Question 1

a

$$\mathcal{L}(\beta,\lambda) = \frac{1}{2n}\sum_{i=1}^{n}(Y_i - X_i'\beta)^2 + \lambda'(R\beta - c)$$

Taking gradients and applying the rules of matrix calculus:

$$\begin{aligned}
\nabla_\beta \mathcal{L} &= -\frac{1}{n}\sum_{i=1}^{n} X_i(Y_i - X_i'\beta) + R'\lambda = 0 \\
\nabla_\lambda \mathcal{L} &= R\beta - c = 0
\end{aligned}$$

Let $\tilde{\beta}_n, \tilde{\lambda}_n$ be such that this system is satisfied.

b

Expanding the condition on β:

$$\begin{aligned}
-\frac{1}{n}\sum_{i=1}^{n} X_iY_i + \frac{1}{n}\sum_{i=1}^{n} X_iX_i'\tilde{\beta}_n + R'\tilde{\lambda}_n = 0 \\
\frac{1}{n}\sum_{i=1}^{n} X_iY_i - R'\tilde{\lambda}_n = \frac{1}{n}\sum_{i=1}^{n} X_iX_i'\tilde{\beta}_n \\
\left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_iY_i\right) - \left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\right)^{-1} R'\tilde{\lambda}_n = \tilde{\beta}_n \\
\widehat{\beta}_n - \left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\right)^{-1} R'\tilde{\lambda}_n = \tilde{\beta}_n
\end{aligned}$$

Note that the inversion is possible because of the same line of reasoning that allowed for it during the calculation of the OLS estimator. The same assumptions and calculations are made.

Applying this formula for $\tilde{\beta}_n$ in the second condition:

$$\begin{aligned}
R\tilde{\beta}_n - c = 0 \\
R\widehat{\beta}_n - R\left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\right)^{-1} R'\tilde{\lambda}_n - c = 0 \\
\left(R\left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\right)^{-1} R'\right)^{-1}\left(R\widehat{\beta}_n - c\right) = \tilde{\lambda}_n
\end{aligned}$$

c

Since R, X, c are fixed, $\tilde{\lambda}_n$ is a continuous function of $\widehat{\beta}_n$, as matrix multiplication, addition and inversion are all continuous operations. It is well known that $\widehat{\beta}_n \xrightarrow{p} \beta$ by the Weak Law of Large Numbers and Continuous Mapping Theorem, so we know that:

$$\tilde{\lambda}_n \xrightarrow{p} \left(R \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} R' \right)^{-1} (R\beta - c) = 0$$

d

The Limiting distribution of $\widehat{\beta}_n$ is given by:

$$\sqrt{n} \left(\widehat{\beta}_n - \beta \right) \xrightarrow{d} \mathcal{N}(0, \Omega)$$
$$\Omega = \mathbb{E}\left[XX'\right]^{-1} \mathbb{V}\left(Xu\right) \mathbb{E}\left[XX'\right]^{-1}$$

Our function $\tilde{\lambda}_n$ is linear in $\widehat{\beta}_n$, and is thus continuous and differentiable. We can then apply the delta method to determine its distribution when $R\beta = c$.

$$\nabla_{\widehat{\beta}_n} \tilde{\lambda}_n = \left(R \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} R' \right)^{-1} R = \Sigma_R$$
$$\tilde{\lambda}_n(\beta) = 0$$

$$\sqrt{n} \tilde{\lambda}_n \xrightarrow{d} \mathcal{N}(0, \Sigma_R' \Omega \Sigma_R)$$

where Ω, Σ_R are given above.

e

Note that our above asymptotic distribution depends on the moments for the estimator of Ω. We may estimate it with the heteroskedasticity-robust estimators used in the notes:

$$\widehat{\Omega}_n = \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' (\widehat{u}_n i)^2 \right) \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1}$$

It is known that $\widehat{\Omega}_n \xrightarrow{p} \Omega$, so under the null hypothesis:

$$\sqrt{n} \left(\widehat{\beta}_n - \beta \right) \xrightarrow{d} \mathcal{N}(0, \widehat{\Omega}_n)$$
$$\sqrt{n} \tilde{\lambda}_n \xrightarrow{d} \mathcal{N}(0, \Sigma_R' \widehat{\Omega}_n \Sigma_R)$$
$$n \tilde{\lambda}_n \left(\Sigma_R' \widehat{\Omega}_n \Sigma_R \right)^{-1} \tilde{\lambda}_n' \xrightarrow{d} \chi_p^2$$

This depends completely on known quantities, and can therefore be tested.

$$n\tilde{\lambda}_n\left(\Sigma_R'\widehat{\Omega}_n\Sigma_R\right)^{-1}\tilde{\lambda}_n' = T_n$$

Our test will take the form of:

$$\mathbb{1}_{\{|T_n|>c_{p,1-\alpha}\}}$$

where $c_{p,1-\alpha}$ is the critical value for the α specified.

To determine if our test is consistent in level:

$$\begin{aligned}
\lim_{n\to\infty}\mathbb{E}\left[\mathbb{1}_{\{|T_n|>c_{p,1-\alpha}\}}\right] &= \lim_{n\to\infty}\Pr(|T_n|>c_{p,1-\alpha}) \\
&= \lim_{n\to\infty}\Pr\left(n\tilde{\lambda}_n\left(\Sigma_R'\widehat{\Omega}_n\Sigma_R\right)^{-1}\tilde{\lambda}_n' > c_{p,1-\alpha}\right) \\
&= \Pr\left(\chi_p^2 > c_{p,1-\alpha}\right) \\
&= \alpha
\end{aligned}$$

f

This test is very similar to the Wald-Test developed before, but the variance estimate is far more complicated. Computationally this test is far more complicated, involving three levels of inversion, two in computing Σ_R and once more in the actual test. For data sets with any reasonably high condition number this test will perform much more poorly than the Wald-Test simply on computation problems.

Both tests are in the limit the most powerful tests, so it is difficult to compare between Wald and LM tests without knowing the potential distributions we would encounter. One would expect that when the likelihood function has a very high peak, an LM test, which relies on comparing the slopes of the likelihood function to do better than a Wald Test which is examining the values of β.

Both tests rely on the asymptotic normality of $\widehat{\beta}_n$ and the converge of $\widehat{O}_n$, so neither makes more assumptions than the other.

Question 2

$$\begin{aligned}
\mathbb{V}\left(\hat{X}\right) &= \mathbb{V}\left(X\right)+\mathbb{V}\left(V\right)+Cov(X,V) \\
&= \mathbb{V}\left(X\right)+\mathbb{V}\left(V\right)+\mathbb{E}\left[XV\right]-\mathbb{E}\left[X\right]\mathbb{E}\left[V\right] \\
&= \mathbb{V}\left(X\right)+\mathbb{V}\left(V\right)
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}\left(Y\right) &= \mathbb{V}\left(\beta_0\right)+\mathbb{V}\left(\beta_1X\right)+\mathbb{V}\left(U\right)+Cov(\beta_0,U)+Cov(\beta_1X,U)+Cov(\beta_0,\beta_1X) \\
&= \beta_1^2\mathbb{V}\left(X\right)+\mathbb{V}\left(U\right)+\beta_1\mathbb{E}\left[XU\right]-\beta_1\mathbb{E}\left[X\right]\mathbb{E}\left[U\right] \\
&= \beta_1^2\mathbb{V}\left(X\right)+\mathbb{V}\left(U\right)
\end{aligned}$$

As β_0 is a constant, and $\mathbb{E}[U] = \mathbb{E}[XU] = 0$.

$$\begin{aligned}
Cov(\hat{X}, Y) &= Cov(X, Y) + Cov(V, Y) \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[VY] - \mathbb{E}[V]\mathbb{E}[Y] \\
&= \mathbb{E}\left[\beta_0 X + \beta_1 X^2 + UX\right] - \mathbb{E}[X](\beta_0 + \beta_1 \mathbb{E}[X]) + \mathbb{E}[\beta_0 V + \beta_1 XV + UV] - 0 \\
&= \beta_0 \mathbb{E}[X] + \beta_1 \mathbb{E}\left[X^2\right] - \beta_0 \mathbb{E}[X] - \beta_1 \mathbb{E}[X]^2 \\
&= \beta_1 \mathbb{V}(X)
\end{aligned}$$

b

Note that $\beta_1 = \frac{Cov(\hat{X},Y)}{\mathbb{V}(X)}$. Let $\beta_1 \geq 0$. If $\beta_1 = 0$, then $Cov(\hat{X}, Y) = 0$ and this result is true trivially. (Assuming that $\frac{c}{0} = \infty$ when $c > 0$). So let $\beta_1 > 0$. When β_1 is positive, $Cov(\hat{X}, Y)$ is positive as well by the final proof.

We know that $\mathbb{V}(X) \leq \mathbb{V}(X) + \mathbb{V}(V) = \mathbb{V}\left(\hat{X}\right)$, so clearly dividing by this larger number can only reduce β_1.

$$\frac{Cov(\hat{X}, Y)}{\mathbb{V}\left(\hat{X}\right)} \leq \beta_1$$

Combining the last two results we get that:

$$\begin{aligned}
\mathbb{V}(Y) &= \beta_1 Cov(\hat{X}, Y) + \mathbb{V}(U) \\
\beta_1 &= \frac{\mathbb{V}(Y) - \mathbb{V}(U)}{Cov(\hat{X}, Y)} \\
\beta_1 &\leq \frac{\mathbb{V}(Y)}{Cov(\hat{X}, Y)}
\end{aligned}$$

c

Result (b) holds when $\beta_1 = 0$, so let us examine the case when $\beta_1 < 0$. Note that since $Cov(\hat{X}, Y) = \beta_1 \mathbb{V}(X)$, it must be the case that $Cov(\hat{X}, Y) < 0$.

$$\beta_1 = \frac{Cov\hat{X}, Y}{\mathbb{V}(X)} \leq \frac{Cov\hat{X}, Y}{\mathbb{V}\left(\hat{X}\right)}$$

as the Covariance is now negative, so increasing the denominator increases the term.

Combining the last two results we get that:

$$\begin{aligned}
\mathbb{V}(Y) &= \beta_1 Cov(\hat{X}, Y) + \mathbb{V}(U) \\
\beta_1 &= \frac{\mathbb{V}(Y) - \mathbb{V}(U)}{Cov(\hat{X}, Y)} \\
\beta_1 &\geq \frac{\mathbb{V}(Y)}{Cov(\hat{X}, Y)}
\end{aligned}$$

Combining these two together we arrive at:

$$\frac{\mathbb{V}(Y)}{Cov(\hat{X}, Y)} \leq \beta_1 \leq \frac{Cov\hat{X}, Y}{\mathbb{V}\left(\hat{X}\right)}$$

Question 3

X, Z are both $k+1$ dimensional vectors, and rank $\mathbb{E}[ZX']$ is $k+1$. We shall approach by contra-positive, showing that perfect colinearity in Z implies that the rank of $\mathbb{E}[ZX']$ is not $k+1$.

Assume that there is perfect colinearity in Z. That is, $\exists c \neq 0$ such that $1 = \Pr(c'Z = 0)$.

$$\begin{aligned}
0 &= \mathbb{E}[c'Z] = \mathbb{E}[Z'c] \\
0 &= \mathbb{E}[\mathbb{E}[Z'c|X]] \\
X0 &= \mathbb{E}[\mathbb{E}[XZ'c|X]] \\
\mathbf{0} &= \mathbb{E}[XZ'c] \\
\mathbf{0} &= \mathbb{E}[XZ']\, c \\
\mathbf{0} &= \mathbb{E}[(ZX')']\, c \\
\mathbf{0} &= \mathbb{E}[ZX']'\, c
\end{aligned}$$

This tells us that the matrix $\mathbb{E}[ZX']'$ has a determinant that is zero, as there is a linear combination of its columns that can be made to equal zero. Note that it is reasonable to talk about determinants since ZX' is a square matrix. Transposing a matrix does not effect its determinants, so the determinant of $\mathbb{E}[ZX']$ is zero as well. This is equivalent to having a rank less than $k+1$.

Question 4

Note that $V = X - BLP(X|Z)$. Note that the $BLP(X|Z) = \Pi' Z$ where Π is a $\ell + 1xk + 1$ matrix.

Consider $BLP(U|V) = V'\gamma$, note that this has an implied first-order condition of orthogonality.

$$\mathbf{0} = \mathbb{E}[V(U - V'\gamma)] = \mathbb{E}\left[V\tilde{U}\right]$$

This is exactly exogeneity for V.

From the first-order conditions on the $BLP(X|Z)$ we know that for all j,

$$\begin{aligned}
\mathbb{E}[Z(X_j - \Pi' Z)] &= 0 \\
\mathbb{E}[ZV_j] &= 0 \\
\mathbb{E}[ZV'] &= 0
\end{aligned}$$

Where the last result accumulates all of the j vector conditions into one matrix condition.

We may also note that:

$$
\begin{aligned}
\mathbb{E}\left[X\tilde{U}\right] &= \mathbb{E}\left[V\tilde{U} + BLP(X|Z)\tilde{U}\right] = \mathbb{E}\left[BLP(X|Z)\tilde{U}\right] \\
&= \mathbb{E}\left[\Pi' Z\left(U - V'\gamma\right)\right] \\
&= \Pi'\mathbb{E}\left[ZU\right] - \Pi'\mathbb{E}\left[ZV'\right]\gamma \\
&= \Pi'\mathbf{0} - \Pi'\mathbf{0}\gamma = \mathbf{0}
\end{aligned}
$$

b

Using the sub-vector results from linear regression before, we may solve for β as:

Define $\tilde{Y} = Y - BLP(Y|V)$ and $\tilde{X} = X - BLP(X|V) = BLP(X|Z) = \Pi' Z$.

The last equality follows from the interpretation of best linear predictors as a linear conditional expectation.

This occurs because V contains no more information than Z, so predicting X on V is the same as predicting it on Z (after handling the subtracting elements.)

Then

$$
\begin{aligned}
\beta &= \mathbb{E}\left[\tilde{X}\tilde{X}'\right]^{-1}\mathbb{E}\left[\tilde{X}\tilde{Y}\right] \\
&= \mathbb{E}\left[\tilde{X}\tilde{X}'\right]^{-1}\mathbb{E}\left[\tilde{X}Y\right] \\
&= \mathbb{E}\left[\Pi' Z Z' \Pi'\right]^{-1}\mathbb{E}\left[\Pi' Z Y\right] \\
&= \Pi'\mathbb{E}\left[ZZ'\right]\Pi^{-1}\Pi'\mathbb{E}\left[ZY\right]
\end{aligned}
$$

This is exactly our instrumental variables estimator that we derived in class.

Question 5

Let $BLP(Y|Z) = Z'\lambda$ and $\Gamma' Z = BLP(X|Z)$

$$
Y = Z'\lambda + \epsilon \qquad X = \Gamma' Z + \eta
$$

Show that $\lambda = \Gamma\beta$ and $\epsilon = \eta'\beta + U$.

$$
\begin{aligned}
Y &= X'\beta + U \\
Z'\lambda + \epsilon &= Z'\Gamma\beta + \eta'\beta + U
\end{aligned}
$$

This must be true for any arbitrary Z satisfying the above constraints, so it must be that the coefficients for the Z are equal, and the coefficients for the terms without the Z are also equal.

$$
\begin{aligned}
\lambda &= \Gamma\beta \\
\epsilon &= \eta'\beta + U
\end{aligned}
$$

b

We know that Γ is an $\ell + 1$ by $k + 1$ matrix, and that η is $k + 1$ vector.

Replace X with its best linear predictor given above. Then show that exogeneity holds.