Assignment 1: OLS and GLS (b11)

A. Colin Cameron U.C.-Davis

Data are at http://cameron.econ.ucdavis.edu/bgpe2011

1. Use **2001 data** in file `musdata10.dta` **IMPORTANT:** `keep if year01==1`

(a) Use commands `describe` and `summarize` to describe and summarize the data.

Which of the variables `private`, `chronic`, `female`, and `income` is most highly correlated with `docvis`? Use command `correlate`.

(b) Perform OLS regression of `docvis` on `private`, `chronic`, `female`, and `income`, obtaining heteroskedastic-robust standard errors.

(c) Which variables are statistically significant at level 0.01 using two-sided tests?

(d) Test the joint statistical significance at level 0.01 of regressors `private` and `chronic` using command `test`.

(e) Do the heteroskedastic-robust standard errors differ much from default standard errors?

(f) State in simple words the impact of private insurance on doctor visits.

(g) Use command `mfx` with `eyex` to obtain the income elasticity of doctor visits, evaluated at the sample mean of the regressors. Is this a large or small effect?

(h) Command `egen x=mean(x)` creates a variable with all observations equal to the mean of variable `x`. Use this command to create demeaned variables $x_i - \bar{x}$ for `income` and $y_i - \bar{y}$ for `docvis`. Use command `regress` to show that OLS of $y_i - \bar{y}$ on $y_i - \bar{y}$ without an intercept is equivalent to OLS of y_i on x_i with an intercept.

2. 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 both $\mathbf{X}$ and $\mathbf{Z}$ are fixed (for simplicity), and $\mathbf{u} \sim [\mathbf{0},\ \sigma^2\mathbf{I}]$.
For asymptotic theory there is no need to use laws of large numbers and central limit theorems. Instead just use plim $\frac{1}{N}\mathbf{Z}'\mathbf{X}$ exists, plim $\frac{1}{N}\mathbf{Z}'\mathbf{u} = \mathbf{0}$ and $\frac{1}{\sqrt{N}}\mathbf{Z}'\mathbf{u} \xrightarrow{d} \mathcal{N}[\mathbf{0},\ \mathbf{B}]$.
To answer this question adapt the proof for OLS.

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

(b) Obtain the limit distribution of $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. [Your derivation can be brief].

(c) Your answer in part (b) will depend in the matrix $\mathbf{B}$. Give the likely form of $\mathbf{B}$ in terms of σ^2, $\mathbf{X}$ and $\mathbf{Z}$. Hence give the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ in terms of σ^2, $\mathbf{X}$ and $\mathbf{Z}$.

3. Consider the estimator $\widetilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$ in the linear 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 of constants, $\mathbf{W}$ is a symmetric $N \times N$ matrix of constants , $\boldsymbol{\beta}$ is a $k \times 1$ vector and $\mathbf{u}$ is an $N \times 1$ vector, and it is assumed that $\mathbf{u}$ has mean $\mathbf{0}$ and variance matrix $\boldsymbol{\Omega} = \text{Diag}[\sigma_i^2]$, so the error term is heteroskedastic.

(a) Show that $\widetilde{\boldsymbol{\beta}}$ is unbiased, stating clearly any assumptions.

(b) Obtain the variance matrix of $\widetilde{\boldsymbol{\beta}}$, stating clearly any assumptions.

(c) Prove that $\widetilde{\boldsymbol{\beta}}$ is consistent, stating clearly any assumptions. [Your derivation can be brief].

(d) Obtain the limit distribution of $\sqrt{N}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$, stating clearly any assumptions. [Your derivation can be brief].

(e) Hence give the asymptotic distribution of $\widetilde{\boldsymbol{\beta}}$.

(f) State how to obtain a consistent estimate of the asymptotic variance of $\widetilde{\boldsymbol{\beta}}$ even if there is no consistent estimate of $\boldsymbol{\Omega}$.

4. Do the same simulation using exactly the same code as that given in file `ct_gls.do` with the following single change:
replace

```
generate y = 1 + 2*x + rchi2(1)-1          // demeaned chi^2 error
```

with

```
generate y = 1 + 2*x + abs(x)*(rchi2(1)-1)
```

(a) What is the d.g.p. for this simulation? Give the model for y_i and the model for the error u_i.

(b) Does the simulation show that the OLS estimator is consistent here? Explain.

(c) Does the simulation show that the default OLS standard errors are correct here? Explain.

(d) Do the default OLS standard errors lead to tests of the correct size? Explain.

(e) Now repeat the same exercise but compute robust standard errors. Do these standard errors lead to tests of the correct size? Explain.

(f) Do the robust standard errors lead to tests of the correct size? Explain.