Question 1. Suppose that annual earnings and marijuana 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 $fine$ assessessed 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
Since it is an SEM using IV, the following assumptions are required to obtain the estimator β_j in a consistancy. Large sample, Random sampling, and IV's basic requirements(validity, relevance).

(b) Explain in detail how you would estimate the β_j assuming the parameters are identified.
Using 2SLS,
Step 1: Get fitted values $\widehat{marijuana}$, by running on OLS on

$$\widehat{marijuana} = \pi_0 + \pi_1 educ + \pi_2 fine + \pi_3 prison + v$$

Step 2: Obtain the 2SLS estimates, by running OLS on

$$log(earnings) = \beta_0 + \beta_1 \widehat{marijuana} + \beta_2 educ + e$$

(c) Do you have overidentification?
Using $fine$ and $prison$ as IV of $log(earnings)$ in marijuana model, earnings model is overidentificated(IVs more than DV). However, marijuana model isn't identified.

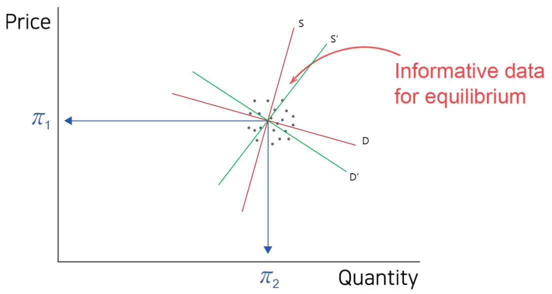
Question 2. Let us consider the demand for fish. Using 97 daily price ($avgprc$) and quantity ($totqty$) observations on fish prices at the Fulton Fish Market in Manhattan, the following results were obtained by OLS.

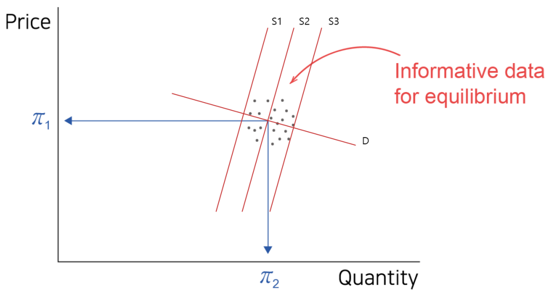
$$log(\widehat{totqty}_t) = \underset{(.163)}{8.244} - \underset{(.176)}{.425}\, log(avgprc_t) - \underset{(.226)}{.311}\, mon_t - \underset{(.223)}{.683}\, tues_t - \underset{(.220)}{.533}\, wed_t + \underset{(.220)}{.067}\, thurs_t$$

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.
First of all, t stat of log($avgprc$) is 2.42. it can be said to be significant. Of course, it would be good to test the significance with the hausman test or the t stat of the reduced form residuals.

(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.
As consequences of simultaneity like price-demand equilibrium, The real world data obtained are scattered around the equilibrium state. Therefore, OLS estimators will not be identifiable without IV.





Since our interest is in the demand, we need to set a supply model with IV to obtain estimators of the demand model.

$$log(\widehat{totqty}_t) = \gamma_0 + \gamma_1 log(avgprc_t) + \gamma_2 mon_t + \gamma_3 tues_t + \gamma_4 wed_t + \gamma_5 thurs_t + \gamma_6 z_t + u \text{ (Supply model)}$$

(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.

$log(\widehat{avgprc_t}) = \pi_0 + \pi_1 mon_t + \pi_2 tues_t + \pi_3 wed_t + \pi_4 thurs_t + \pi_5 wave2_t + \pi_6 wave3_t + v$ (Reduced form)

IVs($wave2$, $wave3$) have to satisfy condition of validity & relevance.

$$Cov(IVs, u) = 0 : \text{condition of validity}$$

$$Cov(log(avgprc), IVs) \neq 0 : \text{condition of relevance}$$

High wave heights can make the boat less stable, making it difficult to fish. Therefore, wave height can affect catch. The correlation between quantity and wave height seems reasonable. I don't think there is a direct relationship between the price and the height of the wave. $wave2$, $wave3$ are suitable as IV of $totqty$, not $avgprc$.

(d) Below we report two sets of regression results, where the dependent variable is $\log(avgprc_t)$. 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

It is weird that constant and wave2 and 3 all have the same estimate and sd. Anyway, t stat are greater than 2, so it can be said to be significant. It would be good to add a significance test as an F test for wave2, 3 presence.

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

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

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

The OLS obtained in (d) is the first step of 2SLS. If it is determined that wave2, 3 are significant, the DV($log(avgprc_t)$) calculated in the first step is added to the demand model to calculate the second step OLS.