240A Cameron Winter 2007
Department of Economics, U.C.-Davis

Midterm Exam: February 7

Compulsory. Closed book. Worth 32% of course grade.
Read question carefully so you answer the question.
Keep answers as brief as possible.

Question scores (total 50 points)

Question	1a	1b	1c	1d	2a	2b	2c	3a	3b	3c	3d	3e	3f	4a	4b	5a	5b	5c	6a	6b	6c
Points	2	2	2	2	2	2	2	2	2	2	4	2	2	2	2	2	2	4	2	4	4

1. A sample of size four has (x, y) equal to $(2, 4)$, $(0, 0)$, $(2, 2)$, and $(0, 2)$. We wish to estimate the model $y = \beta_1 + \beta_2 x_2 + u$. We assume $u \sim \mathcal{N}[0, \sigma^2]$ and independent.

(a) Using the matrix algebra formula for the OLS estimator, compute the OLS coefficients $\widehat{\beta}_1$ and $\widehat{\beta}_2$ for these data.

(b) Compute the standard error for $\widehat{\beta}_2$ for these data.

(c) Compute R^2 for these data, using any method you like (matrix algebra is not required).

(d)(i) Give the Stata command(s) that you would use to manually read the data for this example into Stata.
(ii) Give the Stata command that you would use to compute $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

2. Consider the multiple regression model

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{v},$$

where $\mathbf{y}$ is an $N \times 1$ vector, $\mathbf{Z}$ is an $N \times k$ matrix of constants, $\boldsymbol{\gamma}$ is a $k \times 1$ vector and $\mathbf{v}$ is an $N \times 1$ vector with $\mathbf{v} \sim \mathcal{N}[0, \sigma^2 \mathbf{I}]$.
Consider the estimator $\widetilde{\boldsymbol{\gamma}}$ that minimizes

$$Q(\boldsymbol{\gamma}) = \mathbf{v}'\mathbf{A}\mathbf{v},$$

for nonsingular symmetric positive definite constant matrix $\mathbf{A}$.

(a) Show that $\widetilde{\boldsymbol{\gamma}} = (\mathbf{Z}'\mathbf{A}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{A}\mathbf{y}$.

(b) Derive the variance of $\widetilde{\boldsymbol{\gamma}}$.

(c) Derive $\mathrm{E}[\widetilde{\mathbf{v}}\widetilde{\mathbf{v}}']$, where $\widetilde{\mathbf{v}} = \mathbf{y} - \mathbf{Z}\widetilde{\boldsymbol{\gamma}}$ and state whether $\mathrm{E}[\widetilde{\mathbf{v}}\widetilde{\mathbf{v}}'] = \sigma^2\mathbf{I}$.

3. Consider the estimator

$$\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$$

in the multiple regression model

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

where $\mathbf{y}$ is an $N \times 1$ vector, $\mathbf{X}$ and $\mathbf{Z}$ are $N \times k$ matrices, $\boldsymbol{\beta}$ is a $k \times 1$ vector and $\mathbf{u}$ is an $N \times 1$ vector.
We assume that the regressors $\mathbf{X}$ and $\mathbf{Z}$ are random with components $\mathbf{x}_i$ and $\mathbf{z}_i$ that are iid.
We assume that the error components u_i are iid with mean 0 and variance σ^2 and that $\mathbf{u}$ is independent of $\mathbf{X}$ or $\mathbf{Z}$.
For parts (a)-(d) there is no need to explicitly use laws of large numbers and central limit theorems. Instead state explicitly any additional assumptions that you have made.

(a) Show whether or not $\widehat{\boldsymbol{\beta}}$ is unbiased for $\boldsymbol{\beta}$.
Your answer here should account for $\mathbf{X}$ and $\mathbf{Z}$ being stochastic**.**

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

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

(d) Give the asymptotic distribution for $\widehat{\boldsymbol{\beta}}$.
Your answer should depend only on σ^2, $\mathbf{Z}$ and $\mathbf{X}$.

(e) Suppose we have the same setup as above, except now $k = 1$ so there is just one regressor. Apply a law of large numbers to obtain the probability limit of $N^{-1}\sum_i z_i u_i$.
Provide as much detail as possible, including verifying that a LLN applies.

(f) Suppose we have the same setup as above, except now $k = 1$ so there is just one regressor. Apply a central limit theorem to obtain the limit distribution of $N^{-1/2}\sum_i z_i u_i$.
Provide as much detail as possible, including verifying that a CLT applies.

4. Consider testing q linear restrictions of the form

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

in the linear regression model.

(a) Suppose we have a model with four regressors (the first of which is an intercept). Give $\mathbf{R}$ and $\mathbf{r}$ in each of the following situations
(i) $H_0 : \beta_2 = \beta_3 + 2$.
(i) $H_0 : \beta_2 = 0, \beta_3 = 0, \beta_4 = 0$.

(b) Suppose we have an estimator $\widetilde{\boldsymbol{\beta}} \overset{a}{\sim} \mathcal{N}[\boldsymbol{\beta}, \mathbf{V}]$.
Derive the chisquared form of the Wald test statistic of $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$.

5. Consider the following Stata input and output

```
program simols, rclass
    drop _all
    set obs 5
    gen x = 1 + 2*invnorm(uniform())
    gen y = 1 + 2*x + (invnorm(uniform()))^2 - 1
    regress y x
    return scalar theta =_b[x]
end
simulate "simols" theta=r(theta), reps(1000)
summarize, detail
```

with results

```
. summarize, detail
                          r(theta)
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .3500591      -.1988292
 5%     1.289173      -.1873094
10%     1.534989      -.0204161       Obs                1000
25%     1.823596       .0668301       Sum of Wgt.        1000

50%     2.006446                      Mean           2.016591
                        Largest       Std. Dev.      .6211563
75%     2.177897       4.594079
90%     2.460031       4.769485       Variance       .3858351
95%     2.763496       4.857336       Skewness       6.850154
99%     3.622856       13.72522       Kurtosis       130.4659
```

(a) Give the data generating process in as much detail as possible.

(b) Does the simulation setup correspond to regressors fixed in repeated samples or does it correspond to random sampling of both y and x? Explain. [This is tricky. You need to think hard about this.]

(c) Are the simulation results what you expect? Explain in as much detail as possible.

6. Consider the attached Stata output that uses data on family income and food expenditure from the 1998 Consumer Expenditure Survey.

Dependent Variable

Food = Annual family expenditure on food in 1998 in dollars.

Regressors

Income = Annual family income in 1998 in dollars.

Incomesq = Square of Income.

Famsize = number of members in family (averaged over year so can be non-integer)

Drural = 1 if live in rural (non-urban) area and 0 if urban

(a) Are the regressors jointly statistically significant at level 0.05? Explain.

(b) Does food expenditure appear to be linear in income? Perform **two** appropriate statistical significance tests at level 0.05.

(c) Test at level 0.05 the claim that one more family member is associated with a $700 increase in food expenditures. State clearly details of the test and your conclusion.
[ASIDE: I had meant to ask this as a one-sided test question.]

```
. summarize food income incomesq famsize drural

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        food |       700    4897.343    2855.653        585      20765
      income |       700    41118.78    33096.46        150   192123.1
    incomesq |       700    2.78e+09    4.69e+09      22500   3.69e+10
     famsize |       700    2.469429    1.414257          1          9
      drural |       700    .1285714    .3349643          0          1

. regress food income incomesq famsize drural

      Source |       SS       df       MS              Number of obs =     700
-------------+------------------------------           F(  4,   695) =  134.68
       Model |  2.4890e+09     4   622261100           Prob > F      =  0.0000
    Residual |  3.2111e+09   695  4620327.67           R-squared     =  0.4367
-------------+------------------------------           Adj R-squared =  0.4334
       Total |  5.7002e+09   699   8154752.7           Root MSE      =  2149.5

------------------------------------------------------------------------------
        food |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .0425326   .0071027     5.99   0.000     .0285873    .0564779
    incomesq |  -5.47e-08   4.87e-08    -1.12   0.262    -1.50e-07    4.09e-08
     famsize |    805.701   61.51642    13.10   0.000     684.9207    926.4813
      drural |   -14.3438    246.752    -0.06   0.954    -498.8126     470.125
       _cons |   1313.055   212.3991     6.18   0.000     896.0338    1730.076
------------------------------------------------------------------------------
```