

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:

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

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

b

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

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

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

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

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 $\mathbf{W} = \mathbf{X} \mathbf{A}'$. Similarly, $\mathbf{X} := [X_i']$ and $\mathbf{Y} = [Y_i']$. Then our estimates of β and γ using OLS are:

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

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

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

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 $\mathbf{W} = \mathbf{X}\mathbf{A}'$, 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 $\mathbb{E}[\hat{\gamma}_n^2|\mathbf{W}]$, then the same function is also $\mathbb{E}[\hat{\gamma}_n^2|\mathbf{X}]$. Similarly we have $(\mathbb{E}[\hat{\gamma}_n|\mathbf{W}])^2 = (\mathbb{E}[\hat{\gamma}_n|\mathbf{X}])^2$. Thus, because $\text{Var}(\hat{\gamma}_n|\mathbf{W}) = \mathbb{E}[\hat{\gamma}_n^2|\mathbf{W}] - (\mathbb{E}[\hat{\gamma}_n|\mathbf{W}])^2$, we have $\text{Var}(\hat{\gamma}_n|\mathbf{W}) = \text{Var}(\hat{\gamma}_n|\mathbf{X})$.

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

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

Question 6

a

Due to the first-order condition of $BLP(Y_i|W_i)$, we have $\mathbb{E}[W_i U_i] = 0$, which is equivalent to $\mathbb{E}[U_i] = 0$, $\mathbb{E}[X_i U_i] = 0$ and $\mathbb{E}[Z_i U_i] = 0$. Therefore, we have that $\text{Cov}(U_i, W_i) = \mathbb{E}[U_i W_i] - \mathbb{E}[U_i]\mathbb{E}[W_i] = 0$ 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:

$$\mathbb{E}[Y_i|W_i] = W_i'\beta + \mathbb{E}[U_i|W_i] \implies \mathbb{E}[U_i|W_i] = 0. \tag{7}$$

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:

$$\text{Var}(U_i|W_i) = \text{Var}(Y_i - W_i'\beta|W_i) = \text{Var}(Y_i|W_i) \tag{8}$$

Also, we have that $\text{Var}(U_i|W_i) = \mathbb{E}[U_i^2|W_i] - (\mathbb{E}[U_i|W_i])^2 = \mathbb{E}[U_i^2|W_i]$, 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 $\text{Var}(U_i|W_i) = \text{Var}(Y_i|W_i) = W_i'\beta(1 - W_i'\beta)$. Thus $\text{Var}(U_i|W_i) = \mathbb{E}[U_i^2|W_i]$ depends on W_i , unless $\beta = 0$, making it unreasonable to assume homoskedasticity, since this would imply W_i is not useful in predicting 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 ($1 \leq i \leq n$). Similarly define $Y := [Y_i]$. With that, the OLS estimator of β is:

$$\hat{\beta}_n = (W'W)^{-1}W'Y \quad (9)$$

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

$$\mathbb{E}[\hat{\beta}_n|W] = \mathbb{E}[(W'W)^{-1}W'(W\beta + U)|W] = \beta + (W'W)^{-1}W'\mathbb{E}[U|W] = \beta \quad (10)$$

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 Iterated expectations we obtain unconditional as well: $\mathbb{E}[\hat{\beta}_n] = \mathbb{E}[\mathbb{E}[\hat{\beta}_n|W]] = \mathbb{E}[\beta] = \beta$.

d

We can use CLT, WLLN and CMT to show:

$$\begin{aligned} \sqrt{n}(\hat{\beta}_n - \beta) &= \sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^n W_i W_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n W_i Y_i\right) - \beta\right) \\ &= \sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^n W_i W_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n W_i(W_i'\beta + U_i)\right) - \beta\right) \\ &= \sqrt{n}\left(\beta + \left(\frac{1}{n}\sum_{i=1}^n W_i W_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n W_i U_i\right) - \beta\right) \\ &= \left(\frac{1}{n}\sum_{i=1}^n W_i W_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i U_i\right) \\ &\xrightarrow{d} N(0, \mathbb{E}[W_i W_i']^{-1} \text{Var}(W_i U_i) \mathbb{E}[W_i W_i']^{-1}) = N(0, \Omega), \end{aligned} \quad (11)$$

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

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

$$\hat{\Omega}_n = \left(\frac{1}{n} \sum_{i=1}^n W_i W_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n W_i W_i' \hat{U}_i^2 \right) \left(\frac{1}{n} \sum_{i=1}^n W_i W_i' \right)^{-1}, \quad (12)$$

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 $r := (0 \ 0 \ 1)'$ we have:

$$\begin{aligned} \sqrt{n}(r' \hat{\beta}_n - r' \beta) &\xrightarrow{d} N(0, r' \Omega r) \\ \sqrt{n}(\hat{\beta}_{n,2} - \beta_2) &\xrightarrow{d} N(0, (0 \ 0 \ 1) \Omega \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}) = N(0, \Omega_{3,3}), \end{aligned} \quad (13)$$

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:

$$\frac{\sqrt{n}(|\hat{\beta}_{n,2} - \beta_2|)}{\sqrt{\hat{\Omega}_{3,3}}} \xrightarrow{d} |N(0, 1)|, \quad (14)$$

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:

$$\begin{aligned} \limsup \Pr(T_n > z_{1-\frac{\alpha}{2}}) &\leq \limsup \Pr(T_n \geq z_{1-\frac{\alpha}{2}}) \\ &\leq \Pr(|z| \geq z_{1-\frac{\alpha}{2}}) = \alpha, \end{aligned} \quad (15)$$

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):

$$\begin{aligned}
\hat{\beta}_{n,1} &= \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n) \right) \\
&= \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)(\beta_0 + X_i\beta_1 + Z_i\beta_2 + U_i - \beta_0 - \bar{X}_n\beta_1 - \bar{Z}_n\beta_2 - \bar{U}_n) \right) \\
&= \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)((X_i - \bar{X}_n)\beta_1 + (Z_i - \bar{Z}_n)\beta_2 + (U_i - \bar{U}_n)) \right) \\
&= \beta_1 + \beta_2 \frac{\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)(Z_i - \bar{Z}_n) \right)}{\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right)} + \frac{\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)(U_i - \bar{U}_n) \right)}{\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right)}
\end{aligned} \tag{16}$$

We know the last term in the last line converges in probability 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 is 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:

$$Var(\hat{\beta}_n|W) = (W'W)^{-1}W'Var(U|W)W(W'W)^{-1} \tag{17}$$

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:

$$\begin{aligned}
Var(\hat{\beta}_n^*|W) &= (W'\hat{\Omega}_n^{-1}W)^{-1}W'\hat{\Omega}_n^{-1}Var(U|W)\hat{\Omega}_n^{-1}W(W'\hat{\Omega}_n^{-1}W)^{-1} \\
&\xrightarrow{p} (W'\Omega^{-1}W)^{-1}W'\Omega^{-1}\Omega\Omega^{-1}W(W\Omega^{-1}W)^{-1} = (W\Omega^{-1}W)^{-1}
\end{aligned} \tag{18}$$

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.