Question 1

a

Consider the lagrangian below:

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

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; that is, because there is no perfect colinearity in X and $\left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\right)^{-1} \xrightarrow{p} \mathbb{E}\left[XX'\right]^{-1}$ due to WLLN, the matrix of samples must have an inverse with probability approaching one.

Substituting the expression above for $\tilde{\beta}_n$ in the second condition:

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

c

$\tilde{\lambda}_n$ is a continuous function of $\widehat{\beta}_n$ and $\left(\frac{1}{n}\sum_{i=1}^n X_iX_i'\right)^{-1}$, as matrix multiplication, addition and inversion are all continuous operations. Because $\widehat{\beta}_n \xrightarrow{p} \beta$ and $\left(\frac{1}{n}\sum_{i=1}^n X_iX_i'\right)^{-1} \xrightarrow{p} \mathbb{E}\left[XX'\right]^{-1}$ (by the Weak Law of Large Numbers and Continuous Mapping Theorem), and marginal convergence in probability implies joint convergence, we can aply the CMT to conclude:

$$\tilde{\lambda}_n \xrightarrow{p} \left(R\mathbb{E}\left[XX'\right]^{-1}R'\right)^{-1}(R\beta - c) = 0,$$

when $(R\beta - c) = 0$.

d

As seen in class, the Limiting distribution of $\widehat{\beta}_n$ is given by:

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

Notice then that we can apply CMT to obtain:

$$\sqrt{n}\left(R\widehat{\beta}_n - c - R\beta + c\right) = \sqrt{n}\left(R\widehat{\beta}_n - R\beta\right) \xrightarrow{d} \mathcal{N}(0, R\Omega R')$$

And, when $(R\beta - c) = 0$, this implies: $\sqrt{n}\left(R\widehat{\beta}_n - c\right) \xrightarrow{d} \mathcal{N}(0, R\Omega R')$.

Now, because $\left(\frac{1}{n}\sum_{i=1}^n X_iX_i'\right)^{-1} \xrightarrow{p} \mathbb{E}\left[XX'\right]^{-1}$, and using CMT and slutsky, we can conclude that:

$$\begin{aligned}\sqrt{n}\tilde{\lambda}_n &= \left(R\left(\frac{1}{n}\sum_{i=1}^n X_iX_i'\right)^{-1}R'\right)^{-1}\sqrt{n}\left(R\widehat{\beta}_n - c\right)\\ &\xrightarrow{d} \left(R\mathbb{E}\left[XX'\right]^{-1}R'\right)^{-1}\mathcal{N}(0, R\Omega R') = \mathcal{N}(0, AR\Omega R'A'),\end{aligned}$$

where $A := \left(R\mathbb{E}\left[XX'\right]^{-1}R'\right)^{-1}$.

e

To construct a feasible test, we need to estimate the covariance matrix $\Lambda := AR\Omega R'A'$ in the asymptotic distribution above. First, we may estimate Ω with the heteroskedasticity-robust estimators used in class:

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

We saw in class that $\widehat{\Omega}_n \xrightarrow{p} \Omega$. Also, we saw above that $\left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1} \xrightarrow{p} \mathbb{E}\left[XX'\right]^{-1}$. Thus, by CMT, we have that:

$$\widehat{\Lambda}_n := \left(R\left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1} R'\right)^{-1} R\widehat{\Omega}_n R' \left(R\left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1} R'\right)^{-1} \xrightarrow{p} \Lambda$$

So, $\widehat{\Lambda}_n$ is a consistent estimator of the covariance matrix of $\sqrt{n}\tilde{\lambda}_n$, under the null hypothesis. Therefore, under $(R\beta = c)$, we have, by slutsky and the fact that Λ is invertible (it is a multiplication of invertible matrices):

$$\begin{aligned}\widehat{\Lambda}_n^{-\frac{1}{2}}\sqrt{n}\left(\tilde{\lambda}_n\right) \xrightarrow{d} \mathcal{N}(0,1) \\ n\tilde{\lambda}_n'\widehat{\Lambda}_n^{-1}\tilde{\lambda}_n \xrightarrow{d} \chi_p^2\end{aligned}$$

since we have p restrictions.

This depends completely on known quantities, and can therefore be tested. We define the test statistic as:

$$T_n := n\tilde{\lambda}_n'\widehat{\Lambda}_n^{-1}\tilde{\lambda}_n$$

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 in the chi-square distribution with p degrees of freedom.

And, as shown below, using portmanteau lemma, our test is consistent in level:

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

f

Both tests are actually equivalent. We can see that by substituting the expressions for $\tilde{\lambda}_n$ and $\widehat{\Lambda}_n^{-1}$ into the test statistic used above:

$$T_n = n\left\{\left(R\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right)^{-1}R'\right)^{-1}\left(R\widehat{\beta}_n - c\right)\right\}'$$
$$\left\{\left(R\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right)^{-1}R'\right)^{-1}R\widehat{\Omega}_nR'\left(R\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right)^{-1}R'\right)^{-1}\right\}^{-1}$$
$$\left\{\left(R\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right)^{-1}R'\right)^{-1}\left(R\widehat{\beta}_n - c\right)\right\}$$

$$T_n = n\left(R\widehat{\beta}_n - c\right)'\left(R\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right)^{-1}R'\right)^{-1}$$
$$\left(R\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right)^{-1}R'\right)\left(R\widehat{\Omega}_nR'\right)^{-1}\left(R\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right)^{-1}R'\right)$$
$$\left(R\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right)^{-1}R'\right)^{-1}\left(R\widehat{\beta}_n - c\right)$$

$$T_n = n\left(R\widehat{\beta}_n - c\right)'\left(R\widehat{\Omega}_nR'\right)^{-1}\left(R\widehat{\beta}_n - c\right)$$

Thus, we can see that both test have the same test statistic and the same asymptotic distribution, thus they are equivalent.

Question 2

a

$$\begin{aligned}\mathbb{V}\left(\hat{X}\right) &= \mathbb{V}\left(X\right) + \mathbb{V}\left(V\right) + Cov(X, V)\\ &= \mathbb{V}\left(X\right) + \mathbb{V}\left(V\right) + 2(\mathbb{E}\left[XV\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[V\right])\\ &= \mathbb{V}\left(X\right) + \mathbb{V}\left(V\right)\end{aligned}$$

As $\mathbb{E}\left[V\right] = \mathbb{E}\left[XV\right] = 0$.

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

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

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

As $\mathbb{E}[V] = \mathbb{E}[XV] = \mathbb{E}[U] = \mathbb{E}[XU] = 0$.

b

First let's show that $\frac{Cov(\hat{X},Y)}{\mathbb{V}(\hat{X})} \leq \beta_1$. Let $\beta_1 \geq 0$. If $\beta_1 = 0$, then $Cov(\hat{X}, Y) = 0$ and this result is true trivially. So let $\beta_1 > 0$. Then we get that:

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

since $\mathbb{V}(X) \leq \mathbb{V}(X) + \mathbb{V}(V)$ (otherwise there would be no difference between $X and \hat{X}$) and $\beta_1 > 0$.

Now, we show that $\beta_1 \leq \frac{\mathbb{V}(Y)}{Cov(\hat{X},Y)}$. If $\beta_1 = 0$, then $Cov(\hat{X}, Y) = 0$ and the division is not well-defined (although we may define $\frac{c}{0} := \infty$, for $c > 0$, and then we have the inequality). Consider $\beta_1 > 0$. Then we have:

$$
\frac{\mathbb{V}(Y)}{Cov(\hat{X}, Y)} = \frac{\beta_1^2 \mathbb{V}(X) + \mathbb{V}(U)}{\beta_1 \mathbb{V}(X)} = \beta_1 + \frac{\mathbb{V}(U)}{\beta_1 \mathbb{V}(X)} \geq \beta_1
$$

Because $\beta_1 > 0$ and the variances are positive.

We know that $\frac{Cov(\hat{X},Y)}{\mathbb{V}(\hat{X})}$ is the coefficient from the regression $Y = \beta_0^* + \beta_1^* \hat{X} + U^*$. Thus we can interpret the lower bound as a proof that the coefficient on the "wrong" regression, in this simple case, will be always weakly smaller than the coefficient on the "true" regression. Similarly, the upper bound is the inverse of the coefficient of the reverse regression $\hat{X} = \alpha_0 + \alpha_1 Y + z$, and thus we can interpret it as stating the the true coefficient will be always smaller than the inverse of the coefficient of the reverse regression involving the $\hat{X}$.

c

We already saw that the result holds when $\beta_1 = 0$ (letter b above), so let us examine the case when $\beta_1 < 0$. We can do this similarly to letter b above:

First the upper bound:

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

because now $\beta_1 < 0$, and multiplying it by a number smaller than one decreases its absolute value and makes it bigger, since it is negative.

Then the lower bound:

$$\frac{\mathbb{V}(Y)}{Cov(\hat{X}, Y)} = \frac{\beta_1^2 \mathbb{V}(X) + \mathbb{V}(U)}{\beta_1 \mathbb{V}(X)} = \beta_1 + \frac{\mathbb{V}(U)}{\beta_1 \mathbb{V}(X)} \leq \beta_1$$

Because $\beta_1 < 0$ and the variances are positive, so adding something negative to a negative number decreases the resut.

Question 3

X, Z are both $k+1$ dimensional vectors (thus $\mathbb{E}[ZX']$ is a square matrix with dimension $k+1$), and rank $\mathbb{E}[ZX']$ is $k+1$. Approach by contrapositive, Let us assume that there is perfect colinearity in Z and show that this implies rank $\mathbb{E}[ZX'] < k+1$ (it cannot have rank higher than its dimension). That is, assume $\exists c \neq 0$ such that $1 = \Pr(c'Z = 0)$. Thus using this c, we have:

$$c'\mathbb{E}[ZX'] = \mathbb{E}[(c'Z)X'] = 0$$

But this implies that there exists $c \neq 0$ such that $c'\mathbb{E}[ZX'] = 0$. This means that the rows of $\mathbb{E}[ZX']$ are not linearly independent. As $\mathbb{E}[ZX']$ is a square matrix, a row being linearly dependent implies that the determinant of the matrix is zero. Therefore the transpose of this matrix, $\mathbb{E}[ZX']'$ has a determinant of zero, and linearly dependent rows. Linear dependence in the rows of this matrix is equivalent to linear dependence in the columns of $\mathbb{E}[ZX']$. This means the matrix does not have full rank.

Therefore having rank $k+1$ in $\mathbb{E}[ZX']$ implies no perfect colinearity in Z.

Question 4

We are given $V = X - BLP(X|Z)$. Let us define $BLP(X|Z) = \Pi' Z$ where Π is a $\ell + 1$ by $k+1$ 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,

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

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

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.) Best Linear Predictors, under interpretation one are a linear conditional expectation. Applying the law of iterated expectations:

$$\mathbb{E}\left[\mathbb{E}\left[X|V\right]|Z\right] = \mathbb{E}\left[X|Z\right]$$

as Z has less information than V. So the best linear predictor of X given V is $X - \Pi' Z$.
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

a

Since $Z'\lambda = BLP(Y|Z)$, we can use the expression for the best linear predictor and: (i) that $\mathbb{E}\left[ZZ'\right]$ exists and is invertible, since there is no perfect colinearity in Z, and (ii) $\mathbb{E}\left[ZY\right] =$

$\mathbb{E}[ZX']\beta + \mathbb{E}[ZU]$ exists (since both elements in the summation exist by assumption), to get:

$$\begin{aligned} \lambda &= \mathbb{E}[ZZ']^{-1}\mathbb{E}[ZY] \\ &= \mathbb{E}[ZZ']^{-1}\mathbb{E}[Z(X'\beta + U)] \\ \text{(Because } \mathbb{E}[ZU] = 0) \quad &= \mathbb{E}[ZZ']^{-1}\mathbb{E}[ZX']\beta \end{aligned}$$

Now we also have that $\Gamma' Z = BLP(X|Z)$, and, using the same facts stated in the first paragraph above, we can also get the expression for $\Gamma = \mathbb{E}[ZZ']^{-1}\mathbb{E}[ZX']$. Therefore, using the result above, we get: $\lambda = \Gamma\beta$ as we wanted.

Now we can substitute this into the reduced form equation for Y to get:

$$\begin{aligned} \epsilon &= Y - Z'\lambda \\ &= X'\beta + U - Z'\Gamma\beta \\ &= (X' - Z'\Gamma)\beta + U \\ \text{(Because } \eta = X - \Gamma' Z) \quad &= \eta'\beta + U, \end{aligned}$$

which is the result we wanted.

b

To get an expression for β, we can substitute for λ in the reduced form for Y to get:

$$Y = (\Gamma' Z)'\beta + \epsilon$$

Now we can check that: $\mathbb{E}[\Gamma' Z\epsilon] = 0$

$$\begin{aligned} \mathbb{E}[\Gamma' Z\epsilon] &= \mathbb{E}[\Gamma' Z(\eta'\beta + U)] \\ &= \Gamma'\mathbb{E}[Z\eta']\beta + \Gamma'\mathbb{E}[ZU] = 0, \end{aligned}$$

because $\eta = X - BLP(X|Z)$, so the first order condition for the best linear predictor gives us that $\mathbb{E}[Z\eta'] = 0$, and also we have by $\mathbb{E}[ZU]$ assumption.

Thus, since $\mathbb{E}[\Gamma' Z\epsilon] = 0$, we can obtain β as $BLP(Y|\Gamma' Z)$, which results in: $\beta = \mathbb{E}[\Gamma' ZZ'\Gamma]^{-1}\mathbb{E}[\Gamma' ZY] = \Gamma'\mathbb{E}[ZZ']\Gamma^{-1}\Gamma'\mathbb{E}[ZY]$. Notice that this expression is exactly the one derived in class, except that in class we called $\Gamma' Z = BLP(X|Z)$ as $\Pi' Z$.