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

Final Exam: March 16

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)
Essentially 2 points each part, aside from questions 3 and 4.

Question	1a-c	2a-c	3a-3g	4a-4e	5a-g
Points	6	6	14	10	14

1. Consider the linear regression model

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

where $\mathbf{X}$ is an $N \times k$ full rank regressor matrix. Let $\widehat{\boldsymbol{\beta}}$ denote the OLS estimator in this model. Suppose we rescale the regressors as

$$\mathbf{Z} = \mathbf{X}\mathbf{P},$$

where $\mathbf{P}$ is a $k \times k$ nonsingular matrix and estimate

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

and let $\widehat{\boldsymbol{\gamma}}$ denote the OLS estimator in this model.

(a) Show that $\widehat{\boldsymbol{\gamma}} = \mathbf{P}^{-1}\widehat{\boldsymbol{\beta}}$.

(b) Obtain the relationship between the OLS residuals in the two regressions
i.e., compare $\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$ to $\widehat{\mathbf{v}} = \mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\gamma}}$.

(c) Consider the special case where $\mathbf{Z}$ is obtained from $\mathbf{X}$ by multiplying the last regressor (the k^{th} column of $\mathbf{X}$) by 100, while the other regressors are unchanged. Define explicitly the matrix $\mathbf{P}$ in this special case, and state explicitly the relationship between $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$ in this special case.

2. For the following question on OLS regression of $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ use the results that

$$\begin{aligned} \bar{R}^2 &= 1 - \frac{\text{Residual Sum of Squares}/(N-k)}{\text{Total Sum of Squares}/(N-1)} \\ F &= \frac{(\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}} - \widehat{\mathbf{u}}'\widehat{\mathbf{u}})/q}{\widehat{\mathbf{u}}'\widehat{\mathbf{u}}/(N-k)} \end{aligned}$$

where $\widehat{\mathbf{u}}$ is unrestricted OLS residual, $\widetilde{\mathbf{u}}$ is restricted OLS least squares residual, and there are q restrictions (so the degrees of freedom in the two models differ).

(a) Show that $\bar{R}^2$ does not change if s, the standard error of the regression, is the same in the unrestricted and the restricted models.

(b) Obtain F when s, the standard error of the regression, is the same in the unrestricted and the restricted models.

(c) What do you conclude from your analysis in parts (a) and (b)?

3. Parts (b) and (d) are 3 points each and remaining parts are 2 points each.

(a) Suppose that health status is an explanatory variable that takes three possible values, leading to three mutually exclusive indicator variables
$d1 = 1$ if health is poor (and $= 0$ otherwise)
$d2 = 1$ if health is fair (and $= 0$ otherwise)
$d3 = 1$ if health is good (and $= 0$ otherwise).
State how, if at all, estimated coefficients from the OLS regression: $y = \alpha + \beta x + \gamma d2 + \phi d3 + u$ will differ from estimated coefficients from the OLS regression: $y = a + bx + cd1 + fd3 + v$.

(b) OLS regression on a large sample yields $\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 10 \\ 15 \end{bmatrix}$ and $\widehat{\mathrm{V}} \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$.
Test $H_0 : 2\beta_1 = \beta_2$ against $H_a : 2\beta_1 \neq \beta_2$ at level 0.05.

(c) Suppose $\mathrm{E}[\mathbf{u}|\mathbf{X}] \neq \mathbf{0}$ in the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$.
Propose a method to obtain a consistent estimator of $\boldsymbol{\beta}$ and give the formula for this estimator.

(d) Consider the linear regression model $\mathbf{y}$ where $\mathbf{y}$ has joint density

$$f(\mathbf{y}) = [(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}]^{-1} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

and $\boldsymbol{\Sigma}$ is known. Find the MLE for $\boldsymbol{\beta}$ and its asymptotic variance matrix.

(e) Suppose $y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i$ where the error u_i is iid with mean 0 and variance $\sigma^2 z_i$, where z_i is an observed positive scalar.
Provide a method to obtain an efficient estimator of $\boldsymbol{\beta}$ that can be implemented using only OLS commands.

(f) Suppose $y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t$ where the error u_t follows an AR(1) process.
Provide a method to obtain an efficient estimator of $\boldsymbol{\beta}$ that can be implemented using only OLS commands.

4. Parts (b) and (c) are 3 points each and other parts are 2 points each.

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 $N \times N$ 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 $\mathbf{u}|\mathbf{Z} \sim [\mathbf{0},\ \boldsymbol{\Sigma}]$ where $\boldsymbol{\Sigma} \neq \sigma^2\mathbf{I}$.
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].

5. The following analyzes the Academic Performance Index (API) which is a score for each California school determined by the performance of all students at the school on a standardized test. The lowest possible school score is 200 and the highest possible score is 1000. We consider data for 804 high schools in California in 1999 on the API and on factors that may effect school performance on the API. The variables are:

api99	Academic Performance Index
avg_ed	Average Parent Education Level measured in years of schooling
pact_meals	Percent of students in Free or Reduced Price Lunch Program
pct_el	Percent English Language Learners (i.e. English as second language)
yr_rnd	1 if school is Multi-Track Year-Round School and 0 otherwise
pct_cred	Percent Fully Credentialed Teachers
pact_emer	Percent Emergency Credentialed Teachers (i.e. less qualified)

Use Stata commands and output given on the next page to answer parts (a)-(f).

(a) State which regressors in the full regression are statistically significant at 5 percent.
Do any of these statistically significant regressors have unexpected sign? Explain.

(b) The claim is made that one more year of parental education is associated with more than a 50 point rise in school API. Test this hypothesis at 5 percent using full model results. State your conclusion.

(c) Which model fits the data better? The full model or the reduced model with just avg_ed as a regressor? Explain.

(d) What Stata command(s) would you use to decide whether full model or the reduced model is better? Provide details.

(e) Given the results do you expect that robust standard errors will differ much from the standard errors given here?

(f) Looking at the page of output, what basic fundamental conclusion do you make about the determinants of school API score? Explain.

(g) Consider the following Stata code. What quantities are being displayed by the final command? There is no need to calculate. Just describe what `d` and `g` are measuring.

```
quietly regress api99 avg_ed pct_meal pct_el yr_rnd pct_cred pct_emer
matrix b = e(b)'
matrix V = e(V)
matrix c = (12,20,20,0,80,10,1)'
matrix d = c'*b
matrix e = c'*V*c
scalar g = sqrt(e[1,1])
display d,g
```

```
. summarize api99 avg_ed pct_meal pct_el yr_rnd pct_cred pct_emer

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       api99 |       804    620.9851     107.619        355        966
      avg_ed |       804    14.84199    1.236061      11.62         18
    pct_meal |       804    21.88308    23.65456          0         98
      pct_el |       804    14.00373    12.78206          0         66
      yr_rnd |       804    .0236318    .1519938          0          1
    pct_cred |       804    89.86567    8.419089         33        100
    pct_emer |       804    10.43781    8.198646          0         56

. regress api99 avg_ed

      Source |       SS       df       MS              Number of obs =     804
-------------+------------------------------           F(  1,   802) = 4059.52
       Model |  7765973.98     1  7765973.98           Prob > F      =  0.0000
    Residual |  1534247.84   802  1913.02723           R-squared     =  0.8350
-------------+------------------------------           Adj R-squared =  0.8348
       Total |  9300221.82   803  11581.8454           Root MSE      =  43.738

       api99 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      avg_ed |   79.56099   1.248713    63.71   0.000     77.10986    82.01212
       _cons |  -559.8584   18.59747   -30.10   0.000    -596.3639   -523.3529

. regress api99 avg_ed pct_meal pct_el yr_rnd pct_cred pct_emer

      Source |       SS       df       MS              Number of obs =     804
-------------+------------------------------           F(  6,   797) =  769.74
       Model |  7931492.35     6  1321915.39           Prob > F      =  0.0000
    Residual |  1368729.48   797  1717.35191           R-squared     =  0.8528
-------------+------------------------------           Adj R-squared =  0.8517
       Total |  9300221.82   803  11581.8454           Root MSE      =  41.441

       api99 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      avg_ed |    73.8491   1.886623    39.14   0.000     70.14577    77.55244
    pct_meal |   .0815515    .081216     1.00   0.316    -.0778711    .2409741
      pct_el |  -.3703318   .1676255    -2.21   0.027    -.6993714   -.0412923
      yr_rnd |   26.09651   10.19847     2.56   0.011     6.077471    46.11556
    pct_cred |   .3959038   .3112565     1.27   0.204    -.2150755    1.006883
    pct_emer |  -1.479399   .3152027    -4.69   0.000    -2.098125   -.8606737
       _cons |  -492.4344   42.49818   -11.59   0.000    -575.8559   -409.0128

. estat hettest avg_ed pct_meal pct_el yr_rnd pct_cred pct_emer
         chi2(6)      =    43.52
         Prob > chi2  =   0.0000
```