald Tests
## are not invariant to non-linear Transforms. This means we want to
## use a likelihood-ratio test, which is. However if we do not want to
## assume normality of Y and then the GLM framework to get a
## likelihood-ratio test, we can just stand for the errors in the Wald Test.

## Our test is simply testing if $\beta_1 - \beta_2 = 0$

R <- matrix( c( 0, 1, -1 ), nrow = 1, ncol = 3 )
r <- c(0)

## I just copy and pasted the previous code
c <- qchisq( alpha, df = 1, lower.tail = FALSE )
testStat <- fischerFTest( R, r, N, beta, condVarHetero )
pValue <-  pchisq( testStat, df = 1, lower.tail = FALSE )
```

Our test statistic value is 1.379809 and our p-value is: 0.2401337