Faculty of Economics & Politics

Part2A Paper 3: Microeconometrics and Panel Data

Supervision 2

1. For binary response data in the form $\{y_i, x_i\}$, where y_i is either zero or one, and x_i is a $k \times 1$ vector of independent variables, evaluate the following statements

   a) Within a standard linear regression model the disturbances cannot be normally distributed.

   b) The use of an OLS estimator applied to this data is invalid since the disturbances are heteroscedastic.

   c) Briefly explain why the use of a link (or transformation) function, say $F(x_i\beta)$, where $F(-\infty) = 0$; $F(\infty) = 1$; $dF(x_i\beta)/dx\beta > 0$, is a central component of binary response models.

   d) Unlike the standard linear regression model, the probit model requires an additional assumption for the purpose of estimation. Namely, instead of assuming, that the errors, say ε, are distributed $\varepsilon \sim (0, \sigma^2)$, the probit model requires $\varepsilon \sim N(0, \sigma^2)$. Why is this the case?

2. Consider the following set of unobserved state-specific utilities for full-time (U_1^*) and part-time employment (U_2^*)

$$\begin{aligned} U_{1i}^* &= \alpha_1 + x_i\beta_1 + u_{1i} \\ U_{0i}^* &= \alpha_0 + x_i\beta_0 + u_{0i} \end{aligned}$$

   where x_i denotes a common set of exogenous variables, β_0, and β_1 are unknown parameters, and u_0, and u_1 denote unobservable taste components.

   With reference to the above model, explain how you would specfiy and estimate a model of binary choice. Which elements of the set of mean parameters $\alpha_1, \beta_1, \alpha_0, \beta_0$ are identifiable.

3. Let Y denote a binary indicator of voting behaviour: $Y = 1$ and $Y = 0$, respectively, represents a vote for Party A and Party B. Let X denote the differential evaluation of the two candidates where $X = X^A - X^B$ and X^A (X^B) represents the ranking of Party A (Party B) candidates. The min (max) values for both X^A and X^B are 0 (100).

   i) Show that for $P(Y_i = 1|X_i) = E(Y_i|X_i) = \beta_1 + \beta_2 X_i$ to lie within the unit interval the following constraint must be satisfied

$$0 \leq \beta_1 - 100\beta_2, \ \beta_1 + 100\beta_2 \leq 1$$

ii) Show that if the constraints are exact, then for every unit increase in the relative liking of Party A compared to Party B, the probability of voting for Party A can increase by no more than 0.005.

4. A data file *phillipines.xls*[1], located in the directory `J:\students\mw217,` contains the results from a survey of small scale farmers in the Philippines. The focus of the survey is to obtain insight into the determinants of the use of chemical fertilizers. Potential determinants include: land tenure status and the amount of credit obtained. A full list is given below

   Using the SIGN function in MicroFit to create a dichotomous variable, FERUSE, indicating whether or not fertiliser is used; this will emulate the use of the indciator function, namely $\mathbf{1}(QFER > 0)$, where $QFER$ denotes the *quantity* of chemical fertilizers used, and $\mathbf{1}(.)$ is the indicator function. Estimate two models:

   - a linear probability model
   - a probit model

   using the same set of explanatory variables in each. Discuss your results, interpreting the estimated parameters, and comment on any possible mis-specification.

   Note: This SIGN function when applied to a variable Y returns the value of 1 when Y is positive, and 0 when Y is zero or negative

Data

The *phillipines.xls* dataset is based upon a cross-section of small-scale farmers in the Philippines. There are 491 observations

Dependent Variable

FERUSE	$\mathbf{1}(QFER > 0)$

RHS variables

DMARKET	= distance from farm from nearest market
HOURMEET	= number of hours farmer met with an agricultural 'expert'
OWNER	= 1 if farmer owns some land, 0 otherwise
IRSTAT	= 1 if farm irrigated, 0 otherwise
CREDIT	= amount (pesos/hectare) of loans obtained by farmer

[1] A copy is also on my web page.

5. *SECTION B Question 2002*

   In a random sample of married women 428 report working non-zero hours during a given year. In estimating a labour force participation model using the linear probability model, the following results were obtained.

$$\begin{aligned}\widehat{inlf} &= \underset{(.154)}{.586} - \underset{(.0014)}{.0034}nwifeinc + \underset{(.007)}{.038}educ + \underset{(.006)}{.039}exper - \underset{(.00018)}{.00060}exper^2 \quad (1)\\ &\quad - \underset{(.002)}{.016}age - \underset{(.034)}{.262}kidslt6 + \underset{(.013)}{.013}kidsge6\end{aligned}$$

   where $inlf$ is a binary variable equal to one if the woman works, *nwifeinc* denotes nonwife income in thousands, educ ...., exper ...Age, education, (), *kidsit6* denotes the number of children less than six years of age, and *kidsge6* is the number of kids between 6 and 18 inclusive .

   a. Interpret both the economic and statistical significance of these results. Independent of how many young children a woman has, and regardless of the levels of other independent variables, what is the impact of one more child aged between 6 and 18 on labour force participation on labour force participation.

   (a) Carefully explain why there is no guarantee that the predicted values of $inlf$ lie within the unit interval.

   (b) *Although the linear probability model has heteroscedastic error by construction, the form of the heterosedasticity is known and therefore a weighted least squares estimator may be used for estimation.* Evaluate this statement.

   d. As an alternative to OLS we could have estimated a probit model. Given a resulting log-likelihood value (l^{UN}) using the regressors in (1) of -401.30 and an observed frequency of participation of $\widehat{p} = 428/753$, find the value of the log-likelihood for the most restricted model (l^R) with all slope parameters set at zero and calculate the likelihood ratio *(LR)* statistic, $LR = 2(\ell^{UN} - \ell^R)$. Are the restrictions consistent with the data?

6. *SECTION B Question 2004*

   Consider the linear probability model

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \tag{2}$$

   where $\Pr(Y_i = 1|X_i) = \beta_0 + \beta_1 X_i$.

   (a) Show that $E(u_i|X_i) = 0$

(b) Show that $var(u_i|X_i) = (\beta_0 + \beta_1 X_i)[1 - (\beta_0 + \beta_1 X_i)]$

(c) Is u_i heteroscedastic? Would a weighted least squares estimator be more efficient?

Melvyn Weeks
Lent 2008