Sharif University of Technology
Econometrics. PhD.
Final Exam, fall 2020
Mohammad Hossein Rahamti

1. In Deb and Trivedi (Journal of Health Economics, 2002), data from the Rand Health Insurance Experiment was used to study health services use under different health insurance plans. In this experiment, families were enrolled in one of 14 different health insurance plans. These plans were not freely chosen by the participants, but randomly assigned to the families. The following model for the number contacts with a medical doctor (Yi) for individual i can then be considered:

$$P[Y_i = y|X_i] = \frac{\exp(-\mu(X_i))\,\mu(X_i)^y}{y!}, \quad y = 0,1,\cdots$$

In other words, given Xi, Yi follows a Poisson distribution with parameter $\mu(X_i)$. Remember that, among other things, this implies that

$$E(Y_i|X_i = x) = var(Y_i|X_i = x) = \mu(x)$$

Assume that $\mu(Xi) = exp(X_i'\alpha_0)$ where $X_i = [1\ X_{i1} \cdots X_{i(K-1)}]$is a $K \times 1$ vector of covariates (including a constant) and $\alpha_0 \in R^K$ is the $K \times 1$ parameter vector of interest. The covariate vector includes information on plan characteristics (e.g. coinsurance percentage and deductible) as well as demographic and health characteristics. Throughout the question, assume that (1). Observations are iid; (2). The support of X_i is not contained in any proper linear subspace of R^K(i.e. for any vector $\lambda \in R^K, P(\lambda' X_i = 0) < 1$) (3). $E[X_i X_i']$ is finite (Remark: It is also nonsingular because of the previous item.); (4). $||X_i||$ has a well defined Moment Generating Function.

If you need to invoke any other assumptions, please state them explicitly.

Consider the following estimator:

$$\hat{\alpha} = argmin_\alpha \frac{1}{N}\sum_{i=1}^{N}[Y_i X_i'\alpha - \exp(X_i'\alpha) - lnY_i!] \qquad (1)$$

a) Show that $\hat{\alpha}$ is consistent.
b) Derive its asymptotic distribution.
c) The researcher is interested in the marginal effect of an increase in variable

$$\gamma = E\left(\frac{\partial E[Y_i|X_i]}{\partial X_i^{(k)}}\right) = \alpha_0^{(k)} E[\exp(X_i'\alpha_0)]$$

Where $X_i^{(K)}$is the k-th variable and $\alpha_0^{(K)}$, its coefficient. The following estimator is suggested:

$$\hat{\gamma} = \hat{\alpha}^{(K)} \frac{1}{N}\sum_{i=1}^{N} Y_i$$

What is its asymptotic distribution?

d) Because (conditional on Xi), Yi follows a Poisson distribution

$$E(Yi|Xi = x) = \mu(x) = \exp(x'\alpha_0)$$

Scanned by CamScanner

Suggest an estimator for α_0 based on this fact. How does it compare to the estimator above (i.e. Is it consistent? Is it more or less efficient?)

e) The researcher realizes that Xi is measured with error. Let $\tilde{X}_i = X_i + \varepsilon_i$ denote the observed covariate vector and assume that ε_i is independent of Xi. Suppose that the estimator in (1) relies on the mismeasured data ($\tilde{X}_i$). Show that the objective function converges to the function

$$E\left[-e^{X_i'\alpha} + Y_i X_i'\alpha - \ln(Y_i!)\right] + E\left[-e^{X_i'\alpha}\right]\left(E\left[e^{\varepsilon_i'\alpha}\right] - 1\right)$$

f) It can be shown that the function in (e) is typically not maximized at α_0. Consequently, the identification condition for the estimator in (a) fails when there is measurement error. The econometrician nevertheless is comfortable with the assumption that $\varepsilon_i \sim N(0, \Omega)$ (where Ω is known). Some algebra (based on the moment generating function of a normal random variable) then yields

$$\exp(X_i'\alpha) = E_{\tilde{X}_i|X_i}\left[\exp(\tilde{X}_i'\alpha - \frac{\alpha'\Omega\alpha}{2})\right] \Rightarrow E[\exp(X_i'\alpha)] = E\left[\exp(\tilde{X}_i'\alpha - \frac{\alpha'\Omega\alpha}{2})\right]$$

(Remark: You do NOT need to calculate this!) Consider the following estimator for α:

$$\tilde{\alpha} = argmin_\alpha \frac{1}{N}\sum_{i=1}^{N}[Y_i\tilde{X}_i'\alpha - \exp\left(\tilde{X}_i'\alpha - \frac{\alpha'\Omega\alpha}{2}\right) - lnY_i!]$$

Show that $\tilde{\alpha}$ is consistent.

g) Now back to a setting without measurement error, suppose that the number of medical doctor contacts is "top coded" and the econometrician can only observe the number of contacts up to 5 visits. Although the researcher can learn whether an individual had 5 visits or more, she cannot tell how many. Construct a likelihood function for the estimation of α under this observation scheme and distributional assumption.

2. Assume we have two latent variables (y_1^*, y_2^*), with a distribution

$$(y_1^*, y_2^*) \sim Bivariate(x_1'\beta_1, x_2'\beta_2, 1, \sigma_2^2, \sigma_{12}).$$

We only observe y_2 with the following equation:

$$y_2 = \begin{cases} y_2^* & if\ y_1^* > 0\ \ and\ y_2^* > 0 \\ 0 & otherwise \end{cases}$$

a) Write the likelihood function to estimate the parameters.

b) Now assume that $x_1 = (z_1, y_3)$ and $x_1'\beta_1 = z_1'\gamma + \alpha y_3$ and y_3 is an endogenous variable. A vector of exogenous variables to the disturbance of y_2^* and y_1^* are available as z_2 that are correlated with y_3. Propose a two-step procedure to estimate this equation. If variance adjustment is needed you should highlight how you correct standard errors.

c) Write the likelihood function to estimate the parameters.

d) Construct a test for endogeneity of y_3

3. Consider the static linear panel data model:

Scanned by CamScanner

$$y_{it} = x_{it}\beta_0 + \alpha_i + \varepsilon_{it}$$

Which in matrix form can be written as

$$Y = X\beta_0 + D\alpha + \varepsilon$$

Where as usual Y and ε are $NT \times 1$vectors, X is an $NT \times K$ marix, α is an $N \times 1$ vector and $D = I_N \otimes e_T$ with $r_T$the T-dimensional unit vector. We want to test for random versus fixed effects. Assume that:

$$E(\varepsilon|X, \alpha) = 0$$
$$E(\varepsilon\varepsilon'|X, \alpha) = V(\varepsilon|X, \alpha) = \sigma_\varepsilon^2 I_{NT}$$

While under the null the following two assumptions hold:

$$E(\alpha|X) = E(\alpha) = 0$$
$$E(\alpha\alpha'|X) = V(\alpha|X) = V(\alpha) = \sigma_\alpha^2 I_N$$

(a) Derive the asymptotic distribution of the Hausman test which is based on the difference between the random effect and fixed effects estimators of β_0

(b) Show that the test in (a) is numerically equal to the Huasman test which is based on the difference between the between effects and fixed effects estimators of β_0

(c) Show that the test in (b) is nothing but the Wald test of the hypothesis that $\gamma = 0$ in the following augmented model:

$$Y^* = X^*\beta + \tilde{X}\gamma + \omega$$

Where X^* and $Y^*$are the GLS transformed X and Y, i.e. $X^* = \Omega^{-\frac{1}{2}}X$ and $Y^* = \Omega^{-\frac{1}{2}}Y$and $\tilde{X}$ is the within transformed X, i.e. $\tilde{X} = QX$

Scanned by CamScanner