Econometrics I Comprehensive Exam Questions

Instructions. Read Carefully.

The following 2 questions are from the **second** half of the Econometrics 1 class. You should do both questions

1. The following is the fitted regression of wage (in dollars per hour) on a dummy variable for gender (female=1 if the person is female) a dummy variable for marital status (married=1 if married) and the interaction between the two dummy variables (femmarr=female*married):

$$\widehat{wage} = 5.2 - 0.6\text{female} + 2.8\text{married} - 2.8\text{femmarr}$$

The heteroskedasticity robust variance covariance matrix (lower triangle) is given by,

$$\hat{V} = \begin{pmatrix} 0.086 & & & \\ 0.086 & 0.16 & & \\ 0.086 & 0.086 & 0.19 & \\ 0.086 & 0.16 & 0.19 & 0.29 \end{pmatrix}$$

Note that the diagonals give the estimated variance of the coefficients of the variables in the following order: intercept, female, married, femmarr

(a) (5 points) Provide an interpretation for each of the coefficients in the fitted regression.

(b) (3 points) Test the hypothesis that the coefficient on female is statistically significant at the 5% level (use a the standard normal critical value of 1.96).

(c) (5 points) Suppose I instead used the variable male=1-female and its interaction with married in place of female and femmarr respectively. Derive the fitted regression for this version of the model.

(d) (3 points) Find the standard error for the coefficient on married in the regression in c.

2. Suppose we have a random sample (iid) of n observations $\{y_i, x_{1i}, x_{2i}\}_{i=1}^n$ such that,

$$\begin{aligned} y_i &= x_{1i}\beta_1 + e_i \\ E(e_i|x_{1i}, x_{2i}) &= 0,\ E(e_i^2|x_{1i}, x_{2i}) = \sigma^2 x_{2i}^2 \\ & x_{1i} \text{ and } x_{2i} \text{ are statistically independent} \\ E(x_{1i}) &= E(x_{2i}) = 0 \\ E(x_{1i}^2) &= \omega_1, E(x_{2i}^2) = \omega_2, E(x_{2i}^4) = \kappa_2 \end{aligned}$$

The OLS estimator for β_1 where x_{2i} is omitted and there is no intercept is given by:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n x_{1i} y_i}{\sum_{i=1}^n x_{1i}^2}$$

The OLS estimators for β_1 and β_2 (which is zero given the model) when x_{2i} is included is given by:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \left(\sum_{i=1}^n x_i x_i' \right)^{-1} \sum_{i=1}^n x_i y_i$$

where,

$$x_i' = \begin{pmatrix} x_{1i} & x_{2i} \end{pmatrix}$$

You may assume that enough moments exist to allow one to use the Laws of Large Numbers and Central Limit Theorem for iid data. Statistical Independence of x_{1i} and x_{2i} implies $E(g(x_{1i})h(x_{2i})) = E(g(x_{1i}))E(h(x_{2i}))$ for any functions g and h for which the expectations exist

(a) (5 points) Under the conditions above show that $\tilde{\beta}_1$ is consistent for β_1 and derive its asymptotic distribution and asymptotic variance. The asymptotic variance should be in terms of σ^2, ω_1 and ω_2.

(b) (3 points) Would the usual default (i.e. that assume homoskedastic residuals) standard errors be valid for $\tilde{\beta}_1$? Explain your reasoning.

(c) (5 points) Show that $\hat{\beta}_1$ and $\hat{\beta}_2$ are also consistent. Derive an expression for the asymptotic variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ in terms of σ^2, ω_1, ω_2 and κ_2.

(d) (5 points) Would the usual default standard errors be valid for either $\hat{\beta}_1$ or $\hat{\beta}_2$ or both? Justify your answer and if a default standard error is invalid indicate whether it will be (asymptotically) too large or too small.

June 2021 Comprehensive Exam

3. Suppose that Y is a random variable that has the density
$f(y) = \frac{1}{y\sqrt{2\pi}} \exp(-\frac{(\log y - \mu)^2}{2})$, where μ is the unknown parameter. log is the natural log (base e).

(**3.a**) [8 points] Set up the M-estimation problem that corresponds to maximum likelihood estimation of the parameter μ. Be clear about the specific functional form of the objective function.

(**3.b**) [8 points] Solve the M-estimation problem. What is the maximum likelihood estimator $\hat{\mu}_N$ of μ?

It can be shown that $Z = \log(Y)$ has a Normal distribution with mean μ and variance 1.

(**3.c**) [5 points] Discuss whether $\hat{\mu}_N$ is a consistent estimator of μ. If it is consistent, provide a rigorous justification (e.g., provide a proof). If it is not consistent, explain why not.

(**3.d**) [5 points] Discuss whether $\hat{\mu}_N$ is an unbiased estimator of μ. If it is unbiased, provide a rigorous justification (e.g., provide a proof). If it is not unbiased, explain why not.

(**3.e**) [5 points] What is the distribution of $\hat{\mu}_N$ in repeated samples? Be specific, naming the distribution and giving the values of the parameters of the distribution.

It can be shown that $E(Y) = \exp(\mu + \frac{1}{2})$.

(**3.f**) [5 points] Use this information to come up with a *different* estimator $\tilde{\mu}_N$ of μ, and discuss whether $\tilde{\mu}_N$ is consistent. "Different estimator" means different from the maximum likelihood estimator $\hat{\mu}_N$.

4. This question is about assessing the credibility of particular empirical strategies. There are two unrelated questions along these lines. For each of the following questions, your answer should include the following, in addition to anything else specifically mentioned in that question:

- Give an explanation of the assumption(s) under consideration, with specific reference to the empirical setting. What do the assumption(s) mean in the empirical setting?
- Discuss arguments both for and against the assumption(s). Also, come to a final conclusion: Overall, is there a better argument for or against the assumption(s)?

Be sure to explain and justify your reasoning.

(**4.a**) [15 points] Suppose that we are interested in the causal effect of participating in a job skills program on wages. The outcome Y_i is wages, and the treatment X_i is a binary indicator for participating in the job skills program. Suppose that participation in the job skills program is assigned as part of an experiment, where the probability that an individual is invited to participate in the job skills program is a function of the individual's educational level. The probability of being invited to participate is lower for individuals with greater education. Assume that all individuals that are invited to participate actually do participate in the job skills program. The cross-sectional dataset contains information on wages, job skills program status, education, and age.

First, consider the assumption of *conditional unconfoundedness*, where the "conditioning" is on age. Second, consider the assumption of *conditional unconfoundedness*, where the "conditioning" is on age and education. These two assumptions are to be made individually (i.e., consider just the first assumption and then consider just the second assumption). As detailed above, give an explanation of each of those assumptions, and discuss arguments both for and against those assumptions. Highlight the *differences* in your answers.

(**4.b**) [15 points] Suppose that we are interested in the causal effect of a subsidy for electric cars on pollution levels. The outcome Y_{it} is the pollution level in a particular state i in a particular year t, and the treatment X_{it} is the subsidy an individual receives due to purchasing an electric car in state i in year t. We have panel data, and work with the standard linear fixed effects model $Y_{it} = X_{it}\beta + c_i + u_{it}$. The states are indexed by $i = 1, 2, \dots, N$ and the years are indexed by $t = 1, 2, \dots, T$.

Consider the assumption that the explanatory variables are *strictly exogenous conditional on the unobserved effect*. As detailed above, give an explanation of that assumption, and discuss arguments both for and against the assumption. Highlight the role of *why the subsidy is in effect* in particular states in particular years.