

## Labor Supply of Taxi Drivers:

### Camerer, Babcock, Lowenstein, and Thaler, QJE 97

Recall from the intertemporal model, first order conditions are:

$$u_c(C_t, H_t, X_t) = \lambda_t$$

$$u_H(C_t, H_t, X_t) = -w_t \lambda_t$$

$$\lambda_t = \beta(1 + r_{t+1})E(\lambda_{t+1})$$

where  $\lambda_t$  is the marginal utility of (expected) wealth at time t.

Can use first order conditions to solve implicitly:

$$C_t = C(\lambda_t, w_t, X_t)$$

$$H_t = H(\lambda_t, w_t, X_t)$$

A log linear version of this is:

$$\log H_t = A_t + \eta \log w_t + \gamma \log \lambda_t$$

$$\Delta \log H_t = \Delta A_t + \eta \Delta \log w_t + \gamma [\log(\lambda_t) - \log(\lambda_{t-1})]$$

From first order conditions:

$$\lambda_{t-1} = \beta(1 + r)E_{t-1}(\lambda_t)$$

$$\log(\lambda_t) - \log(\lambda_{t-1}) = \log(\lambda_t) - \log \beta(1 + r) - \log E_{t-1}(\lambda_t)$$

so:

$$\Delta \log H_t = \Delta A_t + \eta \Delta \log w_t - \gamma \log \beta(1 + r) + \gamma [\log(\lambda_t) - \log E_{t-1}(\lambda_t)]$$

$$\Delta \log H_t = \Delta A_t + \eta \Delta \log w_t + \gamma(\rho - r) + \gamma [\log(\lambda_t) - \log E_{t-1}(\lambda_t)]$$

if  $\beta = 1/(1 + \rho)$  and  $\log(1 + x) \approx x$ .

The time preference component does not vanish if  $\rho$  and  $r$  are not equal. If relatively impatient, work more and consume less over time. Vice versa.

We're not quite done yet because  $\log(\lambda_t) - \log E_{t-1}(\lambda_t)$  is not the forecast error in the marginal utility of wealth. Let's define that as  $\varphi_t$ , where

$$\log \lambda_t = E_{t-1} \log \lambda_{t-1} + \varphi_t$$

If  $\varphi_t$  is log normal,  $\log \lambda_t = \log E_{t-1} \lambda_{t-1} + \varphi_t$  and we're done. If not,

$$\lambda_t = \exp[E_{t-1} \log \lambda_{t-1}] \exp[\varphi_t]$$

$$E_{t-1} \lambda_t = \exp[E_{t-1} \log \lambda_{t-1}] E_{t-1} \exp[\varphi_t]$$

$$\log(E_{t-1} \lambda_t) = E_{t-1} \log \lambda_{t-1} + \log E_{t-1} \exp[\varphi_t]$$

subbing in:

$$\Delta \log H_t = \Delta A_t + \eta \Delta \log w_t + \gamma(\rho - r) + \gamma \varphi_t - \gamma \log E_{t-1} \exp[\varphi_t]$$

Estimating  $\eta$  has proven difficult. Of those that are out there, estimates have generally been low and insignificant, but estimation difficulties and the possibility that the wage shocks were more permanent have led to doubt in these results.

Many jobs do not allow much flexibility in labor supply. To adjust labor supply most must change jobs. But a number of recent papers have used data where workers can control their hours and are subject to transitory shocks.

One of those jobs is the taxi driver. Colin Camerer et al. suggest New York cab drivers have good days and bad days. On good days, there's a big conference going on, the weather is bad, subway breakdown, or some other factor leads to an increase in the demand for cabs. Other days there's no action. Rates are the same for every cabbie, but on busier days, the driver spends less time searching for customers and thus earns a higher hourly wage.

So plausibly, if we can measure good and bad days, we can use this potentially exogenous variation as capturing wage shocks and estimate labor supply responses. One beauty of this approach is that it shouldn't matter whether the wage changes are expected or unexpected, because in either case, the change in wealth from a change in daily wage is zero or close to zero.

The data come from 'trip sheets' of New York City cabdrivers. Colin Camerer once drove a cab in new york for a summer, so I presume this helped obtaining the data. Farber had a much harder time. A trip sheet records the list of trips that drivers took on a given day, including the fare (not including tip) and time picked up and dropped off. 12 hour shifts.

3 trip sheet samples

1) 70 cleaned up trip sheets in 1994 from a fleet company that rent the cabs out

The fleet company rents out the caps for about \$80 per 12 hour shift + gas. Driver can return cab anytime within the 12 hour period. 8 drivers on more than one day: allows the authors to create a panel dataset of labor supply and wages.

Smaller of the 3 datasets, but contains info within days.

hours worked is between first passenger and last passenger dropped off.

Average hourly wage is total revenue dived by hours worked.

Important to show that days that start off as good, stay good. So cabbies know they face a positive, unexpected shock, (or negative). The trip sheet data allows the authors to look at the hourly earnings autocorrelation within days

They calculated median hourly wage for all drivers for each hour, and regressed

$$w_{th} = \beta_0 + \beta_1 w_{th-1} + e_{th}$$

b1hat is .5. Correlation between first half and last half is .4.

Also, wages across days appear uncorrelated:

$$w_t = \beta_0 + \beta_1 w_{t-1} + e_t,$$

b1hat is -.07

Thus, wages virtually uncorrelated across days and fairly stable within days.

Datasets 2 and 3 have larger panel, but no info on hour to hour.

Sample characteristics: table 1

Average hours worked, = 9, stdev 1.4,

Main regression is log daily hours worked on log hourly wage:

$$\log H_{it} = \beta X_t + \delta \log \frac{W_{it}}{H_{it}} + e_i + e_{it}$$

where  $H_{it}$  is hours worked for  $i$  in day  $t$ .  $W_{it}$  is total revenue in the day.

Any problems with this? The dependent variable is included in the right hand side! This is not a problem if hours in the day is measured correctly, because then we measure hourly wage correctly and this is the relationship we want to measure. But suppose instead we use:

$$H_{it}^* = H_{it} + u_{it},$$

in the regression, where  $H_{it}$  and  $u_{it}$  is measurement error that is uncorrelated with  $H_{it}$  (classical measurement error). Now whenever daily hours are overestimated, average wages is underestimated, and vice versa. This downward biases the coefficient estimate for  $\delta$ . This is clearly a problem. There may be recording errors on the trip sheet, or time spend talking with friends or lunch may be incorrectly counted in  $H_{it}$ .

Measurement error of independent variables is a general problem as well. If the independent variable of interest has classical measurement error, the OLS coefficient estimate will be downward biased.

Consider first the single variable case:

Suppose that the true model of schooling is:

$$H_i = \beta_0 + \beta_1 W_i^* + e_i,$$

but we use another measure of schooling that is inaccurate:  $W_i = W_i^* + u_i$

Assuming no omitted variables problem, our least squares estimate for  $\beta_1$  is:

$$\begin{aligned} \frac{\text{cov}(W_i, H_i)}{\text{var}(W_i)} &= \frac{\text{cov}(W_i, \beta_0 + \beta_1 W_i^* + e_i)}{\text{var}(W_i)} \\ &= \beta_1 \frac{\text{cov}(W_i, W_i^*)}{\text{var}(W_i)} \\ &= \beta_1 \frac{\text{cov}(W_i^* + u_i, W_i^*)}{\text{var}(W_i^* + u_i)} \\ &= \beta_1 \frac{\text{var}(W_i^*)}{\text{var}(W_i^*) + \text{var}(u_i)} \\ &= \beta_1 R_{W^*, W}^2 \end{aligned}$$

where  $R_{W^*,W}^2$  is the correlation coefficient between  $H^*$  and  $H$ . If the correlation is not 1 (they are not identical),  $B_1$  will be biased downward.

The attenuation bias is towards zero. This is different from the measurement error bias in Camerer's paper because the the m.e. is uncorrelated with the dependent variable.

For the multivariate case:

$$H_i = \beta_0 + \beta_1 W_i + \beta_2 X_i + e_i$$

the attenuation bias will be:

$$p \lim \hat{\beta}_1 = \beta_1 \frac{\text{var}(r_i^*)}{\text{var}(r_i^*) + \text{var}(u_i)}$$

where  $r_i^*$  is residual from the linear projection:

$$W_i^* = \delta_0 + \delta_1 X_i + r_i^*$$

The more collinear  $W_i^*$  with other explanatory variables, the worse the bias (Woodridge's grad text gives more discussion). The coefficient for  $B_2$  will also be biased, but in an unknown direction. It is difficult to determine the direction and size of the bias if  $X$  and  $W$  have measurement error (see Greene).

If we have classical measurement error in the dependent variable (on the left hand side), there is no bias in the OLS estimate. This can easily be verified with a similar approach above.

A general solution to measurement error concerns is to instrument the variable with another variable, even if that variable itself has measurement error. Lagged variables often work, but in our case it's not good because we're trying to measure day to day fluctuations in wages.

Instead, Camerer et al. instrument with cabbie's daily average revenue with the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentile daily revenue for other cabbies. My own concern with this is if 'leisure time while on the job' is correlated across other cabbies. Authors suggest removing breaks more than 30 minutes in TRIP data lead to similar results.

What might be going on? Authors suggest 'targeting': quite when income target reached. Phone survey of 14 owners and managers. Asked to choose one of three sentences that best describe how many hours cab drivers drive each day. 6 said drive until they make certain amount of money, 5 said fixed hours, 1 said work more when doing well, quite early on a bad day.

In the context of the intertemporal model, targeting can be described by big drop in marginal utility of income after the target.

Farber: Is tomorrow another day?

Farber criticizes this approach.

- 1) Instrument of daily wage is invalid if measurement error in hours day specific across all drivers.
- 2) Farber's cab data shows no correlation in hourly wages within days, suggesting no ability to determine 'good days and bad days' from early on. He does not attempt to reconcile this opposite finding with Camerer et al. The two opposite findings are perplexing.

Variation in average hourly wages across days, therefore, is not likely interpreted as transitory fluctuations. More likely random. Such wage fluctuations are not likely predictable and are not associated with the intertemporal model. Farber finds similar negative wage coefficients with his data but attributes them likely to measurement error.

Farber instead tries to test the intertemporal model, relative to the target theory with predictions about quit decisions within a day.

The budget constraint is modified in the intertemporal model:

$$\bar{W}_{t+1} = (1+r)(\bar{W}_t + y_t(H_t) - C_t),$$

where  $y_t(H_t)$  represents daily earnings as a function of hours worked. Wages are allowed to vary throughout the day. This leads to more flexible first order conditions:

$$\frac{-u_h}{u_c} = y_t'(H_t)$$

Within a one day setting, income effects are likely to be small, so smoothed consumption is not likely to change from daily labor supply decisions. If this is approximately true, we can assign a constant to marginal utility of consumption. Let's normalize it to one.

$$-u_h(H_t) = y_t'(H_t)$$

If we consider changes to daily income, we get the usual substitution effect: when daily income is high, a cabbie should work longer hours, whether the higher wage is anticipated or transitory.

But this is not what Farber wants to test. He wants to consider labor supply decisions within a day, with the idea that  $y_t'$  fluctuates throughout the day. After every ride, for example, a cabbie makes a decision of whether to stop or not. The intertemporal model emphasizes that the quit rate will be mostly a function of hours worked, while the target model emphasizes it mostly depends on cumulative income.

Prediction 1:

Intertemporal model

Conditional on cumulated income, marginal disutility of working is increasing every hour. Hours worked will be positively related to quitting. It shouldn't matter whether we condition on income or not, but This is what Farber does to contrast it with the Target model.

#### Target model

It doesn't matter how many hours worked, only whether income target achieved. Conditional on cumulated income, marginal disutility of working is constant over hours. Working 7 or 10 hours does not determine whether a cabbie quits.

#### Prediction 2:

#### Intertemporal model

Conditional on wage opportunities later in the day, or if previous early shift wage fluctuations cannot be used to predict future wage fluctuations, then cumulated income will not be related to quit decisions.  $y'_i(H_i)$  only depends on current and future opportunities in the day. It is not clear whether Farber assumes he adequately controls for future opportunities in his regressions or he relies on his data that shows early shift wage changes are poor predictors of later shift changes, but he needs on of them to make this claim.

#### Target model

Cumulated income determines the decision to quit. So it will clearly be positively related to the quit rate.

Thus, Faber focuses on testing whether cumulated income in the day predicts quit rates better or worse than cumulated hours.

#### Farber's data:

After much inquiry, obtains 584 trip sheets for 21 drivers over various days throughout a 2 year period. Not much overlap in days and drivers, which limits the creation of an instrument for daily wages like Camerer et al. do.

Data issues: cannot match self reported trip sheets with printed meter. When you read the clean up procedure, you get an sense that the sheets were very messy (e.g. a ride ending before it started).

Note, problem if participation itself is correlated with the hours worked, had the cabbie instead decided to work. Individual fixed effects help mitigate this (cabbies less likely to work and less likely to work many hours controlled for). In addition, cabbies are supposed to work on scheduled days and may have to pay a fee if they do not.

Table 1:

Costs \$575 a week to rent cab + fuel. Tips are not included in the table. R-squared from regressing daily income on driver fixed effects is .169. Under target model, we might expect this to be higher. Check out driver 15. Hard life being a cabbie.

Table 3: no autocorrelation, opposite of what camerer et al find. “predicting hours of work with a model that assumes a fixed hourly wage rate during the day does not seem appropriate’

Test target model directly:

$$T_{it} = X_{it}\beta + e_i + e_{it},$$

where T is the average between total shift income and total shift income minus the last fare.

Table 5 shows Rsquared is low, not consistent with target model.

Table 7 main results. After including fixed effects hours significant, income insignificant.

For flexible form in table 8. Hours matter to quid decisions, income doesn't when conditioning on both.

Discussion: what do you think of the two papers?

Cluster Analysis: both papers cluster by driver. What does this mean?

### **A note on the need to 'cluster' standard errors**

OLS assumes no serial correlation or autocorrelation in the error terms when estimating the variance (and standard errors) of the coefficients. This can lead to downward bias in the standard errors if, instead, the errors, or at least some of them, are positively correlated. The bias can sometimes be severe.

Consider the variance of the ordinary least squares regression. Let  $\mathbf{X}$  be the  $n \times p$  design matrix and  $\mathbf{y}$  be the  $n \times 1$  vector of dependent values. The regression model is:  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ , so any fixed effects are defined as dummy variables contained in the  $\mathbf{X}$  matrix, and  $\mathbf{y}$  and  $\mathbf{X}$  are deviations from their means. The ordinary linear regression estimates are  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ , and the variance is:

$$\text{var}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[(\mathbf{y} - E\mathbf{y})(\mathbf{y} - E\mathbf{y})']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

$$\text{var}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

where  $E(\varepsilon_i) = 0$  and  $E(\varepsilon_i\varepsilon_j) = \Omega$ , (the variance-covariance matrix for all  $i$  and  $j$  observations)

The standard OLS assumption to estimate the variance is  $\boldsymbol{\Omega} = \sigma^2\mathbf{I}$ , and  $\hat{\sigma}^2 = \frac{1}{N} \sum_1^N e_i^2$ :

$$\text{var}(\mathbf{b}) = \hat{\sigma}_\varepsilon^2 (\mathbf{X}'\mathbf{X})^{-1}$$

OLS assumes that the variance matrix for the error term is diagonal while in practice it might be block diagonal, with a constant correlation coefficient within each group and time cell. When we want to identify an aggregate group/time effect, within group/time correlation can be substantial. In practice, the correlation is often positive, which leads the OLS results to underestimate the standard error, making it more likely to reject the null hypothesis. It is reasonable to expect that units sharing observable characteristics such as being from the same industry, state, marital status, time period and location, also share unobservable characteristics that would lead the regression disturbances to be positively correlated. With Monte Carlo experiments, several recent papers have suggested using OLS standard error estimates can bias standard errors downwards and lead to rejection that the coefficient is zero, when in fact, it is.

Here, the possibility that  $e_{it}$  is correlated across time is clear: unobservable driver factors may be related to the outcomes, and this may occur whether or not those unobservables are correlated with the independent variable of interest (the issue arises regardless of omitted variables bias).

Fortunately, White (and earlier Eicher and Huber) found a way to estimate robust standard errors, regardless of the form  $\Omega$  takes (provided that  $\Omega$  is well defined). White pointed out that we do not need to estimate every component in the  $n \times n$   $\Omega$  matrix, an obviously impossible task when only  $n$  observations are available. But this way of looking at the problem is misleading. What is actually required is to estimate

$$\text{var}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1} E[\mathbf{X}'\mathbf{e}\mathbf{e}'\mathbf{X}](\mathbf{X}'\mathbf{X})^{-1}$$

(White, 84, Asymptotic Theory for econometricians)

The robust variance-covariance matrix estimator is:

$$\text{var}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1} \left( \sum_1^N [(y_i - \hat{y}_i)\mathbf{x}_i]'[(y_i - \hat{y}_i)\mathbf{x}_i] \right) (\mathbf{X}'\mathbf{X})^{-1}$$

where  $\hat{y}_i$  is the estimated error term, and the sum is over all observations. This variance is computed when the 'robust' option is specified in STATA. When prior knowledge leads the researcher to believe the error terms may be serially correlated within groups, but independent across groups, the variance can be calculated as:

$$\text{var}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1} \left( \sum_1^G u_k' u_k \right) (\mathbf{X}'\mathbf{X})^{-1}, \text{ where } u_k = \sum_{j \in k} [(y_j - \hat{y}_j) \mathbf{x}_j]' [(y_j - \hat{y}_j) \mathbf{x}_j]$$

This variance estimate is computed with STATA's 'cluster' command, specifying groups G.

This estimator is consistent for any arbitrary heteroskedasticity or serial correlation, but it is not efficient when prior information about the form of the matrix is known.

To give you a little intuition for the need to cluster, consider the following example. Suppose we are evaluating the relationship between education attainment and state compulsory school laws. Let  $S_{is}$  be years of schooling for individual  $i$  in state  $S$ , and  $Z_S$  is the dropout age that an individual faced when in high school, from state  $S$ . So the independent variable is the same for everyone from that state. The OLS regression equation is:

$$S_{is} = \beta Z_S + e_{is},$$

It's certainly plausible that individuals from the same state are related in other ways. There could still be no omitted variables bias:  $E(Z_S, e_{is}) = 0$ , but the error terms are serially correlated among individuals from the same state:  $E(e_{is}, e_{js} | S = \bar{S}) \neq 0$ .

One extreme example is we have 100 individuals, 2 from each state.  $Z_S$  is the same for each two individuals from the same state. Suppose also that  $S_{is}$  is the same for both. So what we have is 2 sets of the same 50 values for  $S$  and  $Z$ . Normalize the standard deviation to 1:  $E(e_{is}^2) = 1$ . If the variance-covariance matrix is  $\Omega = \mathbf{I}$ , as in OLS, the variance of  $\hat{\beta}$  is:

$$\text{var}(\hat{\beta}) = \frac{1}{2 \sum_1^{50} Z_i^2} 2 \sum_1^{50} Z_i^2 \frac{1}{2 \sum_1^{50} Z_i^2} = \frac{1}{2 \sum_1^{50} Z_i^2}$$

If, instead,  $e_{iS}$  is perfectly correlated within state,  $E(e_{iS}, e_{jS} | S = \bar{S}) = 1$  and zero otherwise.

Recognizing this, the true variance of  $\hat{\beta}$  is

$$\text{var}(\hat{\beta}) = \frac{1}{2 \sum_1^{50} Z_i^2} 4 \sum_1^{50} Z_i^2 \frac{1}{2 \sum_1^{50} Z_i^2} = \frac{1}{\sum_1^{50} Z_i^2}$$

If the second covariance matrix is correct, we falsely underestimate the variance of  $\hat{\beta}$  using OLS. The second individual in each state adds no new information. If  $e_{iS}$  was only partially correlated within state, the variance would be smaller, but still larger than OLS. Using White's clustering approach leads to a consistent estimate of the variance of  $\hat{\beta}$ , no matter what shape underlies  $\Omega$ .

One should note that this estimator applies asymptotically (as the sample size and the number of groups approaches infinity). Monte carlo experiments reveal that the estimator works reasonably well when the sample size within groups is not especially large relative to the number of groups. Unfortunately, the number of groups is very small, relying on asymptotics can be very misleading. What is small? The references below suggest even groups as high as 40 or 50 can lead to poor estimates. A conservative solution is to aggregate the data up to the group level and run the regressions using the grouped means, weighted by the sample size. In our example, this would be:

$$\bar{S}_s = \beta \bar{Z}_s + e_s,$$

which will generate the same estimate for  $B$  and the variance of  $B$  in our simple example. If there is no cluster effect (no serial correlation within groups), then aggregating to the group level removes information and increases the variance unnecessarily. In practice, results are far more convincing if you can produce robust and significant results with this aggregated approach (if it's applicable).

Note, in the diff-in diff example above, if we aggregated, we only would have 4 observations. And indeed, one criticism that has been put out by some researchers is that the diff in diff approach is just in essence comparing 2 groups over time and we can't be sure that any observed significant difference in means is due entirely to the policy change.