University of Pennsylvania | Gregory Kordas
Economics 705 | Spring 2006

Econometrics Prelim
Part I (Econ 705)

Instructions: Answer all questions. Use a different blue book for each question and remember to write your assigned number (not your name!) on each blue book. There is a total of 330 points. Good Luck!

1. (70 points) [**Joint and Conditional Probabilities**]. Let

$$f_{X|Y}(x|y) = \begin{cases} c_1 x/y^2, & 0 < x < y < 1, \\ 0, & \text{otherwise,} \end{cases}$$

be the conditional p.d.f of $X|Y$ and

$$f_Y(y) = \begin{cases} c_2 y^4, & 0 < y < 1, \\ 0, & \text{otherwise,} \end{cases}$$

be the marginal p.d.f of Y. Determine

- **(a)** (10 points) The constants c_1 and c_2.
- **(b)** (10 points) The joint p.d.f of X and Y.
- **(c)** (10 points) $\Pr(\frac{1}{4} < X < \frac{1}{2}|Y = \frac{5}{8})$
- **(d)** (10 points) $\Pr(\frac{1}{4} < X < \frac{1}{2})$
- **(e)** (10 points) $E(X|Y)$.
- **(f)** (20 points) The cdf and pdf of $Z = E(X|Y)$, F_Z and f_z, respectively.

2. (120 points) [**Random Misclassification in a Binary Choice Model**]. Let y_i^* be a latent variable given by

$$y_i^* = x_i'\beta + \varepsilon_i, \qquad i = 1, ..., n,$$

where the x_i's are a k-vector of individual characteristics, β is a k-vector of parameters to be estimated from data, and the ε_i's are iid F. In what follows we will assume that F is known and symmetric (e.g., normal, logistic, or something similar). Let

$$\widetilde{y}_i = \begin{cases} 1, & \text{if } y_i^* \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

be the binary indicator of the event $\{y_i^* \geq 0\}$, which we will call 'success'. It follows that

$$E(\widetilde{y}_i = 1|x_i) = \Pr(\widetilde{y}_i = 1|x_i) = F(x_i'\beta).$$

Given a random sample $\{(\widetilde{y}_i, x_i), i = 1, ..., n\}$, we could estimate β_0 using a binary choice model (i.e., a logit or a probit depending on the specification of F).

The situation that will concern us here is that of a *random misclassification* of the response variable $\widetilde{y}_i$. More specifically, we will consider the situation where the researcher does not observe $\widetilde{y}_i$ but instead he observes another binary variable y_i that is equal to $\widetilde{y}_i$ for some individuals in the sample but also has errors misclassifying some as 'sucesses' when they are 'failures' and vice versa. A crucial assumption that we are going to make is that *the process governing the misclassification of the $\widetilde{y}_i$'s is random and independent of the x_i's.* (One way of describing this situation is to think of a researcher that, being ready to run his probit, he realizes that a computer virus he downloaded from the internet has, in a completely random way, corrupted his $\widetilde{y}$ data changing some its '1's to '0's, some its '0's to '1's, and leaving some of it unchanged – Another example with an economic interpretation could also be given but the 'virus' scenario is enough for our purposes).

Define the *misclassification probabilities*

$$\begin{aligned}\alpha_0 &= \Pr(y_i = 1|\widetilde{y}_i = 0),\\ \alpha_1 &= \Pr(y_i = 0|\widetilde{y}_i = 1),\end{aligned}$$

so that, the probability that a '0' is misclassified as an '1' is α_0, and the probability that an '1' is misclassified as a '0' is α_1.

(a) (15 points) Show that in the random misclassification model

$$\Pr(y_i = 1|x_i) = \alpha_0 + (1 - \alpha_0 - \alpha_1)F(x_i'\beta).$$

(b) (15 points) Using your result in part (a) give the sample log-likelihood of the model $\ell_n(a_0, a_1, b)$ and define the MLE of $(\alpha_0, \alpha_1, \beta)'$. Hint:

$$\ell_n(a_0, a, b) \quad = \quad n^{-1}\sum_{i=1}^{n} y_i \log \Pr(y_i = 1|x_i) + (1 - y_i) \log\Big(1 - \Pr(y_i = 1|x_i)\Big).$$

(c) (30 points) Apart from the usual identification conditions for binary choice models (e.g. the design matrix X is of full order and the like), an important additional identification condition is that

$$\alpha_0 + \alpha_1 < 1.$$

Using the symmetry of F, first prove that

$$\alpha_0 + (1 - \alpha_0 - \alpha_1)F(x_i'\beta) = (1 - \alpha_1) + (1 - (1 - \alpha_1) + (1 - \alpha_0))F(x_i'(-\beta)),$$

Then argue that if $\alpha_0 + \alpha_1$ is allowed to be arbitrary, the result above means that it would be impossible to distinguish between $(\alpha_0, \alpha_1, \beta)$ and $(1 - \alpha_1, 1 - \alpha_0, -\beta)$, which is a failure of (point) identification. Finally, impose the condition $\alpha_0+\alpha_1 < 1$ and argue that identification is restored.

(d) (30 points) Argue that, under the above and the usual conditions, the sample log-likelihood $\ell_n(a_0, a_1, b)$ converges uniformly to the population log-likelihood $\ell(a_0, a_1, b)$, and show that the later is given by

$$\begin{aligned}\ell(a_0,a_1,b) &= E_x\Big[\Big(\alpha_0+(1-\alpha_0-\alpha_1)F(x'\beta)\Big)\log\Big(a_0+(1-a_0-a_1)F(x'b)\Big)\\ &\quad+\Big(1-\alpha_0+(1-\alpha_0-\alpha_1)F(x'\beta)\Big)\log\Big(1-a_0+(1-a_0-a_1)F(x'b)\Big)\Big]\end{aligned}$$

Finally argue that the MLE is consistent and asymptotically normal, i.e. that $(\widehat{\alpha}_0,\widehat{\alpha}_1,\widehat{\beta})' \xrightarrow{p} (\alpha_0,\alpha_1,\beta)'$ and

$$\sqrt{n}\Big((\widehat{\alpha}_0,\widehat{\alpha}_1,\widehat{\beta})' - (\alpha_0,\alpha_1,\beta)'\Big) \xrightarrow{d} N(0, I^{-1}),$$

where I is the Fisher Information matrix.

(e) (30 points) Determine the three missing elements of the Fisher information matrix for this model

$$\begin{aligned}I &= -E_x\left[\begin{bmatrix} [1] & \dfrac{\partial^2\ell}{\partial a_0\partial a_1} & \dfrac{\partial^2\ell}{\partial a_0\partial b'} \\ \dfrac{\partial^2\ell}{\partial a_1\partial a_0} & \dfrac{\partial^2\ell}{\partial a_1^2} & \dfrac{\partial^2\ell}{\partial a_1\partial b'} \\ [2] & \dfrac{\partial^2\ell}{\partial b\partial a_1} & [3]\end{bmatrix}\right|_{(a_0,a_1,b)=(\alpha_0,\alpha_1,\beta)} \\ &= E_x\frac{1}{P(1-P)}\begin{bmatrix} [1] & -F(1-F) & (1-\alpha_0-\alpha_1)f(1-F)x' \\ -F(1-F) & F^2 & -(1-\alpha_0-\alpha_1)fFx' \\ [2] & -(1-\alpha_0-\alpha_1)fFx & [3]\end{bmatrix},\end{aligned}$$

where $f \equiv f(x'\beta)$, $F \equiv F(x'\beta)$, and $P \equiv \alpha_0 + (1-\alpha_0-\alpha_1)F(x'\beta)$. Denote your answers by the number of the element missing.

3. (70 points) **[Structural Change]**. Consider the classical time-series linear regression model

$$y = X\beta + u, \qquad u \sim \text{iid}N(0, \sigma^2 I).$$

where y is an n vector, X is a $n \times k$ matrix of order k (full order), β is a k vector of coefficients, and u is a homoskedastic normal error term.

Recall that the general linear hypothesis may be written as

$$H_0 : R\beta = r$$

where R is a $q \times k$ restriction matrix (with $q < k$), and r is a q vector of known constants.

(a) (10 points) Starting from the fact that in this model the OLS estimate b is distributed as

$$b \sim N(\beta, \sigma^2(X'X)^{-1})$$

(explain why) show that under the null

$$(Rb - r)'[\sigma^2 R(X'X)^{-1}R']^{-1}(Rb - r) \sim \chi^2(q).$$

(b) (10 points) Using the fact that (explain why)

$$\frac{u'u}{\sigma^2} \sim \chi^2(n-k),$$

determine the distribution of the statistic

$$D = \frac{(Rb - r)'[R(X'X)^{-1}R']^{-1}(Rb - r)/q}{u'u/(n-k)}.$$

Now consider OLS estimation under the constraint. The restricted least squares (ROLS) estimator b_* minimizes the Lagrangian

$$(y - Xb)'(y - Xb) - 2\lambda'(Rb - r)$$

where λ is a q vector of Lagrange multipliers.

(c) (10 points) Show that the ROLS estimator is given by

$$b_* = b + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - Rb).$$

(d) (15 points) Writing u for the OLS and u_* for the ROLS residuals first show that

$$u_*'u_* = u'u + (b_* - b)'X'X(b_* - b)$$

and then that

$$u_*'u_* - u'u = (r - Rb)'[R(X'X)^{-1}R']^{-1}(r - Rb).$$

Thus our statistic above may be written as

$$D = \frac{(u_*'u_* - u'u)/q}{u'u/(n-k)}.$$

Explain briefly the intuition for this statistic.

Now consider the situation where a researcher is worried that at some specified moment of time a *structural change* has occurred, that resulted in a shift in β. Let $y_i, X_i, i = 1, 2$ indicate the partitioning of the data into the two subperiods, which we will call *peace time* and *war time*, and consider the model

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

where $\beta_i, i = 1, 2$ are the relevant k vectors of coefficients for the subperiods and $u_i, i = 1, 2$ are also iid with common variance σ^2. We also assume that the X_i's are of full order too. We are interested in testing the null hypothesis

$$H_0 : \beta_1 = \beta_2$$

(e) (5 points) Specify R and r for this hypothesis.

(f) (20 points) Describe the process you would use to test this hypothesis given a sample of $n = n_1 + n_2$ observations, and give the test statistic and its theoretical distribution under the null.

4. (70 points) Consider the balanced panel data model

$$y_{it} = \beta x_{it} + Z_i \gamma + \alpha_i + u_{it}, \quad i = 1, ..., n, \; t = 1, ..., T$$

where

y_{it} = log gasoline demand in gallons/month of household i at time t.

x_{it} = log of (average) price of gasoline paid by household i at time t.

Z_i = a k-vector ($k > 1$) of time invariant household characteristics including income, family size, etc.

α_i = unobserved household specific effect.

u_{it} = everything else.

A serious potential problem with this model is that the price paid by household i at time t, x_{it} could be correlated with the household-specific intercept α_i. This could happen because some households may search more intensely than others for a low price, so some of the observed price variability may be due to an endogenous "shopping effect" rather than purely exogenous price variability. One might expect that this shopping effect could be correlated with the household-specific demand (fixed effect) α_i.

(a) (15 points) Explain briefly why, if the shopping effect hypothesis is correct, pooled-OLS is an unsatisfactory method of estimating the parameters in this model.

(b) (15 points) If the primary objective of estimating this model is to estimate the price elasticity β, suggest a way to accomplish this which avoids the potential problems alluded to above.

(c) (25 points) Suppose now we are interested in testing for the presence of the "shopping effect" in our data. We estimate the model by the method recommended in part (b) and obtain $\widehat{\beta} = -0.70$ with a standard error of .08. Then, you also estimate the model using the Random Effects estimator, i.e., treating the α_i's as a random sample with mean α_0 and variance σ_α^2. From this you obtain $\widehat{\beta} = -0.40$

with a standard error of .06. Use these results to test for bias due to endogeneity. State your null and alternative and explain briefly your test statistic.

(d) (15 points) What does your conclusions on part (c) means in terms of our ability to estimate the parameter vector γ from the data at hand? Explain.

> I was gratified to answer promptly. I said I don't know.
>
> — *Mark Twain, Life on the Mississippi.*