STA502: Math & Stat for MBA
Problem Set 6

Question 1. Suppose that annual earnings and marjiuana usage are determined jointly by

$$log(earnings) = \beta_0 + \beta_1 marijuana + \beta_2 educ + u_1$$
$$marijuana = \gamma_0 + \gamma_1 log(earnings) + \gamma_2 educ + \gamma_3 fine + \gamma_4 prison + u_2$$

where fine is the typical $fine$ assessed for people possessing small amounts of marijuana and $prison$ is a dummy variable equal to one if a person can serve prison time for being in possession of marijuana for personal use. Assume $fine$ and $prison$ can vary with the country of residence.

(a) If $educ$, $fine$, and $prison$ are exogenous, what do you need to assume about the parameters in the system to consistently estimate the β_j.

(b) Explain in detail how you would estimate the β_j assuming the parameters are identified.

(c) Do you have overidentification?

Solution.

1. Two exogenous variables $fine$ and $prison$ could be candidate of instrumental variables. In first equation, number of endogenous variable = $1(marijuana)$ < number of instrumental variable = 2. It's overidentification case.
But notice that second equation is unidentification case, since there are no exogenous variables appearing in the first equation that aren't also in the second equation.
To avoid unidentification in first equation, γ_3 or γ_4 should not be zero. We can also expect that the coefficients will be positive.

2. (1) 1st stage regression : Regress $marijuana$ on constant, $educ$, $fine$, $prison$ and get estimate $mari\hat{j}uana$
(2) 2nd stage regression : Regress $\log(earnings)$ on constant, $mari\hat{j}uana$ and $educ$ and get $\hat{\beta}_{2SLS}$
$\hat{\beta}_{2SLS}$ are biased, but consistent estimator.

3. First equation is overidentification, but second equation is unidentification.

Question 2. Let us consider the demand for fish. Using 97 daily price ($avgprc$) and quantity ($totqty$) observations on fish prices at the Fulton Fish Market in Manhattan, the following results were obtained by OLS.

$$log(\widehat{totqty_t}) = \underset{(.163)}{8.244} - \underset{(.176)}{.425}\, log(avgprc_t) - \underset{(.226)}{.311}\, mon_t - \underset{(.223)}{.683}\, tues_t - \underset{(.220)}{.533}\, wed_t + \underset{(.220)}{.067}\, thrus_t.$$

The equation allows demand to differ across the days of the week, and Friday is the excluded dummy variable. The standard errors are in parentheses.

(a) Interpret the coefficient of $\log(avgprc)$ and discuss whether it is significant.

(b) It is commonly thought that prices are jointly determined with quantity in equilibrium where demand equals supply. What are the consequences of this simultaneity for the properties of the OLS estimator? In your answer you may want to provide the system of equations that determine quantity and prices and demonstrate these consequences.

(c) The variables $wave2_t$ and $wave3_t$ are measures of ocean wave heights over the past several days. In view of your answer in (b), what two assumptions do we need to make in order to use $wave2_t$ and $wave3_t$ as instruments for $\log(avgprc_t)$ in estimating the demand equation? Discuss whether these assumptions are reasonable.

(d) Below we report two sets of regression results, where the dependent variable is $\log(avgprct)$. Are $wave2_t$ and $wave3_t$ jointly significant? State the test statistic and rejection rule. How is your finding related to your answer in (c)?

	Regressor						
Dependent Variable $log(avgprc_t)$	constant	wave2	wave3	day-of-the-week dummies	R^2	RSS	n
Regression (2.2)	-1.022 (.144)	-1.022 (.144)	-1.022 (.144)	yes	.3041	10.934	97
Regression (2.3)	$-.276$ (.092)	-	-	yes	.0088	15.576	97

(e) the following IV results were obtained in regression:

$$\widehat{log(totqty_t)} = \underset{(.182)}{8.164} - \underset{(.327)}{.815}\, log(avgprc_t) - \underset{(.229)}{.307}\, mon_t - \underset{(.226)}{.685}\, tues_t - \underset{(.223)}{.521}\, wed_t + \underset{(.225)}{.095}\, thrus_t.$$

Discuss how these results can be obtained using Two Stage Least Squares (2SLS).

Solution.

1. In demand equation, the coefficient sign of $\log(avgprc_t)$ should be negative. $\mathrm{T} = \frac{-0.425}{0.176} = -2.414 \sim t_{(91)}$. We could reject the null hypothesis under $\alpha = 0.05$.
   Interpretation : The estimated $avgprc$ elasticity of $totqty$ is -0.425 (or, if $avgprc$ increases 1%, then $totqty$ increases -0.425% on average)

2. If prices are "jointly" determined by demand and supply equation, the estimator will have simultaneity bias, which leads estimator to biased and inconsistent. $\log(avgprc)$ and $\log(totqty)$ turns to be endogenous.

3. If $wave2_t$ and $wave3_t$ are proper instrumental variables, they must satisfy validity(orthogonal with error) and relevance(correlated with $\log(avgprc)$ condition.

4. By F test, we could determine whether two instrumental variables have relevance with $\log(avgprc)$.
   H_0 : The variables $wave2_t$ and $wave3_t$ aren't significant both vs. H_1 : Not H_0
   $F = \frac{(15.576-10.934)/2}{10.934/(97-7)} = 18.37 \sim F_{(2,90)}$. Hence we can reject the null hypothesis at $\alpha = 0.05$.
   We could determine two instrumental variables have relevance, but since $R^2 = 0.3041$, they aren't highly relevant.

5. (1) 1st stage regression : Regress $\log(avgprc_t)$ on constant, $wave2_t$, $wave3_t$, day-of-the-week dummies and get $\log(\hat{avgprc_t})$
   (2) 2nd stage regression : Regress $\log(totqty_t)$ on constant, $\log(\hat{avgprc_t})$, day-of-the-week dummies and get $\hat{\beta}_{2SLS}$
   $\hat{\beta}_{2SLS}$ are biased, but consistent estimator.