

Econometric Methods

Submit your answer in Word (or other word processor) as well as an R executable script.

1. Do more generous workers' compensation cause people to stay out of work longer?

Use the data in *injury* for this exercise. In the early 1980s, Kentucky and Michigan raised the cap on weekly earnings that were covered by workers' compensation. An increase in the cap has no effect on the benefit for low-income workers, but it makes it less costly for a high-income worker to stay on workers' compensation. Using random samples both before and after the policy change, [Meyer, Viscusi, and Durbin \(1995, AER\)](#) tested whether more generous workers' compensation may cause people to stay out of longer (everything else fixed).

(a) Take the control group as low-income workers, and the treatment group as high-income workers.

Estimate the model below for the state of Kentucky.

$$\log(\widehat{durat}) = \beta_0 + \delta_0 afchnge + \beta_1 highearn + \delta_1 afchnge * highearn$$

where *durat* is duration of benefits (measured by weeks of temporary total benefits paid plus anticipated future weeks paid if the claim is still open), *afchnge* is the dummy variable for observations after the policy change and *highearn* is the dummy variable for high earners.

Report your result and interpret the estimate on the interaction term.

(b) If the variable *afchnge* is dropped in the model, does your result for $\widehat{\delta}_1$ change much? Why? What if *afchnge* is included but *highearn* is dropped?

(c) Now re-estimate the model by adding as explanatory variables *male*, *married*, a set of industry (*manuf*, *construc*), and injury type dummy variables (*head*, *neck*, *upextr*, *trunk*, *lowback*, *lowextr*, and *occdis*). How does the estimate on *afchnge***highearn* change when these other factors are controlled for? Is the estimate still significant?

(d) What do you make of the small R-squared from part (c)? Does this mean the equation is useless?

(e) Estimate the same equation in Part (a) but now for the state of Michigan. Compare the estimates on the interaction term for Michigan and Kentucky. Is the Michigan estimate statistically significant? What do you make of it?

2. How does market concentration affect airfare?

Use the data in *airfare* for this exercise. This is a panel data set that contains annual data on airfare in different routes in the US for the years 1997 to 2000. The unit of observation is the route and it is identified in the data set by the variable *id*; years are identified in the data set by the variable *year*.

We are interested in the following population model:

$$\log(\text{fare}_{it}) = \alpha_t + \beta_1 \text{concen}_{it} + \beta_2 \log(\text{dist}_i) + \beta_3 [\log(\text{dist}_i)]^2 + a_i + u_{it},$$

where i indexes routes, t indexes years ($t=1, \dots, 4$), *concen* is a measure of market concentration, *dist* is the length (distance between stops) of the route, α_t indicates aggregate year effects, and a_i denotes a route-specific unobserved effect.

(a) Is this a balanced panel? Estimate the above model by pooled OLS, being sure to include year dummy variables. [Hint: use “factor” to categorize or create year dummies, see the R script for April 7 class]. If $\Delta \text{concen} = 0.1$, what is the estimated percentage increase in fare?

(b) What is the usual OLS 95% confidence interval for β_1 ? Find the fully robust 95% confidence interval for β_1 by using “clustered” standard errors. Compare it to the usual Confidence interval and which one is more appropriate? Why?

(c) Describe what is happening with the quadratic in log-transformed distance, $[\log(\text{dist}_i)]^2$. In particular, for what value of *dist* does the relationship between $\log(\text{fare})$ and *dist* become positive? (hint: turning point.)

(d) Now estimate the model using the random effects estimator. How does the estimate of β_1 change?

(e) Now estimate the model using the fixed effects estimator. How does the estimate of β_1 change? Why is it fairly similar to the random effects estimate?

(f) Name two characteristics of a route (other than distance between stops) that are captured by a_i ? Might these be correlated with *concen*?

(g) Add first difference approach for the analysis too. In general, are you convinced that higher concentration on a route increase airfare?