1. (a) For β_j to be consistently estimated, $A3R_{sru}$ should be satisfied in the first equation. This means that $Cov(marijuana, u_1) = 0$ $Cov(educ, u_1) = 0$ should be satisfied.

However, it isn't due to simultaneity between marijuana and earnings.

Therefore, we use fine, prison as IVs.

IVs should satisfy validity and relevance.

① Validity: $Cov(fine, u_1) = Cov(prison, u_1) = 0$

② Relevance: $Cov(fine, marijuana) \neq 0$, $Cov(prison, marijuana) \neq 0$

The higher the relevance is, the higher the precision of $\hat{\beta}_j$s are.

(b) We can use 2SLS to estimate β_j.

① First stage: $marijuana = \pi_{m1} fine + \pi_{m2} prison + \delta educ + v$

From the first stage, we can obtain $\widehat{marijuana} = \hat{\pi}_{m1} fine + \hat{\pi}_{m2} prison + \hat{\delta} educ$.

② Second stage: $\log(earnings) = \beta_0 + \beta_1 \widehat{marijuana} + \beta_2 educ + u_1$

From the second stage, we can obtain $\hat{\beta}_{0,2SLS}$, $\hat{\beta}_{1,2SLS}$, $\hat{\beta}_{2,2SLS}$ consistently since $\widehat{marijuana}$ and educ are uncorrelated with u_1 and $A3R_{sru}$ is satisfied.

(c) Since we have 1 endogenous variable, marijuana, and 2 IV candidates, fine & prison, we have overidentification

2. (a) The coefficient of $\log(avgprc_t)$ means the price elasticity of demand, which is $\frac{\%\text{ change in totqty}}{\%\text{ change in avgprc}}$. It focuses on the ratio of change "rate" on average.

Since t-statistic of the coefficient is about -2.5 and its magnitude is larger than 2, we can think $\log(avgprc)$ is significant. In addition, the negative sign of the coefficient corresponds with our intuition that quantity for demand decreases as average price increases.

(b) Simultaneity is one of a kind of endogeneity and breaks A3R condition. This can lead to biasedness & inconsistency of coefficients. In this case, $\log(avgprc_t)$ is an endogenous variable. System of equations are described as below.

$$\begin{cases} \log(totqty_t) = \alpha_1 + \alpha_2 \log(avgprc_t) + \alpha_3 mon_t + \alpha_4 tues_t + \alpha_5 wed_t + \alpha_6 thurs_t + u_{dt} & \text{- demand} \\ \log(totqty_t) = \beta_1 + \beta_2 \log(avgprc_t) + u_{st} & \text{- Supply} \end{cases}$$

Since the demand equation does not satisfy order condition, we cannot identify it now. We should introduce exogenous variables included only in supply equation.

(c) To use $wave2_t$ & $wave3_t$ as IVs, those variables should satisfy validity and relevance

① Validity: $Cov(wave2_t, u_{st}) = Cov(wave2_t, u_{dt}) = Cov(wave3_t, u_{st}) = Cov(wave3_t, u_{dt}) = 0$

② Relevance: $\begin{cases} Cov(wave2_t, \log(avgprc_t)) \neq 0 \\ Cov(wave3_t, \log(avgprc_t)) \neq 0 \end{cases}$

$wave2_t$ and $wave3_t$ should not be in demand equation

Wave heights would affect only the quantity for supply, not the quantity for demand because large wave heights for past two or three days make fishermen not be able to go to sea, decreasing the quantity for supply and increasing in daily price.

Therefore, the assumptions seem reasonable.

The structural equations are changed as below.

$$\begin{cases} \log(totqty_t) = \alpha_1 + \alpha_2 \log(avgprc_t) + \alpha_3 mon_t + \alpha_4 tues_t + \alpha_5 wed_t + \alpha_6 thurs_t + u_{dt} & \text{- demand} \\ \log(totqty_t) = \beta_1 + \beta_2 \log(avgprc_t) + \beta_3 wave2_t + \beta_4 wave3_t + u_{st} & \text{- Supply} \end{cases}$$

To obtain αs consistently, we would use 2SLS method

① First stage

$$\log(avgprc_t) = \pi_0 + \pi_1 wave2_t + \pi_2 wave3_t + \pi_3 mon_t + \pi_4 tues_t + \pi_5 wed_t + \pi_6 thurs_t + e_t$$

$$\rightarrow \widehat{\log(avgprc_t)} = \hat{\pi}_0 + \hat{\pi}_1 wave2_t + \hat{\pi}_2 wave3_t + \hat{\pi}_3 mon_t + \hat{\pi}_4 tues_t + \hat{\pi}_5 wed_t + \hat{\pi}_6 thurs_t$$

② Second stage

$$\log(totqty_t) = \alpha_1 + \alpha_2 \widehat{\log(avgprc_t)} + \alpha_3 mon_t + \alpha_4 tues_t + \alpha_5 wed_t + \alpha_6 thurs_t + u^*_{dt}$$

(d) t-statistic of $wave2_t = -\frac{1.022}{0.144} < -2$

t-statistic of $wave3_t = -\frac{1.022}{0.144} < -2$

$$\text{F-statistic} = \frac{(RRSS-URSS)/df}{URSS/n-p} = \frac{(15.576-10.934)/2}{10.934/(97-7)} = 19.1 \sim F(2,90)$$

$H_0: \pi_1 = \pi_2 = 0$ H_A: not H_0

Therefore, we can reject H_0 and conclude that $wave2_t$ & $wave3_t$ are significant both in t-tests & F-test.

This result corresponds to our assumptions that $wave2_t$ & $wave3_t$ have relevance with $\log(avgprc_t)$

(e) As I mentioned in (c), 2SLS method is implemented as below.

① First stage

$$\log(avgprc_t) = \pi_0 + \pi_1 wave2_t + \pi_2 wave3_t + \pi_3 mon_t + \pi_4 tues_t + \pi_5 wed_t + \pi_6 thurs_t + e_t$$

$$\rightarrow \widehat{\log(avgprc_t)} = \hat{\pi}_0 + \hat{\pi}_1 wave2_t + \hat{\pi}_2 wave3_t + \hat{\pi}_3 mon_t + \hat{\pi}_4 tues_t + \hat{\pi}_5 wed_t + \hat{\pi}_6 thurs_t$$

② Second stage

$$\log(totqty_t) = \alpha_1 + \alpha_2 \widehat{\log(avgprc_t)} + \alpha_3 mon_t + \alpha_4 tues_t + \alpha_5 wed_t + \alpha_6 thurs_t + u^*_{dt}$$

As a result, α_js $(j=1,2,3,4,5,6)$ would be same as IV estimates without standard errors. We should use robust standard errors when we use 2SLS method.