BGPE Frontiers in Econometrics 2009

A. Colin Cameron, U.C.-Davis

FINAL EXAM (typos corrected)

Open book. Read question carefully so you answer the question.
Keep answers as brief as possible.
Answer 3 of 4 questions

1. Consider IV regression of y_i on $\mathbf{x}_i$ with instruments $\mathbf{z}_i$, leading to IV estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$. Suppose the true model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{v}$.

(a) Suppose $\mathbf{X}$ and $\mathbf{Z}$ are nonstochastic (constants) and $\mathbf{v} \sim [\mathbf{0}, \boldsymbol{\Omega}]$.
Find $\mathrm{E}[\widehat{\boldsymbol{\beta}}]$.

(b) Find $\mathrm{V}[\widehat{\boldsymbol{\beta}}]$ given the same assumptions as in part (a).

(c) Now suppose $\mathbf{X}$ and $\mathbf{Z}$ are stochastic (random) and plim $N^{-1}\mathbf{Z}'\mathbf{X}$, plim $N^{-1}\mathbf{Z}'\mathbf{W}$, and plim $N^{-1}\mathbf{X}'\mathbf{W}$ exist and are finite nonzero matrices, and plim $N^{-1}\mathbf{Z}'\mathbf{v} = \mathbf{0}$. Find plim $\widehat{\boldsymbol{\beta}}$.

(d) Given your answer to part (b), suggest a way to estimate the variance matrix of $\widehat{\boldsymbol{\beta}}$ in the case that $\boldsymbol{\Omega}$ is a diagonal matrix with i^{th} entry ω_{ii}. A brief explanation will do.

(e) This part unrelated to parts (a)-(d).
Which requires that stronger assumptions be placed on the stochastic properties of x_i in order to make statements about the asymptotic behavior of $X_N = N^{-1}\sum_{i=1}^{N} x_i$ – a law of large numbers or a central limit theorem? Provide an explanation.

2. For parts (a)-(d) consider the following model.
Consider the panel data model

$$y_{it} = \alpha_i + \mathbf{w}_i'\boldsymbol{\gamma} + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it},$$

where $\alpha_i \sim$ i.i.d. $[0, \sigma_\alpha^2]$ and $\varepsilon_{it} \sim$ i.i.d. $[0, \sigma_\alpha^2]$.

(a) Suppose we estimate this model by OLS with default OLS standard errors. What problem(s), if any, do you foresee?

(b) Suppose we estimate this model by the FE (within) estimator with default FE standard errors. What problem(s), if any, do you foresee?

(c) Suppose we estimate this model by the RE estimator with default RE standard errors. What problem(s), if any, do you foresee?

(d) Suppose we perform a Hausman test that compares the FE and RE estimates of $\boldsymbol{\beta}$. We obtain a large value of the test statistic, in excess of the relevant chi-square critical value. What do we conclude?

The following parts are unrelated to parts (a)-(d).

(e) Consider the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where $\mathrm{E}[\mathbf{u}|\mathbf{X}] \neq \mathbf{0}$, but there exist variables $\mathbf{Z}$ (where $\mathbf{Z}$ is of larger dimension than $\mathbf{X}$) that satisfy $\mathrm{E}[\mathbf{u}|\mathbf{Z}] = \mathbf{0}$. Give the objective function for the GMM estimator when the weighting matrix is $(\mathbf{Z}'\mathbf{Z})^{-1}$.

(f) Give the formula for the estimator that minimizes this objective function. Your derivation can be brief.

(g) If the errors are heteroskedastic is the estimator in (f) the most efficient estimator? If not, provide a more efficient estimator.

3.(a) Consider a binary variable y that takes value 1 with probability $\exp(\mathbf{x}_i'\boldsymbol{\beta})$ and probability 0 with probability $1 - \exp(\mathbf{x}_i'\boldsymbol{\beta})$. Give the log-likelihood function.

(b) What problem(s), if any, do you see with the model in part (a).

(c) Consider a multinomial logit model for variable y that takes values 1, 2 and 3. Give the formulae for the probabilities when the explanatory variables $\mathbf{x}_i$ are individual-specific and do not vary across alternatives.

(d) What specialization of the additive random utility model leads to the model in part (c)?

(e) Suppose we observe

$$y_1 = \begin{cases} 1 & \text{if } y_1^* > 0 \\ 0 & \text{if } y_1^* \leq 0. \end{cases}$$

and

$$y_2 = \begin{cases} y_2^* & \text{if } y_1^* > 0 \\ 0 & \text{if } y_1^* \leq 0. \end{cases}$$

where

$$\begin{aligned} y_1^* &= \mathbf{x}_1'\boldsymbol{\beta}_1 + u_1 \\ y_2^* &= \mathbf{x}_2'\boldsymbol{\beta}_2 + u_2. \end{aligned}$$

Find $\mathsf{E}[y_2|y_1^* > 0, \mathbf{x}]$ if $u_2 = \rho u_1 + v$, where v is independent of u_1 and $\mathrm{E}[u_1|u_1 > c] = g(c)$ for some specified function $g(c)$.

(f) Suppose the estimator $\boldsymbol{\beta}$ minimizes $\sum_{i=1}^{N}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^4$. Note that the fourth power is used here. Obtain the first-order conditions for the estimator. What essential condition do you think is needed for the estimator to be consistent?

4. For parts (a)-(e) consider the following Stata output.

```
. use mus10data.dta, clear
. quietly keep if year02==1
. describe docvis income
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
docvis          int    %8.0g                  number of doctor visits
income          float  %9.0g                  Income in $ / 1000
. generate d_docvis = docvis > 0
. generate r_docvis = d_docvis
. replace r_docvis = 2 if docvis > 10
. summarize docvis d_docvis r_docvis income
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      docvis |      4412    3.957389    7.947601          0        134
    d_docvis |      4412    .6359927    .4812052          0          1
```

```
    r_docvis |      4412    .7336809    .6251864          0          2
      income |      4412    34.34018    29.03987    -49.999    280.777
. tabulate r_docvis
   r_docvis |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      1,606       36.40       36.40
          1 |      2,375       53.83       90.23
          2 |        431        9.77      100.00
------------+-----------------------------------
      Total |      4,412      100.00
. probit d_docvis income, vce(robust)
Probit regression                                 Number of obs   =       4412
                                                  Wald chi2(1)    =      74.14
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -2838.4719                 Pseudo R2       =     0.0188
------------------------------------------------------------------------------
             |               Robust
    d_docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .0074536   .0008656     8.61   0.000      .005757    .0091502
       _cons |   .1014859    .033279     3.05   0.002     .0362602    .1667116
------------------------------------------------------------------------------
. mfx
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
  income |   .0027895      .00032    8.67   0.000   .002159    .00342   34.3402
------------------------------------------------------------------------------
. margeff
------------------------------------------------------------------------------
    d_docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .0027474    .000309     8.89   0.000     .0021417    .0033531
------------------------------------------------------------------------------
```

(a) Give the specific formula for the quantity that was computed by command `mfx` after command `probit`.

(b) Give the specific formula for the quantity that was computed by command `margeff` after command `probit`.

(c) If you estimated the same model by command `regress`, would you expect to get similar results to those from command `probit`? Explain.

(d) Question was bad: `private` and `chronic` were not in regression.

(e) Suppose we wish to estimate a model for variable `r_docvis`. Which model do you suggest we estimate? Explain.

(f) What problem(s) do the simulation results indicate exist for the probit estimator and related statistical inference?

(g) Why did these problem(s) arise in this simulation? What aspect(s) of the model was misspecified?