BGPE Frontiers in Econometrics 2008

A. Colin Cameron, U.C.-Davis

Final Exam: August 6

Open book. Read question carefully so you answer the question.

Keep answers as brief as possible.

Answer three of the four questions.

1. Consider the estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{A}\mathbf{X})^{-1}\mathbf{Z}'\mathbf{A}\mathbf{y}$, where $\mathbf{A}$ is a symmetric diagonal $N \times N$ matrix of constants, in the multiple regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

where $\mathbf{y}$ is an $N \times 1$ vector, $\mathbf{X}$ is an $N \times k$ matrix, $\mathbf{Z}$ is an $N \times k$ matrix, $\boldsymbol{\beta}$ is a $k \times 1$ vector and $\mathbf{u}$ is an $N \times 1$ vector.
We assume that $\mathbf{u}|\mathbf{Z} \sim [\mathbf{0},\ \Sigma]$ where $\Sigma \neq \sigma^2\mathbf{I}$.
For asymptotic theory assume that necessary laws of large numbers and central limit theorems can be applied.

(a) Obtain $\mathrm{E}[\widehat{\boldsymbol{\beta}}]$ and $\mathrm{V}[\widehat{\boldsymbol{\beta}}]$. For this part only assume that $\mathbf{Z}$ and $\mathbf{X}$ are nonstochastic.

(b) Show that $\widehat{\boldsymbol{\beta}}$ is consistent, stating any necessary assumptions.

(c) Obtain the limit distribution of $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$.

(d) Suppose the errors are heteroskedastic, so $\mathrm{V}[u_i] = \sigma_i^2$ varies with i and the model for σ_i^2 is unknown. Give a heteroskedastic-consistent estimate of the covariance matrix of the estimator $\widehat{\boldsymbol{\beta}}$ analyzed in parts (a)-(c).
[Your answer here can be very brief. Just give the answer with brief explanation].

(e) Suppose now that $\mathbf{Z} = \mathbf{X}$, so $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\mathbf{y}$. State how you would obtain $\widehat{\boldsymbol{\beta}}$ using only an OLS package.

2. Consider the random variable y with density

$$f(y) = \lambda^2 y \exp\left(-\lambda y\right), \quad y > 0, \lambda > 0.$$

For this distribution it can be shown that $\mathrm{E}[y] = 2/\lambda$ and $\mathrm{V}[y] = 2/\lambda^2$.

Here we introduce regressors and suppose that in the true model the parameter γ depends on regressors according to

$$\lambda_i = 2\exp(-\mathbf{x}_i'\boldsymbol{\beta}),$$

where $\boldsymbol{\beta}$ is an unknown $K \times 1$ parameter vector and $\mathbf{x}_i$ is a nonstochastic $K \times 1$ vector.
Thus $\mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$ and $\mathrm{V}[y_i|\mathbf{x}_i] = [\exp(\mathbf{x}_i'\boldsymbol{\beta})]^2/2$.
In the d.g.p. $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, data are independent over i and $\mathbf{x}_i$ is nonstochastic.

(a) Give the log-likelihood function for $\boldsymbol{\beta}$.

(b) Show that the derivative of the log-likelihood function with respect to $\boldsymbol{\beta}$ can be written as

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum\nolimits_i 2 \times \frac{y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta})}{\exp(\mathbf{x}_i'\boldsymbol{\beta})} \mathbf{x}_i.$$

In answering the subsequent questions use the answer in part (b) even if you cannot derive it.

(c) What essential condition do the first-order conditions indicate needs to be satisfied for $\widehat{\boldsymbol{\beta}}$ to be consistent?

(d) Give the limit distribution of $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, assuming that the density is correctly specified.

(e) For this specific example provide a formula for computation of the MLE by an iterative procedure.

3. **(a)** Consider estimation of the model $y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}$, where $\mathbf{x}_{it}$ are time-varying regressors that are uncorrelated with α_i but are correlated with u_{it}. State how to obtain a consistent estimator for $\boldsymbol{\beta}$, with a brief explanation.

(b) Consider the same model as part (a), except that $\mathbf{x}_{it}$ is also correlated with α_i. State how to obtain a consistent estimator for $\boldsymbol{\beta}$, with a brief explanation.

(c) Consider estimation of a logit model for binary dependent variable y_i with endogenous regressors $\mathbf{x}_i$. Assume that the exist variables $\mathbf{z}_i$ such that $\mathrm{E}[(y_i - \Lambda(\mathbf{x}_i'\boldsymbol{\beta}))|\mathbf{z}_i] = \mathbf{0}$. State how to obtain a consistent estimator for $\boldsymbol{\beta}$, with a brief explanation.

(d) Consider estimation of a three choice model with $\Pr[y_i = j] = F_j(\mathbf{x}_i'\boldsymbol{\beta}_j)$, $j = 1, ..., 3$. Give the log-likelihood function for the model.

(e) Explain what is meant by an optimal estimator.

4. Consider output from the following three regressions. The dependent variable `fairpoor` is equal to one if an individual's health status is fair or poor, and is equal to zero otherwise (health status is good or excellent). The regressors are `age` (age in years / 10), `income` (annual income in thousands of dollars), and `chronic` (indicator equal to 1 if have a chronic disease).

(a) For the first set of regression estimates, are the signs of the coefficients what you expect?

(b) For the first set of regression estimates, are the regressors statistically significant at level 0.05? Explain.

(c) For the first set of regression estimates, provide a meaningful interpretation of the estimated effect of a \$1,000 increase in income on the probability of health status being fair or poor.

(d) Under what circumstances will the first set of regression estimates be consistent and fully efficient?

(e) What is the difference between the results from the first regression and the results from the second regression, and is this difference great?

(f) Which model, if any, fits the data better – the first model or the third model? Explain.

```
. use mus10data.dta, clear
. quietly keep if year02==1
. describe fairpoor age income chronic
              storage  display     value
variable name   type   format      label      variable label
-----------------------------------------------------------------------------------
fairpoor        byte   %8.0g                  = 1 if fair or poor health
age             float  %9.0g                  Age in years / 10
income          float  %9.0g                  Income in $ / 1000
chronic         byte   %8.0g                  = 1 if a chronic condition

. summarize fairpoor age income chronic
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    fairpoor |      4412    .0972348    .2963108          0          1
         age |      4412    4.083454    1.022545        2.5        6.4
      income |      4412    34.34018    29.03987    -49.999    280.777
     chronic |      4412    .3263826    .4689423          0          1

. logit fairpoor age income chronic, nolog
Logistic regression                               Number of obs   =       4412
                                                  LR chi2(3)      =     315.97
                                                  Prob > chi2     =     0.0000
Log likelihood = -1249.2851                       Pseudo R2       =     0.1123
------------------------------------------------------------------------------
    fairpoor |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .2957192   .0517768     5.71   0.000     .1942385       .3972
      income |  -.0236338    .002916    -8.10   0.000     -.029349   -.0179185
     chronic |    1.42776   .1103937    12.93   0.000     1.211393    1.644128
       _cons |  -3.453267   .2319289   -14.89   0.000    -3.907839   -2.998694
------------------------------------------------------------------------------

. logit fairpoor age income chronic, nolog vce(robust)
Logistic regression                               Number of obs   =       4412
                                                  Wald chi2(3)    =     243.69
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -1249.2851                 Pseudo R2       =     0.1123
------------------------------------------------------------------------------
             |               Robust
    fairpoor |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .2957192   .0502612     5.88   0.000     .1972091    .3942294
      income |  -.0236338   .0036903    -6.40   0.000    -.0308666    -.016401
     chronic |    1.42776   .1096906    13.02   0.000     1.212771     1.64275
       _cons |  -3.453267   .2295709   -15.04   0.000    -3.903217   -3.003316
------------------------------------------------------------------------------
```

```
. probit fairpoor age income chronic, nolog
Probit regression                                 Number of obs   =       4412
                                                  LR chi2(3)      =     306.78
                                                  Prob > chi2     =     0.0000
Log likelihood = -1253.8799                       Pseudo R2       =     0.1090
------------------------------------------------------------------------------
    fairpoor |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .1508671   .0273036     5.53   0.000      .097353    .2043812
      income |  -.0106309   .0013185    -8.06   0.000    -.0132152   -.0080467
     chronic |   .7329739   .0564237    12.99   0.000     .6223854    .8435623
       _cons |  -1.934104   .1183627   -16.34   0.000     -2.16609   -1.702117
------------------------------------------------------------------------------
```