Microeconometrics
Supervision 2

Samuel Lee

A. Panel data

1. Like before, "crmrte" is the crimes committed per person, "polpc" is the police per capita, "west", "central" are dummy variables equal to 1 if the county is in western or central N.C. (presumably North Carolina), "urban" is a dummy variable equal to 1 if the county is in a standard metropolitan statistical area (SMSA), and "taxpc" is the tax revenue per capita. Additionally, "density" is the number of people per square mile. Variables with the prefix "l" are the logarithms of the original variables, while variables with the prefix "cl" are the change in the logarithms from the previous period (or the first difference of the log variables).

2. As mentioned, we have the first-differenced log variables so we can use the first-difference method if we want to. If we want to use the fixed effects method we would have to construct new variables for the deviation from the mean of any explanatory variables we think are correlated with some omitted fixed effect.

3.There may be county-specific effects invariant to time that drive some of the variation in crime rates between counties. If these fixed effects are also correlated with any of the regressors, such as the number of police or population density, then there is an omitted variable bias since we do not have measures for these fixed effects. For example, we may estimate a positive correlation between the crime rate and number of police, when some of that correlation is really driven by geographical features which positively affect both the crime rate and number of police.

4. The first column of Table 1 shows the results from regressing "lcrmrte" on "lpolpc", "ldensity", "urban", "west", and "central". The coefficient on "lpolpc" is 0.154, and this reflects that a percentage increase in the police per capita is *associated* with an approximately 0.154% increase in the crime rate.

5. The results of the first differenced regression are reported in the second column of Table 1.

6. The coefficient on "clpolpc" is 0.270, which is larger than the coefficient on "lpolpc" and also more statistically significant; using "clpolpc" did not result in much more imprecise estimates. "clpolpc" is the change in logs of police per capita from the previous period, so the coefficient implies that a percentage change in police per capita from the previous year is associated with an approximately 0.27% increase in the crime rate from the previous year. The first difference estimate is larger than the OLS estimate, which suggests that there might be county-specific fixed effects which bias the OLS estimates downwards.

7. As discussed on 26 February, "lpolpc" is likely to be endogenous as it is plausible that local authorities react to a higher crime rate by strengthening law enforcement. This leads to biased and inconsistent estimates of the coefficient on "lpolpc" and by extension "clpolpc", but the direction and magnitude of bias depends in part on how the crime 'supply' curve and police 'demand' curve have shifted over the years.

Table 1: Log and first difference estimates

	Dependent variable:	
	lcrmrte	clcrmrte
	(1)	(2)
lpolpc	0.154***	
	(0.028)	
ldensity	0.515***	
	(0.027)	
urban	−0.020	
	(0.068)	
west	−0.557***	
	(0.038)	
central	−0.322***	
	(0.036)	
clpolpc		0.270***
		(0.030)
cldensity		0.388
		(0.696)
Constant	−2.350***	−0.006
	(0.187)	(0.011)
Observations	630	540
R^2	0.602	0.128
Adjusted R^2	0.599	0.125
Residual Std. Error	0.363 (df = 624)	0.190 (df = 537)
F Statistic	188.784*** (df = 5; 624)	39.379*** (df = 2; 537)
Note:		*p<0.1; **p<0.05; ***p<0.01

B. Maximum Likelihood Estimation

a. To make things simpler we assume that the order of the outcomes matters. This will not be a problem when maximizing the likelihood or log-likelihood function later since ignoring the order of the outcomes just means multiplying everything by a constant. As such, the likelihood function gives the probability of getting the exact same sample for a given value p which is the probability of H in a single toss. We assign a value of 1 to H and 0 to T, and let k be the sum of all outcomes after tossing the coin 10 times. The likelihood function for 10 tosses is

$$L(p|\mathcal{X}_{10}) = p^k(1-p)^{10-k}$$

where $\mathcal{X}_{10}$ just represents the realized sample. It is easy to verify that the above expression is the probability of getting one specific combination of k heads and $10-k$ tails from 10 tosses.

b. In the sample we have (HHTTHTHHTH), $k=6$ and $10-k=4$. Therefore, the probability of obtaining the above sample if $p=0.5$ is

$$L\left(p=0.5|\mathcal{X}_{10}=\{HHTTHTHHTH\}\right) = 0.5^6 \times 0.5^4 = 0.5^{10}$$

c. The log likelihood is the logarithm of the likelihood function. In this case, for 10 tosses, the log likelihood function is

$$\ell(p|\mathcal{X}_{10}) = k\log p + (10-k)\log(1-p)$$

The value of the maximum likelihood estimator, $\hat{p}_{MLE}$, can be found by maximizing L or ℓ:

$$\begin{aligned}\frac{\partial}{\partial p}\ell\left(p=\hat{p}_{MLE}|\mathcal{X}_{10}\right) = \frac{k}{\hat{p}_{MLE}} - \frac{10-k}{1-\hat{p}_{MLE}} &= 0\\ (1-\hat{p}_{MLE})k &= \hat{p}_{MLE}(10-k)\\ \hat{p}_{MLE} &= \frac{k}{10}\end{aligned}$$

which is just the sample average calculated by number of successes over the total number of tries.

d. In our specific example, the estimator is $\hat{p}_{MLE} = \frac{k}{10}$. The variance of this estimator is $\mathrm{Var}\left[\frac{\sum_{i=1}^{10} X_i}{10}\right]$ where $X_i = 1$ when the i-th outcome is heads. Since X_i are independently and identically distributed, we have $\mathrm{Var}[\hat{p}_{MLE}] = \frac{1}{100}\mathrm{Var}[\sum_{i=1}^{10} X_i] = \frac{1}{10}\mathrm{Var}[X_i] = \frac{p(1-p)}{10}$, since

$$\mathrm{Var}[X_i] = \mathrm{E}[X_i^2] - \mathrm{E}[X_i]^2 = p(1^2) + (1-p)(0^2) - [p(1) + (1-p)(0)]^2 = p - p^2 = p(1-p)$$

We don't know what p is, but we can replace p with the maximum likelihood estimator $\hat{p}_{MLE}$ to get an estimate for the variance. This is also the same as using the method of moments since $\hat{p}_{MLE}$ is the sample first moment of X_i, and the estimate of the variance will converge in probability to the true variance of the MLE estimator.

e. By definition, the probability of obtaining the above sample if the true probability of H in a single draw was equal to $\hat{p}_{MLE}$ must be weakly greater than any other probability for other values of p. In this case, $\hat{p}_{MLE} = \frac{3}{5}$, and the probability of obtaining the above sample (which is the likelihood given $p = \hat{p}_{MLE}$) is $\left(\frac{3}{5}\right)^6\left(\frac{2}{5}\right)^4 = \frac{3^6 \times 2^4}{5^{10}}$ which is higher than the 0.5^{10} calculated in b.

f. With a null hypothesis that $p = 0.5$ and a two-sided alternative hypothesis that $p \neq 0.5$, and with 5% significance level, we need $\Pr(k \geq 6|p=0.5) \leq 0.025$ to reject the null hypothesis. If we had a larger sample size we could use a likelihood ratio test, but we only have 10 tosses so it's probably a good idea to do this the old-fashioned way:

$$\begin{aligned}\Pr(k \geq 6|p = 0.5) &= \binom{10}{6}0.5^6 \times 0.5^4 + \binom{10}{7}0.5^7 \times 0.5^3 \\ &+ \binom{10}{8}0.5^8 \times 0.5^2 + \binom{10}{9}0.5^9 \times 0.5 + 0.5^{10} \\ &= \left[\binom{10}{6} + \binom{10}{7} + \binom{10}{8} + \binom{10}{9} + 1\right] \times 0.5^{10} = \frac{193}{512}\end{aligned}$$

which is obviously not small enough to reject the null hypothesis (and common sense should tell us that getting 6 out of 10 heads is not a very strong case to claim that a coin is unfair).

C. Limited dependent variables

1. fatkids.dta contains cross-section data on various household characteristics, including indicators of obesity.
2. Table 2 shows the variable names and their associated variable labels:

Table 2: Description of variables

Variable name	Variable label
hhsize	household size
povrat	income to needs ratio
bmi	own body mass index (wt for height measure)
ageyrs	age in years
female	=1 if female
white	=1 if white
black	=1 if black
hisp	=1 if hispanic
other	=1 if other race
tvyest	# hours of TV watched yesterday
dadbmi	dad's body mass index
mombmi	mom's body mass index
dadobese	=1 if dad obese
dadovwt	=1 if dad overweight
momobese	=1 if mom obese
momovwt	=1 if mom overweight
ovwtc	=1 if overweight
obesec	=1 if obese
pctcal	calories consumed/2000

3. And Table 3 shows the summary statistics for the data.
4. "obesec" is a dummy for whether the child is obese, although it doesn't seem to be based purely on BMI since there are some people with a BMI of around 18.5 that are classified as obese while some with a BMI of around 29(!) are classified as not obese.
5. Table 4 shows the result of the probit regression.
6. The coefficient on "hisp" is positive and significant, which means that being Hispanic is associated with a higher chance of being obese relative to not being white, black, or Hispanic. The

Table 3: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
hhsize	7,785	4.913	1.826	1	4	6	10
povrat	7,157	1.730	1.316	0.000	0.691	2.556	7.517
bmi	6,329	19.022	4.582	11.161	15.704	21.164	55.191
ageyrs	7,785	9.415	3.366	5	6	12	16
female	7,785	0.507	0.500	0	0	1	1
white	7,785	0.319	0.466	0	0	1	1
black	7,785	0.310	0.463	0	0	1	1
hisp	7,785	0.317	0.465	0	0	1	1
other	7,785	0.054	0.226	0	0	0	1
tvyest	6,350	3.141	1.802	0.000	2.000	4.000	6.000
dadbmi	6,529	26.463	4.310	11.554	23.625	28.728	58.745
mombmi	7,168	26.194	5.913	14.631	21.925	29.261	56.810
dadobese	6,529	0.166	0.372	0.000	0.000	0.000	1.000
dadovwt	6,529	0.614	0.487	0.000	0.000	1.000	1.000
momobese	7,168	0.223	0.416	0.000	0.000	0.000	1.000
momovwt	7,168	0.489	0.500	0.000	0.000	1.000	1.000
ovwtc	6,329	0.275	0.446	0.000	0.000	1.000	1.000
obesec	6,329	0.134	0.341	0.000	0.000	0.000	1.000
pctcal	7,255	0.918	0.445	0.000	0.608	1.137	5.153

exact estimated marginal effects of being Hispanic on the probability of being obese depends on the values of the other regressors:

$$\begin{aligned}
&\Delta\hat{\Pr}(obesec = 1)\\
=&\Phi(-1.56 + 0.008 ageyrs + 0.03 female + 0.024 white + 0.204 black\\
&+ 0.261 hisp + 0.057 tvyest + 0.002 povrat)\\
&- \Phi(-1.56 + 0.008 ageyrs + 0.03 female + 0.024 white + 0.204 black + 0.057 tvyest + 0.002 povrat)
\end{aligned}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. In the dataset the race dummy variables seem to be mutually exclusive, i.e. there is no allowance for biraciality, so if "white" or "black" is equal to 1 in the first half of the above expression then it should be equal to 0 in the second half.

7. The predicted probability for such an individual is $\Phi(-1.56 + 0.008 \times 16 + 0.03 + 0.024 + 0.057 \times 3 + 0.002 \times 4.56) \approx 0.1154819$.

8. If we take someone who was identical in every way except for watching no TV the day before, then the probability becomes $\Phi(-1.56 + 0.008 \times 16 + 0.03 + 0.024 + 0.002 \times 4.56) \approx 0.08551839$, and the fall in probability is about 0.02996347 or roughly 3 percentage points. The fall is statistically significant since the coefficient on "tvyest" is statistically significant.

9. We have to test whether "white", "black", and "hisp" are jointly zero, which now can be done using a likelihood-ratio test since we have a much larger sample size than in that coin toss example. Wilks' theorem suggests that $2(\ell_{ur} - \ell_r)$ converges in distribution to a $\chi^2(q)$ distribution where ℓ_{ur} and ℓr are the maximized log-likelihood functions of the unrestricted and restricted model, and q is the number of restrictions. In this case the unrestricted model is the one estimated in C5., and

Table 4: Probit regression

	Dependent variable:
	obesec
ageyrs	0.008
	(0.006)
female	0.030
	(0.042)
white	0.024
	(0.116)
black	0.204^{*}
	(0.112)
hisp	0.261^{**}
	(0.112)
tvyest	0.057^{***}
	(0.012)
povrat	0.002
	(0.018)
Constant	-1.560^{***}
	(0.132)
Observations	5,770
Log Likelihood	−2,234.689
Akaike Inf. Crit.	4,485.379
Note:	*p<0.1; **p<0.05; ***p<0.01

the restricted model is the same except "white", "black", and "hisp" are not included. $2(\ell_{ur} - \ell_r)$ turns out to be more than 19, and for a $\chi^2(3)$ distributed test statistic this is significant down to the 0.1% level. Using the command `test white black hisp` after the probit regression, we get the same result, and `Prob > chi2 = 0.0002`. Therefore we reject the null hypothesis that "white", "black", and "hisp" are jointly insignificant or that race has no effect on the probability of being obese.

10. The coefficient "tvyest" is highly significant (it is the most significant coefficient among all the regressors), and if we take a causal interpretation of the coefficient then discouraging young children from watching TV could be good advice for obesity prevention. While there is probably some causal link between watching TV and obesity, it is unlikely that the causality explains the entire estimated association in the probit regression. It is more likely that the number of hours of TV watched also proxies for other relevant variables not included in the regression, such as overall activity levels or the amount of parental contact/supervision. It could be worth checking if the estimated effects of "tvyest" are robust to including other variables such as "dadobese" or "momobese", or perhaps dummy variables indicating whether the child is involved in sports. However, this may lead to more imprecise estimates if doing so results in more missing observations and an effectively smaller sample size.