1 Replicating Alsan (2015, AER) (35 points)

This exercise is based on the paper "The Effect of the TseTse Fly on African Development" by Marcella Alsan, which was published in the American Economic Review as **?**.

The abstract says: "The TseTse fly is unique to Africa and transmits a parasite harmful to humans and lethal to livestock. This paper tests the hypothesis that the TseTse reduced the ability of Africans to generate an agricultural surplus historically. Ethnic groups inhabiting TseTse-suitable areas were less likely to use domesticated animals and the plow, less likely to be politically centralized, and had a lower population density. These correlations are not found in the tropics outside of Africa, where the fly does not exist. The evidence suggests current economic performance is affected by the TseTse through the channel of precolonial political centralization."

The idea that "deep" historical roots have lasting impacts on current variation in economic development has become very popular in recent years in economics. This is a local historical approach to explaining why some parts from the world are rich and others are poor, and it complements "general" theories of economic growth, such as the one that was empirically examined in the previous problem set.

The data concerns several outcomes for different ethnic groups in Africa that were collected by anthropologists during the 19th century and first half of the 20th century. We are interested in the effect of the TseTse on several variables. In particular, in the file `precolonial.csv`, we have the variables listed in Table 1. The unit of observation, i, corresponds to an ethnic group. Please read the paper for more details.

Table 1: Variable definitions

Variable name	Variable description
Outcome variables (y_i)	
animals	large domesticated animals
intensive	intensive agriculture
plow	plow use
female_ag	female participation in agriculture
ln_popd_murdock	log population density
slavery_indigenous	indigenous slavery
central	political centralization
Main explanatory variable ($x_{2,i}$)	
tse	TseTse Suitability Index (TSI)
Climate controls ($x_{3,i}, x_{4,i}, x_{5,i}, x_{6,i}$)	
meanrh	humidity
meantemp	temperature
itx	interaction between humidity and temperature
prop_tropics	proportion of land area in tropics
Malaria control ($x_{7,i}$)	
malaria_index	malaria ecology index
Waterway controls ($x_{8,i}, x_{9,i}$)	
river	whether a river was located within the ethnic group boundaries
coast	whether a coast was included in the ethnic group boundaries
Geography controls ($x_{10,i}, x_{11,i}, x_{12,i}, x_{13,i}$)	
lon	longitude of the ethnic group
abslat	absolute latitude of the ethnic group
meanalt	mean altitude of the ethnic group
SI	FAO Agricultural Sustainability Index

Note: The data are available in the replication package of **?** at https://www.openicpsr.org/openicpsr/project/112921/version/V1/view. For details, see **?**.

In all exercises, conduct hypothesis tests at the 5% significance level, and report all results rounded to three decimal places. As you will see, there are missing values, as is common in many studies. Drop these observations on a regression-by-regression basis (so drop them if any only if there are missing values in the specific regression you are estimating). The standard errors you are asked to report will not match exactly the standard errors reported in the paper, but that is not a problem (**?** used clustered robust standard errors).

1.1 Linear regression without controls (7 points)

Consider the linear regression model

$$y_i = \beta_1 + \beta_2 x_{2,i} + \varepsilon_i \,. \tag{1}$$

1. (7 points) Using OLS, estimate Equation (1) separately for all *seven* outcome variables described above. Present your results in a table. Report b_2, s^2 and the standard error $\mathrm{SE}(b_2)$ for each regression. Test $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$, and report the t-statistic, p-value and your decision for each regression.

Solution:

Table 2: Statistical Results

Variable	b_2	s^2	SE_{b_2}	T_stat	P_Value	Decision
animals	-0.222	0.000	0.020	-11.100	0.000	Reject H_0
intensive	-0.206	0.000	0.019	-10.842	0.000	Reject H_0
plow	-0.096	0.000	0.011	-8.727	0.000	Reject H_0
female_ag	0.126	0.001	0.027	4.667	0.000	Reject H_0
ln_popd_murdock	0.078	0.007	0.083	0.940	0.348	Fail to Reject H_0
slavery_indigenous	0.104	0.000	0.017	6.118	0.000	Reject H_0
central	-0.116	0.000	0.021	-5.524	0.000	Reject H_0

1.2 Linear regression with controls (14 points)

Consider the linear regression model (equation 1 on p. 391 in **?**):

$$y_i = \beta_1 + \beta_2 x_{2,i} + \sum_{j=3}^{13} \beta_j x_{j,i} + \varepsilon_i \,, \tag{2}$$

i.e., now you are adding the control variables, replicating Column (4) of Table I on p. 394 in **?**.

1. (7 points) Using OLS, estimate Equation (2) separately for all *seven* outcome variables described above. Present your results in a table. Report b_2, s^2 and the standard error $\mathrm{SE}(b_2)$ for each regression. Test $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$, and report the t-statistic, p-value and your decision for each regression.

Solution:

Table 3: Statistical Results

Variable	b_2	s^2	SE_{b_2}	T_Stat	P_Value	Decision
animals	-0.230	0.002	0.039	-5.897	0.000	Reject H0
intensive	-0.090	0.001	0.038	-2.368	0.018	Reject H0
plow	-0.057	0.000	0.019	-3.000	0.003	Reject H0
female_ag	0.204	0.003	0.051	4.000	0.000	Reject H0
ln_popd_murdock	-0.747	0.022	0.148	-5.047	0.000	Reject H0
slavery_indigenous	0.103	0.001	0.033	3.121	0.002	Reject H0
central	-0.079	0.002	0.043	-1.837	0.067	Fail to Reject H0

2. (2 points) Discuss the differences between b_2 for the models with and without control variables.

Solution: A model with control variables provides a more robust foundation for understanding the relationship between the main explanatory variable and the outcome variable. With control variables, the interpretation of b_2 is clearer and focuses solely on the effect of tse while holding other variables constant. Including control variables helps mitigate omitted variable bias by isolating the effect of tse from the influences of other relevant factors, as in the case of the model without control variables part of the influence attributed to b_2 is actually explained by other variables.

When comparing b_2 for both models the results vary. For example, the variable of the log of the population density switches from a positive value, meaning that a higher TseTse suitable Index is associated with a higher level of population density, a strange result; to a negative value, meaning a higher TseTse is associated with a lower population density as expected since TseTse is "a parasite harmful to humans and lethal to livestock" which should then reduce population density. Assuming this reasoning is correct, that the presence of TseTse should impact population density negatively, this is a practical example of correcting an omitted variable bias.

3. (1 point) Discuss the differences between s^2 for the models with and without control variables.

Solution:

When comparing s^2 for both models we can observe that in general s^2 is higher for the simpler model without control variables. In this case, as more variables are added, although the omitted variable bias seems to decrease, the likelihood of including irrelevant variables increases. Theoretically, as more irrelevant variables are included there is an efficiency loss and consequently, the variance of the estimators increases. Then, what is observed in the data aligns with what should be expected from econometric literature.

4. (4 points) For each category of control variables (Climate, Malaria, *etc.*), explain why they need to be included in the regression. What is the consequence of not including them?

Solution: By including control variables, the analysis aims to isolate the effect of the tse-tse fly on these outcomes, ensuring that any observed relationship is not confounded by other factors.

Climate controls (meanrh, meantemp, itx, prop_tropics) are included to account for environmental factors that could influence both the presence of the tse-tse fly and the dependent variable (such as health outcomes). For instance, variations in humidity and temperature can affect the habitat of the tse-tse fly as well as impact agricultural productivity and health conditions.

Including malaria-related controls (malaria_index) helps to address the potential overlap between malaria and tse-tse fly infections, as both are significant health concerns in many African regions. Ignoring malaria prevalence could bias the estimated effect of the tse-tse fly if those areas are more susceptible to both diseases, as the simple estimator associated with the tse-tse fly would also indirectly express the impact of malaria in the dependent variables.

The same applies for waterway controls; Alsan accounts for the waterways because rivers and coasts might influence the productivity conditions of a certain area, as well as the propensity of increasing the rate of tse-tse fly propagation.
Finally, if geographic controls are not included in the analysis, not considering the impacts of certain altitudes, latitudes, and longitudes—such as tropical areas or extremely humid areas that could be fixed habitats for the tse-tse fly—could bias the results.

However, this decrease in the bias of the results is counterbalanced by the increase in the standard error associated with the estimates as irrelevant variables are possibly included, leading to efficiency losses.

Given this, the consequence of not including the control variables is that the estimators will more likely be biased, but their standard error will be lower, as some control variables seem to be relevant, decreasing the bias; while others seem to be irrelevant, decreasing the efficiency of the model.

1.3 The differential effect of the TseTse Suitability Index in tropical Africa (14 points)

A concern is that the TseTse Suitability Index is identifying a generic relationship between climate and agriculture. That means that in areas where the TseTse fly is present, the climate is such that agriculture is less possible. To find out whether this is the case, we use the fact that the TseTse fly only exists in Africa.

In particular, we add several other ethnic groups from tropical areas outside Africa to the dataset. This new dataset is in `placebo.csv`. Note that the log population density (ln_popd_murdock) variable is only produced for Africa, hence it cannot enter the analysis. Also, the proportion of land area in the tropics has to be omitted as a control, since it is unity in this sample. The key new variable is a dummy variable, "africa" in the dataset: $A_i = 1$ if the ethnic group is located in Africa and zero otherwise.

Consider the regression (equation 2 on p. 399 in **?**):

$$y_i = \beta_1 + \beta_2 x_{2,i} + \gamma_1 A_i + \gamma_2 (A_i x_{2,i}) + \sum_{\substack{j=3 \\ j \neq 6}}^{13} \beta_j x_{j,i} + \sum_{\substack{j=3 \\ j \neq 6}}^{13} \gamma_j (A_i x_{j,i}) + \varepsilon_i \,, \tag{3}$$

replicating Table 5 on p. 402 in **?**. Note that the interactions with A_i have already been added to the dataset with the "africa_" prefix.

1. (6 points) Using OLS, estimate the regression for all *six* outcome variables, and present the results for β_2 and γ_2 in a table with parameter estimates, standard errors, t-statistics ($H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$, and $H_0 : \gamma_2 = 0$ vs. $H_1 : \gamma_2 \neq 0$), accompanying p-values and your decisions.

Solution:

Table 4: Testing $\beta_2 = 0$

Metric	b_2	SE_{b_2}	T-Stat	P-Value	Decision
Animals	0.036	0.050	0.729	0.466	Fail to Reject H0
Intensive	-0.015	0.048	-0.311	0.756	Fail to Reject H0
Plow	0.069	0.028	2.440	0.015	Reject H0
Female_ag	-0.039	0.056	-0.693	0.489	Fail to Reject H0
Slavery_Indigenous	-0.003	0.046	-0.062	0.951	Fail to Reject H0
Centralization	0.010	0.053	0.183	0.855	Fail to Reject H0

Table 5: Testing $\gamma_2 = 0$

Metric	$\hat{\gamma}_2$	$SE_{\hat{\gamma}_2}$	T-Stat	P-Value	Decision
Animals	-0.214	0.067	-3.206	0.001	Reject H0
Intensive	-0.075	0.064	-1.161	0.246	Fail to Reject H0
Plow	-0.070	0.038	-1.841	0.066	Fail to Reject H0
Female_ag	0.247	0.081	3.049	0.002	Reject H0
Slavery_Indigenous	0.105	0.062	1.698	0.090	Fail to Reject H0
Centralization	-0.116	0.070	-1.646	0.100	Fail to Reject H0

2. (8 points) Interpret β_2 and γ_2, and their sum $\beta_2 + \gamma_2$. In each of the *six* regressions, test the null hypothesis that $\beta_2 + \gamma_2 = 0$ against the alternative that $\beta_2 + \gamma_2 \neq 0$. In a table, report the F-statistics, p-values and your decisions. Do these results support the claim that the TseTse fly had an effect on African development?

Solution:

Table 6: Testing $\beta_2 + \gamma_2 = 0$

Variable	$b_2 + \hat{\gamma}_2$	SE	T-stat	P-Value	F-stat	Decision
Animals	-0.177	0.083	-2.133	0.033	4.548	Reject H0
Intensive	-0.090	0.080	-1.117	0.264	1.248	Fail to Reject H0
Plow	-0.000	0.047	-0.015	0.988	0.000	Fail to Reject H0
Female_ag	0.208	0.099	2.113	0.035	4.463	Reject H0
Slavery_Indigenous	0.103	0.077	1.328	0.184	1.763	Fail to Reject H0
Centralization	-0.106	0.088	-1.209	0.227	1.462	Fail to Reject H0

Since we are testing that tse in Africa have non significance, we get the we reject the null hypothesis in the animals and female_ag regressions. In other words, it implies that the combined effect of the tse and its interaction with the Africa dummy variable on the dependent variables is statistically significant. This suggests that the suitability of the tse-tse fly has a meaningful impact on the dependent variable (animals and female_ag), and that this impact varies based on whether the ethnic group is located in Africa.

However, in accordance to the results, we don't have sufficient evidence to reject the null hypothesis, hence failing to reject that there is any statistical significance between the tse in Africa through the (intensive, plow, slavery_indigenous, centralization) hypothesis testing that $\beta 2 = \gamma 2$.

2 Maximum Likelihood (15 points)

2.1 Cramer–Rao lower bound: Bernoulli (12 points)

Consider Example 2 from class: we observe an *iid*. sample of Bernoulli random variables: $\boldsymbol{y} = (y_1, \ldots, y_n)'$, with $y_i \overset{iid}{\sim} \text{Ber}(p)$. The Maximum Likelihood estimator is $\check{p} = \frac{\sum_{i=1}^{n} y_i}{n}$, as we saw.

1. (6 points) Derive the Cramer–Rao lower bound.

Solution:

From class, we know that for an i.i.d. sample of $y_i \sim Ber(p)$, the log-likelihood function is the following:

$$\log L(\widetilde{p}) = \left(\sum_i y_i\right) \log(\widetilde{p}) + \left(n - \sum_i y_i\right) \log(1 - \widetilde{p})$$

The score function is the following:

$$\begin{aligned} s(\widetilde{p}) = \frac{\partial \log L(\widetilde{p})}{\partial \widetilde{p}} &= \frac{\sum_i y_i}{\widetilde{p}} - \frac{n - \sum_i y_i}{1 - \widetilde{p}} \\ &= \frac{(1-\widetilde{p}) \sum_i y_i - \widetilde{p}(n - \sum_i y_i)}{\widetilde{p}(1-\widetilde{p})} = \frac{\sum_i y_i - n\widetilde{p}}{\widetilde{p}(1-\widetilde{p})} \end{aligned}$$

Then the Cramer-Rao lower bound is $I(p)^{-1}$:

$$\begin{aligned} I(p) &= \mathbb{E}\left[\left(\frac{\sum_i y_i - n \cdot p}{p(1-p)}\right)^2\right] = \\ &= \frac{1}{p^2(1-p)^2} \mathbb{E}\left[\left(\sum_i y_i\right)^2\right] - \frac{2np}{p^2(1-p)^2} \mathbb{E}\left[\sum_i y_i\right] + \frac{n^2 p^2}{p^2(1-p)^2} \end{aligned}$$

We know that $\mathbb{E}\left[\sum_i y_i\right] = n \cdot p$ for a binomial distribution and $\mathbb{E}\left[\left(\sum_i y_i\right)^2\right] = np(1-p) + n^2p^2$

$$I(p) = \frac{np(1-p)}{p^2(1-p)^2} + \frac{n^2p^2}{p^2(1-p)^2} - \frac{2n^2p^2}{p^2(1-p)^2} + \frac{n^2p^2}{p^2(1-p)^2} = \frac{n}{p(1-p)}$$

Thus,

$$I(p)^{-1} = \frac{p(1-p)}{n}$$

2. (6 points) Derive the expected value and the variance of the Maximum Likelihood estimator. Is the estimator unbiased? Does the variance achieve the Cramer–Rao lower bound?

Solution:

The ML estimator is $\check{p} = \frac{\sum_i y_i}{n}$.

$$\mathbb{E}[\check{p}] = \mathbb{E}\left[\frac{\sum_i y_i}{n}\right] = \frac{1}{n}\mathbb{E}\left[\sum_i y_i\right] = \frac{1}{n} \cdot n \cdot p = p$$

The estimator is unbiased.

$$Var(\check{p}) = \mathbb{E}\left[(\check{p} - \mathbb{E}(\check{p}))^2\right] = \mathbb{E}[\check{p}^2] - 2\mathbb{E}(\check{p})\mathbb{E}(\check{p}) + \mathbb{E}(\check{p})^2 =$$

$$= \mathbb{E}[\check{p}^2] - (\mathbb{E}(\check{p}))^2 = \frac{1}{n^2}\mathbb{E}\left[\left(\sum_i y_i\right)^2\right] - \frac{1}{n^2}\mathbb{E}\left[\sum_i y_i\right]^2$$
$$= \frac{1}{n^2}\left(np(1-p) + n^2p^2\right) - \frac{1}{n^2}(np)^2$$
$$= \frac{p(1-p)}{n} + p^2 - p^2 = \frac{p(1-p)}{n}$$

The variance achieves the C-R lower bound.

2.2 Cramer–Rao lower bound: s^2 (3 points)

Recall that $s^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-K}$ is an unbiased estimator of σ^2. In class, we derived the Cramer–Rao lower bound under Assumptions 1–5: $\frac{2(\sigma^2)^2}{n}$.

1. (3 points) Does s^2 achieve the Cramer–Rao lower bound? *Hint*: What is $\mathbb{V}\text{ar}(s^2|\boldsymbol{X})$ under Assumptions 1–5? Recall that $(q|\boldsymbol{X}) \sim \chi^2_{n-K}$, where $q = \frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}$, and the variance of a random variable following the chi-squared distribution with m degrees of freedom is $2 \cdot m$.

Solution:

$$Var(s^2|X) = Var\left(\frac{e'e}{n-k}\middle| X\right) = Var\left(\frac{\sigma^2}{n-k}q\middle| X\right)$$
$$= \frac{(\sigma^2)^2}{(n-k)^2}Var(q|X) = \frac{(\sigma^2)^2}{(n-k)^2} \cdot 2 \cdot (n-k) = \frac{2(\sigma^2)^2}{(n-k)}$$

Which is clearly greater than $\frac{2(\sigma^2)^2}{n}$ due to the fact that $n > n-k$.

So s^2 does not achieve the Cramér-Rao lower bound.

3 Restricted least squares (5 bonus points)

When estimating a classical linear regression model with the normality assumption using restricted least squares, the sum of squared residuals is minimized subject to the constraint implied by the null hypothesis

$\boldsymbol{R\beta} = \boldsymbol{r}$. The objective function is the Lagrangian, which is given by

$$\mathcal{L}(\widetilde{\boldsymbol{\beta}}) = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}) + \boldsymbol{\lambda}'(\boldsymbol{R}\widetilde{\boldsymbol{\beta}} - \boldsymbol{r}), \tag{4}$$

where $\boldsymbol{\lambda}$ is the $\#r$-dimensional vector of Lagrange multipliers (recall: $\boldsymbol{R}$ is $(\#r \times K)$, $\widetilde{\boldsymbol{\beta}}$ is $(K \times 1)$, and $\boldsymbol{r}$ is $(\#r \times 1)$). Let $\widehat{\boldsymbol{\beta}}$ be the restricted least squares estimator of $\boldsymbol{\beta}$. It is the solution to the constrained minimization problem.

1. (2 points) Let $\boldsymbol{b}$ be the unrestricted OLS estimator. Show the following:

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{b} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{Rb} - \boldsymbol{r}), \tag{5}$$

$$\boldsymbol{\lambda} = [\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{Rb} - \boldsymbol{r}). \tag{6}$$

Hint: Obtain the FOC by differentiating the Lagrangian with respect to $\widetilde{\boldsymbol{\beta}}$ and setting it equal to $\boldsymbol{0}$. Then combine that with the constraint $\boldsymbol{R}\widehat{\boldsymbol{\beta}} = \boldsymbol{r}$, and solve for $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\lambda}$.

Solution:

$$\mathcal{L}(\widetilde{\beta}) = \frac{1}{2}(y - X\widetilde{\beta})'(y - X\widetilde{\beta}) + \lambda'(R\widetilde{\beta} - r) =$$

$$= \frac{1}{2}\left[y'y - y'X\widetilde{\beta} - \widetilde{\beta}'X'y + \widetilde{\beta}'X'X\widetilde{\beta}\right] + \lambda'R\widetilde{\beta} - \lambda'r$$

We take the derivative of the objective function:

$$\frac{\partial\mathcal{L}(\widetilde{\beta})}{\partial\widetilde{\beta}} = \frac{1}{2}\left[-X'y - X'y + 2X'X\widetilde{\beta}\right] + R'\lambda$$

Applying the first order conditions for when $\widetilde{\beta} = \hat{\beta}$:

$$-X'y + X'X\hat{\beta} + R'\lambda = 0$$

Premultiplying all this expression by $(X'X)^{-1}$, we end up with:

$$\hat{\beta} = b - (X'X)^{-1}R'\lambda$$

We will find the value of λ by substituting this expression in the constraint $R\hat{\beta} = r$:

$$Rb - R(X'X)^{-1}R'\lambda = r \rightarrow \lambda = (R(X'X)^{-1}R')^{-1}(Rb - r)$$

And, finally:

$$\hat{\beta} = b - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(Rb - r)$$

2. (3 points) Let us define $\widehat{\boldsymbol{\epsilon}} \equiv \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$, the residuals from the restricted regression. Let $\text{SSR}_R = \widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}}$ and recall $\text{SSR}_U = \boldsymbol{e}'\boldsymbol{e}$ with $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{Xb}$. Show that

$$\text{SSR}_R - \text{SSR}_U = \widehat{\boldsymbol{\epsilon}}'\boldsymbol{P}\widehat{\boldsymbol{\epsilon}}, \tag{7}$$

where $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Verify that you have just shown that the F-statistic that we introduced in class can also be written as

$$F = \frac{(\mathrm{SSR}_R - \mathrm{SSR}_U)/\#r}{\mathrm{SSR}_U/(n-K)}. \tag{8}$$

Solution: Since P is symmetric and idempotent, we know that:

$$\hat{\varepsilon}' P \hat{\varepsilon} = (P\hat{\varepsilon})' P \hat{\varepsilon}$$

Where $P\hat{\varepsilon} = P(y - X\hat{\beta}) = \hat{y} - X\hat{\beta} = y - e - X\hat{\beta}$, therefore:

$$\hat{\varepsilon}' P \hat{\varepsilon} = (y - e - X\hat{\beta})'(y - e - X\hat{\beta}) = (y - X\hat{\beta})'(y - X\hat{\beta}) - y'e - e'y + \hat{\beta}'X'e + e'X\hat{\beta} + e'e$$

We already know from PS1 that e is orthogonal to every regressor and also to $\hat{y}$, so $X'e = 0$ and $\hat{y}'e = 0$. Then, from the expression above and knowing that $y = \hat{y} + e$, we get:

$$\hat{\varepsilon}' P \hat{\varepsilon} = \hat{\varepsilon}'\hat{\varepsilon} - e'e = SSR_R - SSR_U$$

Now, from class we know that

$$F = \frac{(Rb - r)' \left[R(X'X)^{-1}R'\right]^{-1} (Rb - r)}{s^2/\#r} = \frac{(Rb - r)' \left[R(X'X)^{-1}R'\right]^{-1} (Rb - r)(n-k)}{e'e/\#r}$$

So, in essence we need to prove that

$$(Rb - r)^T \left[R(X^TX)^{-1}R^T\right]^{-1} (Rb - r) = SSR_R - SSR_U$$

Given the results in a, we can rewrite this expression:

$$(Rb - r)^T \left[R(X^TX)^{-1}R^T\right]^{-1} (Rb - r) = (b - \hat{\beta})'R'\lambda$$

From, the FOC in a, we also know that $R'\lambda = X'y - X'X\hat{\beta} = X'\hat{\varepsilon}$. Therefore:

$$(Rb - r)^T \left[R(X^TX)^{-1}R^T\right]^{-1} (Rb - r) = (b - \hat{\beta})'X'\hat{\varepsilon} = (Xb - X\hat{\beta})'\hat{\varepsilon} = (\hat{\varepsilon} - e)'\hat{\varepsilon} = \hat{\varepsilon}'\hat{\varepsilon} - e'e$$

due to the identity $e'\hat{\varepsilon} = e'(y - X\hat{\beta} = e'y = e'e$.

Finally we have proven what we wanted to prove, that is:

$$(Rb - r)^T \left[R(X^TX)^{-1}R^T\right]^{-1} (Rb - r) = \hat{\varepsilon}'\hat{\varepsilon} - e'e = SSR_R - SSR_U$$