Solutions to Problem Set 4 – KAN-CCMVV2401U

Part 1

1 (a) Assuming the structural equation represents a causal relationship, $100\cdot\beta_1$ is the approximate percentage change in income if a person smokes one more cigarette per day.

(b) Since consumption and price are, ceteris paribus, negatively related, we expect $\gamma_5 \leq 0$ (allowing for γ_5) = 0. Similarly, everything else equal, restaurant smoking restrictions should reduce cigarette smoking, so $\gamma_5 \leq 0$.

(c) We need γ_5 or γ_6 to be different from zero. That is, we need at least one exogenous variable in the *cigs* equation that is not also in the log(*income*) equation.

OLS estimation of the log(*income*) equation gives

$$\begin{array}{llllll} \log(income) & = 7.80 & +.0017\ cigs & +.060\ educ & +.058\ age & -.00063\ age^2 \\ & (0.17) & (.0017) & (.008) & (.008) & (.00008) \end{array}$$

$n = 807, \quad R^2 = .165.$

The coefficient on *cigs* implies that cigarette smoking causes income to increase, although the coefficient is not statistically different from zero. Remember, OLS ignores potential simultaneity between income and cigarette smoking.

(d) The estimated reduced form for *cigs* is

$$\begin{array}{llllll} cigs & = 1.58 & -.450\ educ & +.823\ age & -.0096\ age^2 & -.351\ \log(cigpric) \\ & (23.70) & (.162) & (.154) & (.0017) & (5.766) \\ & & -2.74\ restaurn & & & \\ & & (1.11) & & & \end{array}$$

$n = 807, \quad R^2 = .051.$

While log(*cigpric*) is very insignificant, *restaurn* had the expected negative sign and a *t* statistic of about –2.47. (People living in states with restaurant smoking restrictions smoke almost three fewer cigarettes, on average, given education and age.) We could drop log(*cigpric*) from the analysis but we leave it in. (Incidentally, the *F* test for joint significance of log(*cigpric*) and *restaurn* yields *p*-value $\approx$ .044.)

(e) Estimating the log(*income*) equation by 2SLS gives

$$\begin{array}{llllll} \log(income) & = 7.78 & -.042\ cigs & +.040\ educ & +.094\ age & -.00105\ age^2 \\ & (0.23) & (.026) & (.016) & (.023) & (.00027) \end{array}$$

$n = 807.$

Now the coefficient on *cigs* is negative and almost significant at the 10% level against a two-sided alternative. The estimated effect is very large: each additional cigarette someone smokes lowers predicted income by about 4.2%. Of course, the 95% CI for β_{cigs} is very wide.

(f) Assuming that state level cigarette prices and restaurant smoking restrictions are exogenous in the income equation is problematical. Incomes are known to vary by region, as do restaurant smoking restrictions. It could be that in states where income is lower (after controlling for education and age), restaurant smoking restrictions are less likely to be in place.

2

(a) The demand function should be downward sloping, so $\alpha_1 < 0$: as price increases, quantity demanded for air travel decreases.

(b) The estimated price elasticity is −.391 (t statistic = −5.82).

(c) We must assume that passenger demand depends only on air fare, so that, once price is controlled for, passengers are indifferent about the fraction of travel accounted for by the largest carrier.

(d) The reduced form equation for log(*fare*) is

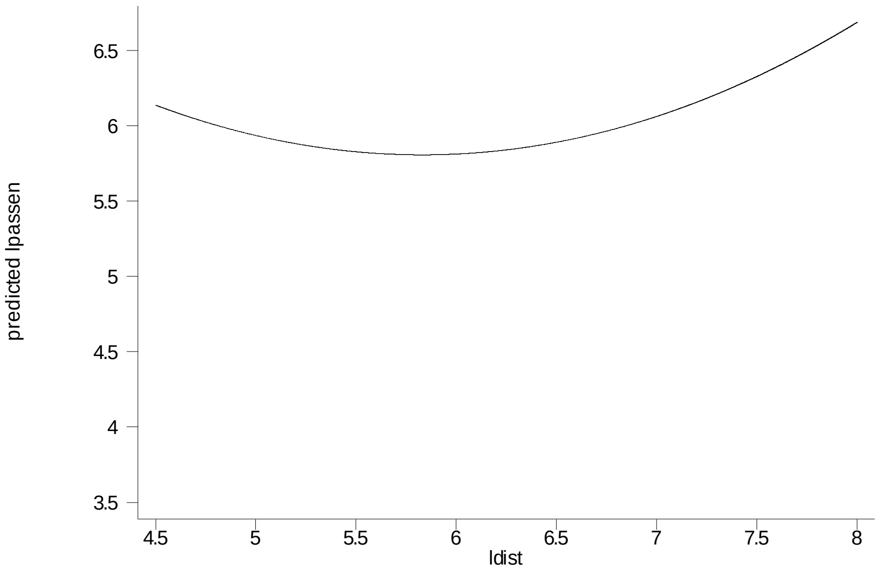
$$\widehat{\log(fare)} = \underset{(0.89)}{6.19} + \underset{(.063)}{.395}\, concen - \underset{(.272)}{.936}\log(dist) + \underset{(.021)}{.108}\,[\log(dist)]^2$$

n = 1,149, R^2 = .408

The coefficient on *concen* shows a pretty strong link between concentration and fare. If *concen* increases by .10 (10 percentage points), *fare* is estimated to increase by almost 4%. The t statistic is about 6.3.

(e) Using *concen* as an IV for log(*fare*) [and where the distance variables act as their own IVs], the estimated price elasticity is −1.17, which shows much greater price sensitivity than did the OLS estimate. The IV estimate suggests that a one percent increase in fare leads to a slightly more than one percent increase drop in passenger demand. Of course, the standard error of the IV estimate is much larger (about .389 compared with the OLS standard error of .067), but the IV estimate is statistically significant (t is about −3.0).

(f) The relationship between log(*fare*) and log(*dist*) has a U-shape, as given in the following graph:



The minimum is at about *ldist* = log(*dist*) = 5.82 (which, in terms of distance, is about 340 miles. About 11% of the routes are less than 336 miles long. If the estimated quadratic is believable, the lowest demand occurs for short, but not very short, routes (holding price fixed). It is possible, of course, that we should ignore the quadratic to the left of the turning point, but it does contain a nontrivial fraction of the observations.

Part 2

3 (a) Other things equal, homes farther from the incinerator should be worth more, so $\delta_1 > 0$. If $\beta_1 > 0$, then the incinerator was located farther away from more expensive homes.

(b) The estimated equation is

$$\begin{array}{lllll} \log(\widehat{price}) & = 8.06 & -.011\, y81 & +.317 \log(dist) & +.048\, y81 \cdot \log(dist) \\ & (0.51) & (.805) & (.052) & (.082) \end{array}$$

$$n = 321, \quad R^2 = .396, \quad \bar{R}^2 = .390.$$

While $\hat{\delta}_1$ = .048 is the expected sign, it is not statistically significant (t statistic ≈ .59).

(c) When we add the list of housing characteristics to the regression, the coefficient on $y81 \cdot \log(dist)$ becomes .062 (se = .050). So the estimated effect is larger – the elasticity of *price* with respect to *dist* is .062 after the incinerator site was chosen – but its t statistic is only 1.24. The p-value for the one-sided alternative H_1: $\delta_1 > 0$ is about .108, which is close to being significant at the 10% level.