FINAL EXAM
BGPE Frontiers in Econometrics 2011
A. Colin Cameron, U.C.-Davis
Open book. 90 minutes.
Answer 3 of 4 questions.
Read question carefully so you answer the question.
Keep answers as brief as possible.

1. Linear estimation and asymptotic theory.
Consider the OLS estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ when $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ and $\text{plim}\, N^{-1}\mathbf{X}'\mathbf{X} = \mathbf{A}$, $\text{plim}\, N^{-1}\mathbf{X}'\mathbf{u} = \mathbf{0}$, $N^{-1/2}\mathbf{X}'\mathbf{u} \xrightarrow{d} \mathcal{N}[\mathbf{0}, \mathbf{B}]$.

(a) Obtain $\text{plim}\, \widehat{\boldsymbol{\beta}}$.

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

(c) State how you would consistently estimate the variance matrix of the limit distribution you obtained in part (b) when u_i is independent and conditionally (on $\mathbf{x}_i$) heteroskedastic.

(d) Suppose that in fact $\mathbf{u} = \mathbf{X}\boldsymbol{\delta} + \mathbf{v}$ where $\text{plim}\, N^{-1}\mathbf{X}'\mathbf{v} = \mathbf{0}$. Obtain $\text{plim}\, \widehat{\boldsymbol{\beta}}$.

(e) Suppose we find that $\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \\ \widehat{\boldsymbol{\beta}}_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$ with $\widehat{\text{V}}[\widehat{\boldsymbol{\beta}}] = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$.
Test $H_0 : \beta_2 - \beta_3 = 0$ against $H_0 : \beta_2 - \beta_3 \neq 0$ at significance level 0.05.

(f) Consider OLS regression of the model $y_i = \beta x_i + u_i$ where x_i is set in an experiment and u_i is an i.i.d. $(0, \sigma^2)$ error. Could we use the Lindeberg-Levy central limit theorem to obtain the limit distribution of $\sqrt{N}(\widehat{\beta} - \beta)$? Explain your answer.

2. IV and panel data
Consider the first set of output on the final page, which comes from model $y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i$ where y_i is `ldruxexp` (log drug expenditure) and $\mathbf{x}_i$ is `hi_empunion` (health insurance through employer or union) and `age` (age in years). Additional the variables `firmsz` (firm size) and `multlc` (multiple firm locations) are used.

(a) Under what properties of u_i would you prefer the output in column OUTPUT1 to that in column OUTPUT2?

(b) What additional analysis in Stata would you do to see whether this was the case?

(c) Provide an explanation for why the estimated standard error of `hi_empunion` is so much higher in column OUTPUT2 than in column OUTPUT1.

(d) Why might you prefer the output in column OUTPUT3 to that in column OUTPUT2?

(e) Provide the mathematical formula for the estimator in column OUTPUT2.

..... Question continues on next page .....

Now consider the second set of output on the final page, which comes from model $y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}$ where y_{it} is `wage` (the level of hourly wage) and $\mathbf{x}_{it}$ is `wks` (annual weeks worked) and `ed` (years of schooling).

(f) Explain under what circumstances the output in column OUTPUT5 is preferred to that in column OUTPUT4.

(g) Explain under what circumstances the output in column OUTPUT6 is preferred to that in column OUTPUT5.

(h) Explain the purpose of option `vce(robust)` used to obtain the results in OUTPUT6.

3. Nonlinear. Consider the following density for the continuous positive random variable y

$$f(y) = \exp(-\theta/y)\theta^2(1/y)^3/2; \quad y > 0,\ \theta > 0,$$

where it can be shown that

$$\mathsf{E}[y] = \theta;\ \mathsf{V}[y] = \infty;\ \mathsf{E}[1/y] = 2/\theta;\ \mathsf{V}[1/y] = 2/\theta^2.$$

Suppose we have a random sample $(y_i, \mathbf{x}_i)$, $i = 1, ...N$, where $\mathbf{x}_i$ is a $k \times 1$ nonstochastic regressor vector and y_i has the above density with

$$\theta_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}),$$

where $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ in the data generating process.

(a) Obtain the log-likelihood function.

(b) Give the first-order conditions for the MLE $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$.

(c) Obtain the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ assuming the density is correctly specified. Obtain this as easily as you can.

(d) Does consistency of $\widehat{\boldsymbol{\beta}}$ in this example require correct specification of the density of y, or can consistency be obtained under weaker assumptions on y? Explain your answer. A brief explanation will do and there is no need to apply a LLN or CLT.

(e) Give the Newton-Raphson algorithm for computing $\widehat{\boldsymbol{\beta}}$ in this specific example.

(f) Give the formula for the average marginal effect on $\mathsf{E}[y|\mathbf{x}]$ of a change in the j^{th} regressor.

4. Simulation in a nonlinear model.
Consider the output from generated data given on the last page.
NOTE: The command `poisson` yields the MLE for $y_i|\mathbf{x}_i$ Poisson distributed with mean $\exp(\mathbf{x}_i'\boldsymbol{\beta})$.

(a) Does the slope estimator appear to be unbiased? Explain your answer.

(b) Do the standard errors appear to be correctly calculated? Explain your answer.

(c) What value do you expect to find for the mean of `reject2f`?

(d) Provide Stata code that would determine whether the Poisson MLE was consistent.

(e) Suppose we are using real data, rather than generated data, and command `poisson` did not report standard errors for $\widehat{\boldsymbol{\beta}}$. Provide with some detail a simulation-type method to obtain the correct standard error assuming independent observations.

(f) Suppose Stata does not have a command `poisson`. Is there another way to program up the Poisson MLE in Stata? A short answer will do.

```
**** OUTPUT FOR QUESTION 2 parts (a)-(c) ****
. correlate ldrugexp hi_empunion ssiratio multlc
(obs=10391)
             | ldrugexp hi_emp~n ssiratio   multlc
-------------+------------------------------------
    ldrugexp |   1.0000
 hi_empunion |   0.0345   1.0000
    ssiratio |   0.0611  -0.1963   1.0000
      multlc |  -0.0489   0.1191  -0.1744   1.0000
. quietly regress ldrugexp hi_empunion age, vce(robust)
. estimates store OUTPUT1
. quietly ivregress 2sls ldrugexp (hi_empunion = multlc ssiratio) age, vce(robust)
. estimates store OUTPUT2
. quietly ivregress gmm ldrugexp (hi_empunion = multlc ssiratio) age, vce(robust)
. estimates store OUTPUT3
. estimates table OUTPUT1 OUTPUT2 OUTPUT3, b(%9.4f) se
----------------------------------------------------
    Variable |  OUTPUT1     OUTPUT2     OUTPUT3
-------------+--------------------------------------
 hi_empunion |    0.1050     -1.0524     -1.0478
             |    0.0278      0.1623      0.1622
         age |    0.0039     -0.0088     -0.0087
             |    0.0021      0.0028      0.0028
       _cons |    6.1459      7.5390      7.5339
             |    0.1578      0.2571      0.2569
----------------------------------------------------
                                       legend: b/se


**** OUTPUT FOR QUESTION 2 parts (a)-(c) ****
. quietly regress wage wks ed, noheader vce(cluster id)
. estimates store OUTPUT4
. quietly xtreg wage wks ed, re vce(robust)
. estimates store OUTPUT5
. quietly xtreg wage wks ed, fe vce(robust)
. estimates store OUTPUT6
. estimates table OUTPUT4 OUTPUT5 OUTPUT6, b(%9.4f) se
----------------------------------------------------
    Variable |  OUTPUT4     OUTPUT5     OUTPUT6
-------------+--------------------------------------
         wks |    4.2499      0.0819     -0.7968
             |    1.8025      1.2714      1.4696
          ed |   59.6670     59.6154   (omitted)
             |    5.1873      5.0526
       _cons |  -82.4569    113.3130    920.2310
             |  111.8610     87.6570     68.7932
----------------------------------------------------
                                       legend: b/se
```

```
**** OUTPUT FOR QUESTION 4 ****
. program simpoisson, rclass
  1.     version 10.1
  2.     drop _all
  3.     set obs $numobs
  4.     generate double x = rnormal(0,1)
  5.     generate mu = exp(-2 + 1*x)
  6.     generate double y = rpoisson(mu)
  7.     poisson y x
  8.     return scalar b2 = _b[x]
  9.     return scalar se2 = _se[x]
 10.     return scalar t2 = (_b[x]-1)/_se[x]
 11.     return scalar r2 = abs(return(t2))>invttail($numobs-1,.025)
 12. end
. global numobs 50
. set seed 10101
. quietly simulate b2f=r(b2) se2f=r(se2) t2f=r(t2) reject2f=r(r2),  ///
>    reps(1000) saving(chi2datares, replace) nolegend nodots: simpoisson

. summarize b2f se2f t2 reject2f
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         b2f |      1000    1.022277    .4063652  -.6024085   3.722705
        se2f |      1000    .3620188     .124872   .1252486   1.401412
         t2f |      1000    -.024073    1.016284  -2.829997   2.699198
    reject2f |      1000       ???       .2095899      0          1
. mean b2f se2f t2 reject2f
Mean estimation                     Number of obs    =    1000
--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
         b2f |   1.022277   .0128504        .99706    1.047494
        se2f |   .3620188   .0039488        .35427    .3697677
         t2f |   -.024073   .0321377     -.0871382    .0389922
    reject2f |       ???       --              --          --
--------------------------------------------------------------
```