240A Cameron Winter 2006
Department of Economics, U.C.-Davis

Final Exam: March 21

Compulsory. Closed book. Worth 50% of course grade.
Read question carefully so you answer the question.
Lengthy Exam. Keep answers as brief as possible.

Question scores (total 50 points)

Question	1a	1b	1c	2a	2b	2c	3a	3b	3c	3d	3e	3f	4a	4b	4c	4d	5a	5b	5c	5d	5e	5f	5g
Points	4	2	2	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

1. Various topics.

(a) Consider OLS regression in the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where $\mathbf{X}$ is $N \times k$ and $\mathbf{u} \sim \mathcal{N}[\mathbf{0},\, \sigma^2\mathbf{I}]$. Suppose we wish to predict the actual value of y given $\mathbf{x} = \mathbf{x}^*$. Give a 95% confidence interval for this prediction.

(b) Suppose we regress $\mathbf{y}$ on $\mathbf{X}_1$ only yielding OLS estimator $\widehat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$. In fact the true model is $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}$, where $\mathbf{u} \sim \mathcal{N}[\mathbf{0}, \sigma^2\mathbf{I}]$. Find $\mathrm{E}[\widehat{\boldsymbol{\beta}}_1]$. What do you conclude?

(c) Consider the usual multiple regression model with intercept, which can be written as $\mathbf{y} = \beta_1\mathbf{l} + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}$, where $\mathbf{l}$ is an $N \times 1$ vector of ones, $\mathbf{X}_2$ is $N \times (k-1)$. Assume $\mathbf{u} \sim \mathcal{N}[\mathbf{0},\, \sigma^2\mathbf{I}]$. By subtracting means of regressors, an equivalent transformed model is

$$\mathbf{y} = \alpha_1\mathbf{l} + \mathbf{X}_2^*\boldsymbol{\beta}_2 + \mathbf{u}, \text{ where } \alpha_1 = \beta_1 + \bar{\mathbf{x}}_2'\boldsymbol{\beta}_2,$$

where $\mathbf{X}_2^* = \mathbf{X}_2 - \mathbf{l}\bar{\mathbf{x}}_2'$, with $\bar{\mathbf{x}}_2' = \frac{1}{N}\mathbf{l}'\mathbf{X}_2$ a $1 \times (k-1)$ vector of means of the regressors. Write this transformed model as $\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}$, for appropriate $\mathbf{Z}$ and $\boldsymbol{\gamma}$, use the result that $\mathbf{X}_2^{*\prime}\mathbf{l} = \mathbf{0}$, and the general result that $\widehat{\boldsymbol{\gamma}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$ to show that OLS regression in the transformed model yields, after some algebra, gives estimates

$$\begin{bmatrix} \widehat{\alpha}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \bar{y} \\ (\mathbf{X}_2^{*\prime}\mathbf{X}_2^*)^{-1}\mathbf{X}_2^{*\prime}\mathbf{y}. \end{bmatrix}$$

2. Consider OLS regression in the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where $\mathbf{X}$ is $N \times k$ and $\mathbf{u} \sim \mathcal{N}[\mathbf{0},\, \sigma^2\mathbf{I}]$. Consider test of $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ against $H_a : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$, where R is $q \times k$.

(a) Suppose σ^2 is known. Given knowledge of the distribution of the OLS estimator in this setting, derive a test statistic for H_0 and gives its distribution under H_0.

(b) Use this test statistic to test $H_0 : \beta_1 = 2\beta_2$ against $H_a : \beta_1 \neq 2\beta_2$ at level 0.05 when $\widehat{\beta}_1 = 5$, $\widehat{\beta}_2 = 2$, $\sigma^2 = 0.1$ and $\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$. What do you conclude?

(c) Suppose now that σ^2 is unknown. Propose a variation to your test statistic in part (a) that can be used if $N \to \infty$.

3. Various topics

(a) Give a mathematical definition for convergence in probability.

(b) Give a mathematical definition for a multivariate central limit theorem.

(c) Give the formula for the IV estimator and state any desirable properties of this estimator.

(d) Give the formula for the GLS estimator and state any desirable properties of this estimator.

(e) For regression with heteroskedasticity give the formula that permits valid inference based on the OLS estimator.

(f) Find the estimator that minimizes $(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{Z}\mathbf{Z}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$, where $\mathbf{X}$ is $N\times k$, $\mathbf{Z}$ is $N\times m$, $m>k$ and both matrices are of full column rank.

4. Consider the estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\mathbf{y}$, where $\mathbf{A}$ is a symmetric $N\times N$ 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, $\boldsymbol{\beta}$ is a $k\times 1$ vector and $\mathbf{u}$ is an $N\times 1$ vector. We assume that $\mathbf{X}$ is fixed, and $\mathbf{u}\sim[\mathbf{0},\ \boldsymbol{\Sigma}]$ where $\boldsymbol{\Sigma}\neq\sigma^2\mathbf{I}$.
For asymptotic theory there is no need to use laws of large numbers and central limit theorems. Instead just use $\operatorname{plim}\frac{1}{N}\mathbf{X}'\mathbf{A}\mathbf{X}$ exists, $\operatorname{plim}\frac{1}{N}\mathbf{X}'\mathbf{A}\mathbf{u}=\mathbf{0}$ and $\frac{1}{\sqrt{N}}\mathbf{X}'\mathbf{A}\mathbf{u}\xrightarrow{d}\mathcal{N}[\mathbf{0},\ \mathbf{B}]$.

(a) Obtain $\mathrm{E}[\widehat{\boldsymbol{\beta}}]$ and $\mathrm{V}[\widehat{\boldsymbol{\beta}}]$.

(b) Show whether or not $\widehat{\boldsymbol{\beta}}$ is consistent.

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

(d) Your answer in part (c) will depend in the matrix $\mathbf{B}$. Give the likely form of $\mathbf{B}$ in terms of $\mathbf{X}$, $\mathbf{A}$ and $\boldsymbol{\Sigma}$. Hence give the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ in terms of $\mathbf{X}$, $\mathbf{A}$ and $\boldsymbol{\Sigma}$.

5. The following analyzes data on Median Housing Rents and related data for 58 California Counties from the 2000 Census. The variables are:

rent = Monthly rent (median in the county) for Occupied Housing in 2000 in $
value = Value (median in the county) of an Occupied House in the county in 2000 in $
vacrate = Percentage of Rental Housing Units in the county unoccupied in 2000
hhsize = number of persons per household in Occupied Housing in the county
percrent = Percentage of Occupied Housing in the county that is rental (rather than nonrental)

Key Stata output is provided on the last page. All three OLS regressions report heteroskedastic-robust standard errors.

(a) Consider the first regression. Show that the predicted value of rent when value equals it s sample average is equal to the sample average of rent. Are you surprised? Explain.

(b) Suppose we wish to do a one-sided test of the claim that rent increases with household size. Provide the p-value for this test and state whether or not we reject the claim at five percent.

(c) Does the second model appear to be an improvement on the first model with just value as a regressor? Give two explanations, aside from a formal F-test.

(d)(i) Suppose you wanted to do the F-test of the second model versus the first. What Stata command would you use: give as much detail as possible.
(ii) Give the complete Stata command that produces the estimates for the first regression output.

(e) Test at level 0.05 the hypothesis that the elasticity of house rent with respect to house value equals 1.
[Hint: Do this the easiest possible way].

(f) For these data provide two distinct methods that you would use, aside from a formal hypothesis test of heteroskedasticity, to determine whether the errors are heteroskedastic in OLS regression of level of rent on level of value? [Note: that the data provided does not answer this question: you need to propose two methods].

(g) Suppose heteroskedasticity is determined to be present and we assume that V[rent | value] = $\sigma^2 \times$value^2. Provide a simple method to obtain parameter estimates in the model rent = $\beta_1 + \beta_2$value that are more efficient than the OLS estimates.

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      county |         0
        rent |        58    668.2414    180.9138        429       1185
       value |        58    179420.7     96920.5      72900     493300
     vacrate |        58     5.62931    3.427086        1.8       20.9
      hhsize |        58    2.720172    .3093428       2.29       3.33
    percrent |        58    36.57241    7.425839       21.3         65
     logrent |        58    6.472257    .2509759   6.061457   7.077498
    logvalue |        58      11.982     .464486   11.19684   13.10887

Linear regression                                      Number of obs =      58
                                                       F(  1,    56) =  274.18
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.8983
                                                       Root MSE      =  58.219

             |               Robust
        rent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       value |   .0017691   .0001068    16.56   0.000     .0015551    .0019831
       _cons |   350.8251   17.23109    20.36   0.000     316.3071    385.3431

Linear regression                                      Number of obs =      58
                                                       F(  4,    53) =  136.15
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9194
                                                       Root MSE      =  53.276

             |               Robust
        rent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       value |   .0018117   .0001071    16.92   0.000     .0015969    .0020265
     vacrate |  -2.003118   2.654439    -0.75   0.454    -7.327249    3.321014
      hhsize |   69.26762   35.20702     1.97   0.054     -1.34872     139.884
    percrent |  -2.904163   1.310289    -2.22   0.031    -5.532269   -.2760572
       _cons |   272.2524   90.60301     3.00   0.004     90.52576     453.979

Linear regression                                      Number of obs =      58
                                                       F(  1,    56) =  610.32
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9082
                                                       Root MSE      =  .07671

             |               Robust
     logrent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    logvalue |   .5149396   .0208438    24.70   0.000     .4731844    .5566947
       _cons |   .3022529   .2487881     1.21   0.230    -.1961295    .8006354
```