Second Midterm Review Sheet

Ec240a – Second Half, Fall 2020

In preparing for the exam review your lecture notes, assigned course readings, problem sets and prepare answers to the questions on the review sheet(s). The exam is open book, however you are not allowed to consult with classmates or other people (you may freely uses books, notes etc.).

[1] You observe a simple random sample of size N from the population

$$Y_0 \sim N\left(\mu, \sigma^2\right)$$

as well as a second, independent, simple random sample, also of size N, from the population

$$Y_1 \sim N\left(\mu, 4\sigma^2\right).$$

The value of σ^2 is known. Consider the family of estimates of μ

$$\hat{\mu}\left(c_0, c_1\right) = c_0 \bar{Y}_0 + c_1 \bar{Y}_1,$$

where $\bar{Y}_0 = \frac{1}{N}\sum_{i=1}^N Y_{0i}$ and $\bar{Y}_1 = \frac{1}{N}\sum_{i=1}^N Y_{1i}$.

[a] Show that mean squared error equals

$$\mathbb{E}\left[\left(\hat{\mu}\left(c_0, c_1\right) - \mu\right)^2\right] = \frac{c_0^2\sigma^2}{N} + \frac{c_1^2 4\sigma^2}{N} + \left(1 - c_0 - c_1\right)^2 \mu^2. \tag{1}$$

[b] Derive the oracle estimator (within the family) which minimizes (1).

[c] Show that

$$\hat{R}\left(c_0, c_1\right) = \frac{c_0^2\sigma^2}{N} + \frac{c_1^2 4\sigma^2}{N} + \left(1 - c_0 - c_1\right)^2 \frac{1}{2}\left\{\bar{Y}_0^2 + \bar{Y}_1^2 - \frac{5\sigma^2}{N}\right\} \tag{2}$$

is an unbiased estimate of of (2). Can you propose another unbiased risk estimate? Why would you prefer one unbiased risk estimate over another?

[d] Describe in *words* how one might use (2) to construct an implementable estimator of μ.

[2] Let Y denote log-earnings and X years of completed schooling for a cohort of workers. Assume a random sample of size N is available from this population. Let $D_x = 1$ if $X = x$ and zero otherwise. Assume that $X \in \{0, \ldots, 16\}$ with positive probability attached to each support point.

[a] Let

$$\mathbb{E}^*\left[Y | D_1, \ldots, D_L\right] = \alpha_0 + \sum_{l=1}^{16} \gamma_{0l} D_l.$$

What is the relationship between this linear predictor and $\mathbb{E}\left[Y | X = x\right]$?

[b] Assume that $\Pr\left(X = 6\right) = 0$. Is the linear predictor defined in part [a] still well-defined? Why or why not?

[c] You hypothesize that $\mathbb{E}\left[Y | X = x\right]$ is linear in x. Consider the linear predictor in part [a] and let $\beta = \left(\alpha, \gamma_1, \ldots, \gamma_{16}\right)'$. Show how your hypothesis may be equivalently expressed as set of linear restrictions of the form $C\beta_0 = c$. Provide explicit expressions for C and c. Describe how you would construct a test statistic for your hypothesis. What is the asymptotic sampling distribution of your statistic under the null?

Assume that your have a consistent estimate $\hat{\Lambda}$ of the asymptotic variance-covariance matrix of $\sqrt{N}\left(\hat{\beta}-\beta\right)$, with $\hat{\beta}$ the least squares estimate.

[3] Let $X \in \{0,1,2\}$ and $Y \in \{0,1,2\}$. The probability of the event $X = x$ and $Y = y$ for all possible combinations of x and y is given in the following table:

X\Y	0	1	2
0	$\frac{15}{100}$	$\frac{10}{100}$	$\frac{5}{100}$
1	$\frac{10}{100}$	$\frac{20}{100}$	$\frac{10}{100}$
2	$\frac{5}{100}$	$\frac{10}{100}$	$\frac{15}{100}$

[a] Calculate $\mathbb{E}[Y]$ and $\mathbb{E}[Y|X=1]$. Are X and Y independent?

[b] Calculate $\mathbb{E}[X]$, $\mathbb{E}\left[X^2\right]$ and hence $\mathbb{V}(X)$.

[c] Calculate $\mathbb{C}(X,Y)$ and also the coefficient on X in $\mathbb{E}^*[Y|X]$.

[d] Calculate the intercept of $\mathbb{E}^*[Y|X]$.

[e] Repeat [a] to [d] above for the following joint distribution

X\Y	0	1	2
0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

[4] Let $X_1,...,X_K$ be a set of regressors with the property that $\mathbb{C}(X_k,X_l)=0$ for all $k \neq l$. We will show that

$$\mathbb{E}^*[Y|X_1,\ldots,X_K] = \sum_{k=1}^{K}\mathbb{E}^*[Y|X_k] - (K-1)\,\mathbb{E}[Y].$$

[a] First show that

$$\mathbb{E}^*\left[\mathbb{E}^*[Y|X_k]|X_l\right] = \mathbb{E}[Y]$$

for every $k \neq l$.

[b] Second verify the orthogonality conditions

$$\mathbb{E}[UX_l] = 0$$

for $U = \left(Y - \sum_{k=1}^{K}\mathbb{E}^*[Y|X_k] + (K-1)\,\mathbb{E}[Y]\right)$ and $l = 1,\ldots,K$.

[5] Let Y be a scalar random variable, X a K vector of covariates (which includes a constant), and W a vector of additional covariates (which excludes a constant). Consider the long (linear) regression

$$\mathbb{E}^*[Y|W,X] = X'\beta_0 + W'\gamma_0. \tag{3}$$

Next define the short and auxiliary regressions

$$\mathbb{E}^*[Y|X] = X'b_0 \tag{4}$$

$$\mathbb{E}^*[W|X] = \Pi_0 X. \tag{5}$$

[a] Let $V = W - \mathbb{E}^*[W|X]$ be the projection error associated with the auxiliary regression. Show that

$$\begin{aligned}\mathbb{E}^*[Y|V,X] &= \mathbb{E}^*[Y|X] + \mathbb{E}^*[Y|1,V] - \mathbb{E}[Y] \\ &= \mathbb{E}^*[Y|X] + \mathbb{E}^*[Y|V]\end{aligned}$$

where $\mathbb{E}^*[Y|1,V]$ denotes the linear regression of Y onto a constant and V, while $\mathbb{E}^*[Y|V]$ denotes the corresponding regression without a constant (HINT: Observe that $\mathbb{C}(X,V) = 0$).

[b] Next show that $\mathbb{E}^*[Y|V,X] = \mathbb{E}^*[Y|W,X]$ and hence that the coefficient on V in $\mathbb{E}^*[Y|V,X]$ coincides with that on W in $\mathbb{E}^*[Y|W,X]$.

[c] Let $U = Y - \mathbb{E}^*[Y|X]$ be the projection error associated with the short regression. Derive the coefficient on V in the linear regression of U onto V (excluding a constant).

[d] Discuss the possible practical value of the results shown in [b] and [c] above.

[6] Consider the following model of supply and demand:

$$\begin{aligned}\ln Q_i^D(p) &= \alpha_1 + \alpha_2 \ln(p) + U_i^D \\ \ln Q_i^S(p) &= \beta_1 + \beta_2 \ln(p) + U_i^S,\end{aligned}$$

with i indexing a generic random draw from a population of 'markets' ; U_i^D and U_i^S are market-specific demand and supply shocks. We assume that $\left(U_i^S, U_i^D\right) \overset{i.i.d}{\sim} F$ for $i = 1, 2, \ldots, N$. In each market the observed price and quantity pair (P_i, Q_i) coincides with the solution to market clearing condition

$$Q_i^D(P_i) = Q_i^S(P_i) = Q_i.$$

[a] Provide an economic interpretation of the parameters α_2 and β_2. What signs do you expect them to take? Why?

[b] Depict the market equilibrium graphically. Solve for the equilibrium values of $\ln Q_i$ and $\ln P_i$ algebraically. How is the market price and quantity related to the demand and supply shocks, U_i^D and U_i^S? Provide some economic content for your answer. Can you use a figure to illustrate it?

[c] Calculate $\mathbb{E}^*[\ln Q|\ln P]$. You may assume that $\mathbb{C}\left(U^D, U^S\right) = 0$. Evaluate the coefficient on $\ln(P)$, does it coincide with an economically interpretable parameter? Assume that $\mathbb{V}\left(U_i^S\right) / \left(\mathbb{V}\left(U_i^S\right) + \mathbb{V}\left(U_i^D\right)\right) \approx 1$, does your answer change? Why?

[7] This question is about the Rambly Shambly Hex Bolt Corporation. Let Y_t be the number of hex bolts produced by Rambly Shambly Hex in year t, M_t tons of steel used in production, K_t total factory capital stock, and L_t total person-hours worked. We assume that

$$Y_t = A_t M_t^{\alpha} K_t^{\beta} L_t^{\gamma}.$$

[a] Rambly Shambly Hex is owned by an eccentric billionaire who chooses M_t, K_t and L_t each year randomly using an eternally unchanging roulette-wheel-like-device (i.e., inputs are chosen independently of each other and independently of A_t). Further assume that the distribution of A_t is i.i.d. over time. Show that under this input choice mechanism that

$$\mathbb{E}^*[\ln Y_t|\ln M_t, \ln K_t, \ln L_t] = \lambda + \alpha \ln M_t + \beta \ln K_t + \gamma \ln L_t$$

with $\lambda = \mathbb{E}[\ln A_t]$. Is this same result likely to hold if Rambly Shambly instead chose input levels to maximize profits? Why or why not?

[b] Further show that under the completely random input choice scheme described above that:

$$\mathbb{E}^*[\ln Y_t | \ln M_t, \ln K_t, \ln L_t] = \mathbb{E}^*[\ln Y_t | \ln M_t] + \mathbb{E}^*[\ln Y_t | \ln K_t] + \mathbb{E}^*[\ln Y_t | \ln L_t] - 2\mathbb{E}[\ln Y_t].$$

[c] Let $X_t = (\ln M_t, \ln K_t, \ln L_t)'$ and $\sigma^2 = \mathbb{V}(\ln A_t)$. Argue that under the completely random input choice scheme:

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{D} N(0, \Lambda),$$

for $\theta = (\alpha, \beta, \gamma)'$, $\Lambda = \sigma^2 \mathbb{V}(X_t)^{-1}$ and $\hat{\theta}$ estimated by the OLS fit of $\ln Y_t$ onto a constant and X_t for $t = 1, \ldots, T$. What are the values of the off-diagonal elements of $\mathbb{V}(X_t)$?

[d] For $T = 3,859$, an OLS fit gives

$$\hat{\theta} = \begin{pmatrix} 0.32 \\ 0.36 \\ 0.42 \end{pmatrix}, \quad \hat{\Lambda} = \begin{pmatrix} 5 & 1/200 & 2/1000 \\ 1/200 & 3 & 3/1000 \\ 2/1000 & 3/1000 & \frac{198}{100} \end{pmatrix}.$$

Construct a Wald Statistic (carefully explaining each step in the construction) for the null hypothesis of constant returns to scale (i.e., $H_0 : \alpha + \beta + \gamma = 1$). What is the appropriate reference distribution and critical value for a two-sided test with size $\alpha = 0.05$? Do you reject the null?

[8] You are given a random sample from South Africa in the late 1980s. Each record in this sample includes, Y, an individual's log income at age 40, X the log permanent income of their parents, and D a binary indicator equaling 1 if the respondent is White and zero if they are Black. Let the best linear predictor of own log income at age forty given parents' log permanent income and own race be

$$\mathbb{E}^*[Y | X, D] = \alpha_0 + \beta_0 X + \gamma_0 D.$$

[a] Let $Q = \Pr(D = 1)$, assume that $\mathbb{V}(X | D = 1) = \mathbb{V}(X | D = 0) = \sigma^2$ and recall the analysis of variance formula $\mathbb{V}(X) = \mathbb{V}(\mathbb{E}[X | D]) + \mathbb{E}[\mathbb{V}(X | D)]$. Show that

$$\mathbb{V}(X) = Q(1 - Q)\{\mathbb{E}[X | D = 1] - \mathbb{E}[X | D = 0]\}^2 + \sigma^2.$$

[b] Let $\mathbb{E}^*[D | X] = \kappa + \lambda X$. Show that

$$\lambda = \frac{Q(1 - Q)\{\mathbb{E}[X | D = 1] - \mathbb{E}[X | D = 0]\}}{Q(1 - Q)\{\mathbb{E}[X | D = 1] - \mathbb{E}[X | D = 0]\}^2 + \sigma^2}.$$

[c] Assume that $\beta_0 = 0$. Show that in this case $\gamma_0 = \mathbb{E}[Y | D = 1] - \mathbb{E}[Y | D = 0]$.

[d] Let $\mathbb{E}^*[Y | X] = a + bX$. Maintaining the assumption that $\beta_0 = 0$ show that

$$b = \frac{Q(1 - Q)\{\mathbb{E}[Y | D = 1] - \mathbb{E}[Y | D = 0]\}\{\mathbb{E}[X | D = 1] - \mathbb{E}[X | D = 0]\}}{Q(1 - Q)\{\mathbb{E}[X | D = 1] - \mathbb{E}[X | D = 0]\}^2 + \sigma^2}.$$

[e] Let $Q(1-Q) = 1/10$, $\sigma^2 = 3/10$ and $\mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0] = \mathbb{E}[X|D=1] - \mathbb{E}[X|D=0] = 3$. Provide a numerical value for $\mathbb{V}(X)$ and b.

[f] On the basis of β_0 a member of the National Party argues that South Africa is a highly mobile society. One the basis of b a member of the African National Congress argues that it is a highly immobile one. Comment on the relative merits of these two assertions.

[9] Consider the following joint probability density function for x and y

$$f_{X,Y}(x,y) = \begin{cases} \frac{3}{11}\left(x^2+y\right), & (x,y) \in [0,2] \times [0,1] \\ 0 & (x,y) \notin [0,2] \times [0,1] \end{cases}$$

[a] Show that $f_{X,Y}(x,y)$ is a valid probability density function.

[b] Compute the conditional density $f_{Y|X}(y|x)$.

[c] Compute the conditional expectation function $\mathbb{E}[Y|X=x]$.

[d] Compute the linear predictor $\mathbb{E}^*[Y|X=x]$.

[e] Consider a joint distribution with a conditional that coincides with the one derived in part [b] above, but where the marginal distribution of X is uniform on $[0,2]$. Explain, qualitatively, how this change would affect your answers in parts [c] and [d] above?

[10] Consider the population of married men. Let Y denote log earnings for a generic random draw from this population, X his years of completed schooling and W the schooling of his spouse. Assume that the conditional mean of own log earnings given own and spouse's schooling is

$$\mathbb{E}[Y|X,W] = \alpha_0 + \beta_0 X + \gamma_0 W,$$

while the best linear predictor of spouse's schooling given own schooling is

$$\mathbb{E}^*[W|X] = \delta_0 + \zeta_0 X.$$

You may assume that the joint distribution of (W,X,Y) is such that these objects are well-defined. You may assume that all the slope coefficients in the two equations above are positive.

[a] Show that $\zeta_0 = \rho_{WX}\frac{\sigma_W}{\sigma_X}$, with ρ_{WX} the correlation of W with X, and σ_W and σ_X respectively the standard deviation of W and X. Further show that $\delta_0 = \mu_W - \rho_{WX}\frac{\sigma_W}{\sigma_X}\mu_X$ with μ_W and μ_X denoting the population means of W and X.

[b] Using your answers in [a] above, as well as the form of $\mathbb{E}[Y|W,X]$, provide an expression for $\mathbb{E}^*[Y|X]$.

[c] Consider another population of married men where $F_{Y|W,X}(y|W=w,X=x)$, $F_W(w)$ and $F_X(x)$ coincide with those for the population described above, but where $F_{W,X}(w,x)$ differs. Assume that in this alternative population $\rho_{WX} = 0$. Solve for $\mathbb{E}[Y|X,W]$, $\mathbb{E}^*[W|X]$ and $\mathbb{E}^*[Y|X]$. Use the notation established in parts [a] and [b] to formulate your answer.

[d] Assume that $F_W(w)$ and $F_X(x)$ are identical and that marriage is homogamous in terms of education so that $W = X$ for all couples (i.e., individuals choose partners with identical levels of education). Show that in this world $\rho_{WX} = 1$. Solve for $\mathbb{E}[Y|X,W]$, $\mathbb{E}^*[W|X]$ and $\mathbb{E}^*[Y|X]$. Use the notation established in parts [a] and [b] to formulate your answer.

[e] Compare the form of $\mathbb{E}^*[Y|X]$ in the original population with that in the two alternative populations of parts [c] and [d]. In which population does log earnings rise most steeply with years of schooling? Provide some intuition for your answer (5 sentences).

[f] Assume that schooling is binary valued, taking on the values 0,1. Let R_W be a 2×1 vector equal to $(1,0)'$ if $W = 0$ and $(0,1)'$ if $W = 1$. Let S_X be the analogous 2×1 vector defined using X. Let $T_{WX} = (R_W \otimes S_X)$ and

$$\mathbb{E}^*[Y|T_{WX}] = T'_{WX}\pi,$$

where a constant is *not* included and $\pi = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})'$. Show that

$$\pi_{jk} = \mathbb{E}[Y|W = j, X = k].$$

[g] Consider the null hypothesis that $\mathbb{E}[Y|W, X] = \alpha_0 + \beta_0 X + \gamma_0 W$. Maintaining this null find an explicit expression for each component of π in terms of α_0, β_0 and γ_0. Express this null in the form $C\pi = c$ for some matrix of constants C and vector of constants c.

[h] Let $W = 1$ if a wife has completed primary school and zero otherwise, let $X = 1$ if a husband has completed primary school and zero otherwise. A least squares fit, loosely based on data from Brazil, of log husband's earnings on T_{WX} as defined in [f] using a random sample of size $N = 50,000$ yields point estimate of

$$\hat{\pi} = \begin{pmatrix} 5.50 \\ 6.00 \\ 5.00 \\ 7.00 \end{pmatrix}$$

with an estimated asymptotic variance-covariance matrix of

$$\hat{\Lambda} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 7 & 0 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Can you reject the null hypothesis (at the $\alpha = 0.05$ level) formulated in part [g] on the basis of this sample? For your reference the 0.95 quantiles of χ^2 random variables with parameters 1, 2 and 3 are, respectively, 3.84, 5.99 and 7.81.

[11] Available is a random sample of $i = 1, \ldots, N$ farmers. We have measures of output, Y_t, and (labor) input, X_t, for each farmer (on a per hectare basis) for each of $t = 1, \ldots, T$ years. The sample is

$$\{(Y_{i1}, \ldots, Y_{iT}, X_{i1}, \ldots, X_{iT})'\}_{i=1}^{N}.$$

We assume that, for $Y_t = \ln O_t$ and $X_t = \ln L_t$, output per hectare, O_t, equals

$$O_t = L_t^{\beta} Q^{\gamma} \exp(U_t)$$

with L_t labor, Q soil quality − which is unobserved by the econometrician − and U_t a stochastic, and also unobserved, input outside of the farmer's control (e.g., rainfall).

Our behavioral assumption is that the farmer chooses L_t to maximize expected profits. She knows period t output and input prices, respectively P_t and W_t, as well as her soil quality, Q.

$$L_t = \arg\max_{l} \mathbb{E}\left[P_t l^{\beta} Q^{\gamma} \exp\left(U_t\right) - W_t l \middle| P_t, W_t, Q\right]$$

She does not know rainfall, assumed non-forecastable by anything in her information set, with marginal distribution

$$U_t \sim N\left(0, \sigma_U^2\right).$$

You may also assume that (P_t, W_t) varies independently of all other variables in the model and is also independently and identically distributed across farms and over time.

[a] Show that the farmer's $t$labor input is

$$X_t = \mu + \frac{1}{1-\beta} A + V_t$$

with $\mu = \frac{1}{1-\beta}\left(\ln\beta + \frac{\sigma_U^2}{2}\right)$, $A = \gamma \ln Q$ and $V_t = \frac{1}{1-\beta} \ln\left(\frac{P_t}{W_t}\right)$. How does rainfall risk affect the farmer's chosen labor input level? Soil quality?

[b] Show that

$$\mathbb{E}\left[Y_t \middle| X_t, A\right] = \beta X_t + A.$$

Why is conditioning on land quality alone sufficient to identify β? What maintained assumption is important for this result?

[c] Show that, for $\sigma_A^2 = \mathbb{V}(A)$, and $\sigma_V^2 = \mathbb{V}(V_t)$ for $t = 1, \ldots, T$, that

$$\begin{aligned} \mathbb{C}(A, X_t) &= \frac{1}{1-\beta}\sigma_A^2 \\ \mathbb{V}(X_t) &= \left(\frac{1}{1-\beta}\right)^2 \sigma_A^2 + \sigma_V^2 \\ \mathbb{E}[X_t] &= \mu + \frac{1}{1-\beta}\mathbb{E}[A] + \mathbb{E}[V_t]. \end{aligned}$$

[d] Next use your results in part [c] to show that

$$\mathbb{E}^*\left[A \middle| X_t\right] = \eta_0 + \eta_1 X_t$$

where

$$\begin{aligned} \eta_1 &= (1-\beta)\left[1 + (1-\beta)^2 \frac{\sigma_V^2}{\sigma_A^2}\right]^{-1} \\ \eta_0 &= E[A] - \eta_1\left[\mu + \frac{1}{1-\beta}\mathbb{E}[A] + \mathbb{E}[V_t]\right]. \end{aligned}$$

[e] Using your results from parts [b] and [d] solve for the coefficient, say b, on X_t in $\mathbb{E}^*\left[Y_t \middle| X_t\right]$. Does $b = \beta$? Discuss economic conditions under which $b \approx \beta$ as well as conditions where $b >> \beta$.

[f] Let $X = (X_1, \ldots, X_T)'$. Solve for $\mathbb{V}(X)$, $\mathbb{C}(A, X)$ and hence the coefficients $\delta = (\delta_1, \ldots, \delta_T)'$ on

$X_1, \ldots, X_T$ in the linear regression

$$\mathbb{E}^*\left[A \middle| X\right] = \lambda + X'\delta.$$

[g] Let $Y = (Y_1, \ldots, Y_T)'$ and $W = (\mathbf{1}, X')'$ Solve for the multivariate linear predictor

$$\mathbb{E}^*\left[Y \middle| X\right] = \Pi W.$$

Specifically, express the elements of the $T \times (1+T)$ matrix Π in terms of β, λ and δ.

[h] Let $\theta = (\beta, \lambda, \delta')'$. Let $\pi = \operatorname{vec}(\Pi')$ be the $T(T+1) \times 1$ vector constructed by vertically stacking the columns of Π. Let G be a $T(T+1) \times (1+1+T)$ matrix such that

$$\pi = G\theta.$$

Derive the form of G.

[i] Consider the estimator

$$\hat{\pi} = \left[\frac{1}{N}\sum_{i=1}^{N}(I_T \otimes W_i)(I_T \otimes W_i)'\right]^{-1} \times \left[\frac{1}{N}\sum_{i=1}^{N}(I_T \otimes W_i)Y_i\right].$$

Further define $U = Y - (I_T \otimes W_i)'\pi$ and

$$\Gamma = I_T \otimes \mathbb{E}\left[WW'\right]$$

and

$$\Omega = \mathbb{E}\left[(I_T \otimes W_i)UU'(I_T \otimes W_i)'\right].$$

Argue that $\hat{\pi} \xrightarrow{p} \pi$ and furthermore also argue that $\sqrt{N}(\hat{\pi} - \pi) \xrightarrow{D} \mathcal{N}(0, \Lambda)$ for $\Lambda = \Gamma^{-1}\Omega\Gamma^{-1}$.

[12] Consider two large populations of households and firms, both of which may costlessly migrate across $i = 1, \ldots, N$ different localities. Conditional on choosing to reside in location i, households $t = 1, \ldots, M_i$, choose the amount of unimproved land, l, and composite commodity, x, to consume in order to maximize utility:

$$\max_{l,x} Q_i l^{\gamma} x^{1-\gamma} \ s.t.\ H_{it}W_i \leq R_i l + x, \tag{6}$$

where Q_i, W_i and R_i respectively denote the "quality of life", wage rate per efficiency unit of labor, and rent per unit of unimproved land, in locality i. The price of the composite commodity is fixed on world markets and normalized to one. A household's endowment of efficiency units of labor is given by H_{it} so that their total budget conditional on residence in locality i is $H_{it}W_i$. Let L_{it} and X_{it} denote the household's utility-maximizing land and composite commodity consumption.

Equilibrium requires that households are indifferent between locations, or that the indirect utility associated with one efficiency unit of labor is constant across communities. Under (6) this condition takes the form

$$V(W, R; Q) = \psi\frac{QW}{R^{\gamma}} = \bar{u},\ \psi = \gamma^{\gamma}(1-\gamma)^{1-\gamma}, \tag{7}$$

where $\bar{u}$ it the common, nationwide, utility level (available to an owner of an efficiency unit of labor).

Firm $f = 1, \dots, L_i$ in locality i produces X_{if} units of the composite commodity using a constant returns to scale production technology requiring three inputs: (i) efficiency units of labor, n ; (ii) units of improved land, l , and (iii) capital, k. Labor and land prices in a locality are determined in equilibrium, while the cost of capital is constant and equal to σ. The firm's problem is to minimize costs:

$$\min_{n,l,k} W_i n + R_i l + \sigma k, \ s.t. \ A_i n^{\alpha} l^{\gamma} k^{1-\alpha-\beta} = X_{if}, \tag{8}$$

where A_i is locality-specific total factor productivity. Free entry ensures that profits are zero in equilibrium. The zero profit condition requires that unit costs equal the normalized price of the composite good:

$$C(W, R, A) = \xi \frac{W^{\alpha} R^{\beta}}{A} = 1, \ \xi = \left(\frac{1}{\alpha}\right)^{\alpha} \left(\frac{1}{\beta}\right)^{\beta} \left(\frac{1}{1-\alpha-\beta}\right)^{1-\alpha-\beta} \sigma^{1-\alpha-\beta} \tag{9}$$

[a] Use (7) and (9) to show that the equilibrium wage and rent in locality i is given by

$$\begin{aligned}
\ln W_i &= \kappa_W + \frac{\gamma}{\alpha\gamma + \beta} \ln A_i - \frac{\beta}{\alpha\gamma + \beta} \ln Q_i \\
\ln R_i &= \kappa_R + \frac{1}{\alpha\gamma + \beta} \ln A_i + \frac{\alpha}{\alpha\gamma + \beta} \ln Q_i,
\end{aligned} \tag{10}$$

with

$$\begin{aligned}
\kappa_W &= \frac{1}{\alpha\gamma + \beta} (\beta \ln \bar{u} - \beta \ln \psi - \gamma \ln \xi) \\
\kappa_R &= \frac{1}{\alpha\gamma + \beta} (\alpha \ln \bar{u} - \alpha \ln \psi - \ln \xi).
\end{aligned}$$

HINT: You do not need to verify the form of the intercepts in (10).

[b] Assume that A_i and Q_i vary independently. Let $\phi = \frac{\gamma^2 \mathbb{V}(\ln A_i)}{\gamma^2 \mathbb{V}(\ln A_i) + \beta^2 \mathbb{V}(\ln Q_i)}$. Show that the coefficient on $\ln W_i$ in the (mean squared error minimizing) linear predictor of $\ln R_i$ onto a constant and $\ln W_i$ equals

$$b = \frac{1}{\gamma}\phi - \frac{\alpha}{\beta}(1 - \phi).$$

Interpret this expression. Under what conditions is b positive? Negative? Why?

[c] Assume that the $\gamma = 1/3$, $\alpha = 3/5$, and $\beta = 1/5$. A macro economist, with considerable central banking experience, asserts that the "quality-of-life" differences across cities are "overblown". An econometrician, with almost no "real world" experience, asserts that "firms are equally unproductive in all places". An urban economist claims that "actually both are equally important".

Using a random sample of US cities you compute an estimate of b equal to -3. Who is correct? The macro-economist, econometrician or urban economist?

[d] It turns out that you made a Python coding error and that the correct estimate of b is zero. Who do you believe to be correct now?

[e] Use (10) to solve for $\ln A_i$ and $\ln Q_i$ (again you may ignore constants). Use the parameter values given in (c) to simplify your expressions.

[f] You observe mean January temperature for each city in your sample. Can you compute consistent

estimates of the semi-elasticity of total factor productivity and quality of life with respect to temperature? Describe your procedure.

[13] Let $X \sim \text{Uniform}\,[-1,1]$ and assume that

$$Y = -\frac{2}{3} + X^2 + V, \;\; V|\, X \sim \mathcal{N}\left(0, \sigma^2\right) \tag{11}$$

[a] Calculate $\mathbb{E}\left[Y|\, X\right]$

[b] Calculate $\mathbb{E}\left[X^2\right]$ and $\mathbb{V}\left(X\right)$

[c] Calculate $\mathbb{E}\left[Y\right]$

[d] Calculate $\mathbb{C}\left(X,Y\right)$ and also the coefficient on X in $\mathbb{E}^*\left[Y|\, X\right]$.

[e] Let $U = Y - \mathbb{E}^*\left[Y|\, X\right]$. Show that $\mathbb{C}\left(U,X\right) = 0$. Give an intuitive explanation for this result.

[f] Find a function $g\left(X\right)$ such that that $\mathbb{C}\left(U, g\left(X\right)\right) = \mathbb{V}\left(X^2\right)$. Give an intuitive explanation for your answer.

[g] Describe how your answers in (d) to (f) would change if (11) held but now $X \sim \text{Uniform}\,[0,2]$. You may find it helpful to sketch a figure.