Final Exam (January 2021)

Problem 1. (15 Points) The family of Pareto distributions has been used as a model for a density function with a slowly decaying tail:

$$f(x|x_0, \theta) = \theta x_0^{\theta} x^{-\theta-1}$$

for $x \geqslant x_0$, $\theta > 1$. Assume that x_0 is given and that $X_1, ..., X_n$ is an i.i.d. sample with density $f(\cdot|x_0, \theta_*)$.

(i) We can find that $\mathrm{E}[X_1] = \frac{x_0\theta_*}{\theta_*-1}$ (you are not required to prove this). Find an estimator of θ_* based on the method of moments.

(ii) Find the maximum likelihood estimator of θ_*, $\hat{\theta}_n^{MLE}$.

(iii) Find the asymptotic variance of the maximum likelihood estimator. i.e. Find σ^2 such that

$$\sqrt{n}\left(\hat{\theta}_n^{MLE} - \theta_*\right) \to_d \mathrm{N}\left(0, \sigma^2\right).$$

Problem 2. (15 Points) Consider the following linear model:

$$Y = \beta_0 + \beta_1 X + U.$$

Suppose $\mathrm{E}[U] = 0$ and $\mathrm{E}[XU] \neq 0$, but you have two valid instruments, Z_1 and Z_2 ($\mathrm{E}[Z_1 U] = \mathrm{E}[Z_2 U] = 0$). Write the first-stage regression as

$$X = \pi_0^X + \pi_1^X Z_1 + \pi_2^X Z_2 + \epsilon, \tag{1}$$

where $\mathrm{E}[\epsilon] = \mathrm{E}[\epsilon Z_1] = \mathrm{E}[\epsilon Z_2] = 0$.

(i) Explain how you compute the 2SLS estimator of β_1.

(ii) Also write the linear regression of Y on Z_1 and Z_2 as

$$Y = \pi_0^Y + \pi_1^Y Z_1 + \pi_2^Y Z_2 + \nu, \tag{2}$$

where $\mathrm{E}[\nu] = \mathrm{E}[\nu Z_1] = \mathrm{E}[\nu Z_2] = 0$. Let $\hat{\pi}_1^X$ and $\hat{\pi}_1^Y$ denote the OLS estimators of the regressions (1) and (2). Show that

$$\hat{\beta}_1^{ILS,1} = \frac{\hat{\pi}_1^Y}{\hat{\pi}_1^X}$$

is a consistent estimator of β_1. You may take as given that $\hat{\pi}_1^X \to_p \pi_1^X$ and $\hat{\pi}_1^Y \to_p \pi_1^Y$.

(iii) Using the coefficients on Z_2 instead of Z_1, we define

$$\hat{\beta}_1^{ILS,2} = \frac{\hat{\pi}_2^Y}{\hat{\pi}_2^X}.$$

Like $\hat{\beta}_1^{ILS,1}$, $\hat{\beta}_1^{ILS,2}$ is a consistent estimator of β_1. In some data, suppose we find $\hat{\beta}_1^{ILS,1} - \hat{\beta}_1^{ILS,2}$ is large. What might this indicate about our instruments?

Problem 3. (6 Points) Suppose we observe the i.i.d. random sample $\{(Y_i, X_i)\}_{i=1}^n$. Denote $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$, $\overline{Y}_n = n^{-1}\sum_{i=1}^n Y_i$, $\mu_X = \mathrm{E}[X_i]$ ($\mu_X \neq 0$) and $\mu_Y = \mathrm{E}[Y_i]$. We are interested in μ_Y/μ_X. Derive the asymptotic distribution of $\sqrt{n}\left(\overline{Y}_n/\overline{X}_n - \mu_Y/\mu_X\right)$. Hint: Write

$$\frac{\overline{Y}_n}{\overline{X}_n} = \frac{\overline{Y}_n}{\mu_X} \cdot \left(\frac{\mu_X}{\overline{X}_n} - 1\right) + \frac{\overline{Y}_n}{\mu_X}.$$

You may use the following result: $W_n \to_d N\left(0, \sigma^2\right)$ and $\theta_n \to_p 0$, then $\theta_n W_n \to_p 0$.

Problem 4. (15 Points) Suppose that the linear model

$$\mathrm{PS} = \beta_0 + \beta_1 \mathrm{Funds} + \beta_2 \mathrm{Risk} + U$$

satisfies $\mathrm{E}[U] = \mathrm{E}[U \cdot \mathrm{Funds}] = \mathrm{E}[U \cdot \mathrm{Risk}] = 0$. PS is the percentage of a person's savings invested in the stock market, Funds is the number of mutual funds that the person can choose from, and Risk is some measure of risk tolerance (larger Risk means the person has a higher tolerance for risk).

(i) If Funds and Risk are positively correlated, does the slope coefficient in the simple regression of PS on Funds overestimate or underestimate β_1, in large samples?

(ii) We are unable to observe Risk directly, but we have data on the amount of life insurance a worker has, Insurance. Assume that Insurance is noisy measure of Risk, Insurance = Risk + e, with $\mathrm{E}[e] = \mathrm{E}[\mathrm{Risk} \cdot e] = \mathrm{E}[\mathrm{Funds} \cdot e] = \mathrm{E}[eU] = 0$. Will the OLS estimate of the coefficient on Funds in a regression of PS on Funds and Insurance be a consistent estimate of β_1?

(iii) Suppose we also have data on how often a worker gambles, Gamble. Assume that Gamble is an independent noisy measure of Risk, Gamble = Risk + v, with $\mathrm{E}[v] = \mathrm{E}[vU] = \mathrm{E}[ve] = \mathrm{E}[\mathrm{Risk} \cdot v] = \mathrm{E}[\mathrm{Funds} \cdot v] = 0$. Explain how we can consistently estimate β_1 using our data on PS, Funds, Insurance, and Gamble.

Problem 5. (10 Points) A researcher has data on the following variables for 5,061 respondents in the US National Longitudinal Survey of Youth:

- MARRIED, marital status in 1994, defined to be 1 if the respondent was married with his/her spouse present and 0 otherwise (a man/woman may not be in marriage legally with his/her spouse);
- MALE, defined to be 1 if the respondent was male and 0 if female;
- AGE in 1994 (the range being 29-37);
- S, years of schooling, defined as highest grade completed, and
- ASVABC, score on a test of cognitive ability, scaled so as to have mean 50 and standard deviation 10.

She uses Probit analysis to regress MARRIED on the other variables. The sample means of the explanatory variables and their (average) marginal effects evaluated at the sample means are shown in the table.

```
. probit MARRIED MALE AGE S ASVABC

Probit estimates                                  Number of obs   =       5061
                                                  LR chi2(4)      =     229.78
                                                  Prob > chi2     =     0.0000
Log likelihood = -3286.1289                       Pseudo R2       =     0.0338

------------------------------------------------------------------------------
     MARRIED |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        MALE |  -.1215281    .036332    -3.34   0.001    -.1927375   -.0503188
         AGE |    .028571   .0081632     3.50   0.000     .0125715    .0445705
           S |  -.0017465     .00919    -0.19   0.849    -.0197587    .0162656
      ASVABC |   .0252911   .0022895    11.05   0.000     .0208038    .0297784
       _cons |  -1.816455   .2798724    -6.49   0.000    -2.364995   -1.267916
------------------------------------------------------------------------------
```

Variable	Mean	Marginal effect
MALE	0.4841	−0.0467
AGE	32.52	0.0110
S	13.31	−0.0007
ASVABC	48.94	0.0097

(i) Discuss the conclusions one may reach, given the Probit output and the table, comment on whether you think they are reasonable.

(ii) The researcher considers including CHILD, a dummy variable defined to be 1 if the respondent had children and 0 otherwise, as an explanatory variable. When she does this, its t-statistic is found to be 33.65 and its average marginal effect is found to be 0.5685. Discuss these findings.

Problem 6. (12 Points) Which of the following can and which cannot cause the usual OLS-based t-statistic to be invalid (that is, not to have the t distributions under H_0, even in large samples)? Explain briefly.

(i) Heteroskedasticity.

(ii) A sample correlation coefficient of 0.95 between two regressors.

(iii) Omitting an important explanatory variable.

Problem 7. (12 Points)

Dependent Variable: $\log(salary)$			
Independent Variables	(1)	(2)	(3)
$\log(sales)$	.224 (.027)	.158 (.040)	.188 (.040)
$\log(mktval)$	—	.112 (.050)	.100 (.049)
$profmarg$	—	−.0023 (.0022)	−.0022 (.0021)
$ceoten$	—	—	.0171 (.0055)
$comten$	—	—	−.0092 (.0033)
intercept	4.94 (0.20)	4.62 (0.25)	4.57 (0.25)
Observations	177	177	177
R-squared	.281	.304	.353

The variable $mktval$ is market value of the firm, $profmarg$ is profit as a percentage of sales, $ceoten$ is years as CEO with the current company, and $comten$ is total years with the company.

(i) Comment on the effect of $profmarg$ on CEO salary.

(ii) Does market value have a significant effect? Explain.

(iii) Interpret the coefficients on $ceoten$ and $comten$. Are these explanatory variables statistically significant?

Problem 8. (15 Points) Aggregate demand Q_D for a certain commodity is determined by its price P, aggregate income Y, and population, POP,

$$Q_D = \beta_1 + \beta_2 P + \beta_3 Y + \beta_4 POP + U^D$$

and aggregate supply is given by

$$Q_S = \alpha_1 + \alpha_2 P + U^S$$

where U_D and U_S are independently distributed error terms: U_D and U_S are independent from all other variables and they are also independent from each other. Remember that the quantity and the price are determined simultaneously in the equilibrium $Q_S = Q_D = Q$. We observe only the equilibrium values Q so that the observed price must satisfy the equation (demand = supply):

$$\beta_1 + \beta_2 P + \beta_3 Y + \beta_4 POP + U^D = \alpha_1 + \alpha_2 P + U^S.$$

(i) Show that the OLS (ordinary least squares) estimator of α_2 will be inconsistent if OLS is used to fit the supply equation.

(ii) Show that a consistent estimator of α_2 is

$$\widetilde{\alpha}_2 = \frac{\sum_{i=1}^n \left(Y_i - \overline{Y}\right)\left(Q_i - \overline{Q}\right)}{\sum_{i=1}^n \left(Y_i - \overline{Y}\right)\left(P_i - \overline{P}\right)}.$$

($\overline{Y} = n^{-1}\sum_{i=1}^n Y_i$, $\overline{Q} = n^{-1}\sum_{i=1}^n Q_i$, $\overline{P} = n^{-1}\sum_{i=1}^n P_i$.)

(iii) Explain how to construct a bootstrap percentile confidence interval for α_2, using $\widetilde{\alpha}_2$.