Econ 836 Final Exam

RULES: No notes or other source materials are permitted. All such items must remain on the floor. You may use only: pens, pencils, rulers, erasers, calculators and watches. Be sure to carefully write down all assumptions that you use to answer the question, including any calculations that you may make. Write your NAME on the front page of your exam booklet. Show your work, including numerical calculations. The exam is out of 40 points.

1. [4 points] Let

$$Y = X\beta + \varepsilon$$
$$E[X'\varepsilon] = 0_K$$
$$E[\varepsilon\varepsilon'] = \sigma^2 I_N$$

Consider the "ridge estimator" $\widehat{\beta} = \left(X'X + \lambda I_K\right)^{-1} X'Y$. Note that

$$\left(X'X + \lambda I_K +\right)^{-1} = \left(X'X\right)^{-1}\left(I_K + \lambda\left(X'X + \lambda I_K\right)^{-1}\right)$$
$$= \left(X'X\right)^{-1}\left(I_K + G\right)$$

where $G = \lambda\left(X'X + \lambda I_K\right)^{-1}$ and is symmetric.

What is the bias and variance of this estimator?

a) bias.

$$E\left[\widehat{\beta}\right] = E\left[\left(X'X + \lambda I_K\right)^{-1} X'Y\right] = E\left[\left(X'X\right)^{-1}(I+G)X'Y\right]$$
$$= E\left[(I+G)\left(X'X\right)^{-1} X'Y\right] = E\left[(I+G)\left(X'X\right)^{-1} X'X\beta\right] + E\left[(I+G)\left(X'X\right)^{-1} X'\varepsilon\right]$$
$$= E[(I+G)\beta] + \left[(I+G)\left(X'X\right)^{-1}\right]E[X'\varepsilon] = [(I+G)\beta]$$

$bias = G\beta$

b) variance

$$E\left[\left(\widehat{\beta} - E\left[\widehat{\beta}\right]\right)^2\right] = E\left[\left((I+G)\left(X'X\right)^{-1} X'\varepsilon\varepsilon' X\left(X'X\right)^{-1}(I+G)\right)\right]$$
$$= \sigma^2 I_N\left((I+G)\left(X'X\right)^{-1} X'X\left(X'X\right)^{-1}(I+G)\right)$$
$$= \sigma^2\left((I+G)\left(X'X\right)^{-1}(I+G)\right)$$

2. [4 points] Suppose you have an unbiased estimator of a parameter vector. Suppose also you have estimated the parameter vector with a very large sample, and, invoking a central limit theorem, have found the sampling distribution of your estimated parameters to be

$$\hat{\beta} = \begin{bmatrix} \widehat{\beta_1} \\ \widehat{\beta_1} \end{bmatrix} \sim N\left(\begin{bmatrix} -2 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix}\right).$$

a) Test the hypothesis that the two parameters are equal to each other.

i) d=-2-2=-4
ii) V(a-b)=v(a)+v(b)-2cov(a,b)=4+9-2=11
iii) t=-4/sqrt(11) ~ -4/3.5 is distributed normally under the null. this is less than 2, so don't reject

b) Use a Wald Test Statistic to test the hypothesis that $\beta_1 = \beta_2 = -1$. Do you reject the hypothesis?
   i) joint hypothesis.

$$H_0 : R\beta + r = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \beta + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0$$

$$T = \left(R\hat{\beta} + r \right)' \left(R V(\hat{\beta})' R \right)^{-1} \left(R\hat{\beta} + r \right) \sim \chi_2^2$$

   ii) $T = \begin{bmatrix} -1 & 3 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} -1 \\ 3 \end{bmatrix}$

$$= \begin{bmatrix} -1 & 3 \end{bmatrix} \left(\begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix} \right)^{-1} \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} -1 & 3 \end{bmatrix} \left(\begin{bmatrix} 9/35 & -1/35 \\ -1/35 & 4/35 \end{bmatrix} \right) \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \left(\begin{bmatrix} -12/35 & 13/3 \end{bmatrix} \right) \begin{bmatrix} -1 \\ 3 \end{bmatrix} = 51/35$$

   iii) about 1.5, much less than the crit for a chi2 with 2 df.

c) Construct a test statistic for the hypothesis that $\beta_1 + \beta_2 = -2$. Do you reject the hypothesis?
   i) d=2
   ii) v(d)=v(a+b)=v(a)+v(b)+2cov(a,b)=4+9+2=15
   iii) d/sqrt(v(d))=2/4=0.5, is less than 2 the crit for a normal.

d) If the hypothesis in b) is true, then the hypothesis in c) is true. Why are test statistics different?
   i) one is a normal, the other is a chi-square.

3. [4 points] Suppose you wanted to know the variance of a LAD (Least Absolute Deviations, i.e., conditional median) estimator, but didn't know the formula for it. Write Stata code for a Monte Carlo experiment that would give an estimate of the variance of the estimated LAD coefficients. If you cannot write Stata code, then (for partial credit) you may write out how you would implement this more generically.

```
use "some data set.dta"
* x is a regressor, y is a regressand
qreg y x, q(0.5)
matrix coefs=e(b)
* could write "LAD y x" or "median y x" or whatever they think the stata command is.
forval i=1/1000 {
    qreg y x, q(0.5)
```

```
    matrix coefs=coefs\e(b)
}
* could loop any way they want; could get the coefs any way they want
svmat coefs_, matrix(coefs)
* could use any command they want to turn the matrix of coefficients by iteration into a pair of series
summ coefs*
*look at the standard deviations of the constant and x coefficient
```

any halfway good attempt at writing stata code is okay here.

4. [4 points] Suppose you have an AR1 regression model with a time trend:

$$Y_t = X_t\beta + \delta t + \varepsilon_t$$
$$\varepsilon_t = \rho\varepsilon_t + u_t$$
$$u_t \sim N(0,\sigma_u^2)$$

   a) Let $\rho = 0.5$. What is the covariance between the error term in period *t* and the error term in period *t-3*?
      i) $$\varepsilon_t = \rho\varepsilon_{t-1} + u_t = \rho\left(\rho\varepsilon_{t-2} + u_{t-1}\right) + u_t = \left(\rho^2\varepsilon_{t-2} + \rho u_{t-1}\right) + u_t = \rho^3\varepsilon_{t-3} + \rho^2 u_{t-2} + \rho u_{t-1} + u_t$$
      $$E\left(\varepsilon_t\varepsilon_{t-3}\right) = E\left(\varepsilon_t\left(\rho^3\varepsilon_{t-3} + \rho^2 u_{t-2} + \rho u_{t-1} + u_t\right)\right) = \rho^3\sigma_\varepsilon^2$$
      ii) can also write out in terms of the variance of u:
      iii) $\rho^3\sigma_\varepsilon^2 = \rho^3\frac{\sigma_u^2}{1-\rho^2}$
   b) Describe the Newey-West variance matrix for the estimated coefficients (how does it differ from the standard OLS variance matrix).
      i) $V(\widehat{\beta}_{OLS}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$ where Ω is the expectation of the outer product of the errors.
      ii) You can get an estimate of Ω by regressing the residual on its lag (or, in this case, you can use our prior knowledge that this coefficient is 0.5).
      iii) Using this, construct the entire matrix Ω, and stick it in to the variance estimate.
      iv) if they propose a GLS estimator rather than a robust variance, and do it right, give them the point.
   c) Suppose $\rho = 1$. What is the variance of ε?
      i) it is infinite in the limit. if $\rho = 1$, then $\varepsilon_t = \sum_{s=1}^{t} u_t$ and $V\left(\varepsilon_t\right) = t\sigma^2$
   d) Suppose $\rho = 1$. How might you estimate the parameters β?
      i) you could difference the equation, which would yield

$$\Delta Y_t = \Delta X_t\beta + \delta + \Delta\varepsilon_t = \delta + \Delta X_t\beta + u_t$$
$$u_t \sim N(0,\sigma_u^2)$$

this model has a constant term, and normally distributed errors

5. [4 points] Consider the binary choice probit model:

$$Y_i^* = X_i\beta + \varepsilon_i$$

$$Y_i = 1 \;\; if \;\; Y_i^* > 0$$

$$\varepsilon_i \sim N(0,1)$$

a) Write out the log of the likelihood for a single observation of *Y,X*.

$$P[Y=1 \mid X] = P\left[X\beta + \varepsilon > 0\right] = P\left[\varepsilon > -X\beta\right] = 1 - \Phi(-X\beta)$$

i) $P[Y=0 \mid X] = P\left[X\beta + \varepsilon <= 0\right] = P\left[\varepsilon <= -X\beta\right] = \Phi(-X\beta)$

$$\ln L_i = I(Y_i = 1) * \ln\left(1 - \Phi(-X_i\beta)\right) + (1 - I(Y_i = 1)) * \Phi(-X_i\beta)$$

ii) anything approaching the above is okay here.

b) Write out the first-order condition of maximum likelihood approach to estimating the parameters of this model.

i) FOC, take the derivative of the likelihood over all observations with respect to the parameter vector

$$\ln L = \sum_{i=1}^{N} Y_i \left[\ln\left(1 - \Phi(-X_i\beta)\right)\right] + (1 - Y_i)\ln\Phi(-X_i\beta)$$

ii) $FOC:$

$$\sum_{i=1}^{N} \frac{X_i Y_i \phi(-X_i\beta)}{\left(1 - \Phi(-X_i\beta)\right)} - \frac{X_i(1 - Y_i)\phi(-X_i\beta)}{\Phi(-X_i\beta)}$$

iii) anything approaching the above is okay here.

c) Suppose *X* has 10 columns, and that one of those columns is a dummy variable equal to 1 for men, and that another of those columns is a dummy variable equal to 1 for women. You cannot run the regression with all 10 columns of *X*---you have to drop one of those dummy variables. Does it matter which one you drop? Why or why not?

i) it doesn't matter. it just shifts the constant term.

d) What is the probability that $|\varepsilon_i|$ (the absolute value of the error) is bigger than 2.56?

i) check the distribution of the normal. this is the 99% critical value. So, that probability is 1%.

6. [20 points] Long Stata Question

```
. *Notes: p* are logged commodity prices, varying by province and year; z3 and z8-z22 are
demographic dummy variables; totcucon and totexpend are measures of total consumption; num* are
numbers of bedrooms and bathrooms; s* are nominal expenditures; s3 is nominal expenditure on
shelter; p3 is the logged price of rental shelter. z1 is age of HH head minus 40, and z2 is its
square; z4 is year  minus 2002, and z5 is its square; z6, z7 are heating and cooling degrees less
their averages.

. replace totcucon=ln(totcucon)
. replace totexpen=ln(totexpen)
. summ totcucon

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
```

```
    totcucon |    155713    10.40319    .6543996   4.127134   12.73976

. replace totcucon=totcucon-r(mean)
. summ totexpen

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    totexpen |    155713    10.67065    .7489869   4.127134   13.29606

. replace totexpen=totexpen-r(mean)
. summ totcucon totexpen

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    totcucon |    155713   -2.50e-16    .6543996   -6.27606   2.336567
    totexpen |    155713   -1.56e-16    .7489869  -6.543515   2.625408

. replace mortgage_rate=mortgage_rate-6
. replace unemployment=unemployment-6
. g totcucon2=totcucon^2
. g totexpen2=totexpen^2
. rename z9   Married
. rename z10  Married_w_children
. rename z11  Married_non_children
. rename z12  Single_parent
. tab redurent

          reduced rent reason |      Freq.     Percent        Cum.
------------------------------+-----------------------------------
                   not rented |    111,335       65.49       65.49
government subsidized housing |      6,833        4.02       69.51
                        other |      6,007        3.53       73.04
              no reduced rent |     45,834       26.96      100.00
------------------------------+-----------------------------------
                        Total |    170,009      100.00

. tab redurent, nol

    reduced |
rent reason |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    111,335       65.49       65.49
          1 |      6,833        4.02       69.51
          2 |      6,007        3.53       73.04
          3 |     45,834       26.96      100.00
------------+-----------------------------------
      Total |    170,009      100.00

. g renter=redurent==3
. g real_rent=s3/exp(p3)

. global v1 "p1-p13 z1-z8 z13-z22 totcucon totcucon2 totexpen totexpen2 i.yearbuip i.repairs
i.numrmp i.numbedrp "
. global v2 "Married Married_w_children Married_non_children Single_parent pricerural
pricebigcity mortgage_rate unemployment"

. xi: regress renter $v1 $v2
i.yearbuip        _Iyearbuip_1-6      (naturally coded; _Iyearbuip_1 omitted)
i.repairs         _Irepairs_1-3       (naturally coded; _Irepairs_1 omitted)
i.numrmp          _Inumrmp_1-11       (naturally coded; _Inumrmp_1 omitted)
i.numbedrp        _Inumbedrp_0-5      (naturally coded; _Inumbedrp_0 omitted)

      Source |       SS       df       MS              Number of obs =  155710
-------------+------------------------------           F( 65,155644) = 1197.80
       Model |   10175.138    65  156.540585           Prob > F      =  0.0000
    Residual |  20341.1096155644   .13068997           R-squared     =  0.3334
-------------+------------------------------           Adj R-squared =  0.3332
       Total |  30516.2477155709  .195982555           Root MSE      =  .36151

------------------------------------------------------------------------------------
              renter |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
```

p1	-.4250508	.0699921	-6.07	0.000	-.5622338	-.2878678
p2	.0918609	.0761752	1.21	0.228	-.0574409	.2411626
p3	.0725983	.0191935	3.78	0.000	.0349795	.1102171
p4	-.0781673	.0045453	-17.20	0.000	-.087076	-.0692586
p5	.0178939	.0094488	1.89	0.058	-.0006255	.0364133
p6	-.3269421	.0665116	-4.92	0.000	-.4573035	-.1965808
p7	.1876591	.0726753	2.58	0.010	.045217	.3301011
p8	.1194037	.0433886	2.75	0.006	.0343629	.2044445
p9	.1570672	.0264089	5.95	0.000	.1053064	.2088281
p10	-.0800533	.1582483	-0.51	0.613	-.3902166	.2301101
p11	.1806891	.1220854	1.48	0.139	-.0585958	.4199739
p12	.028451	.0441631	0.64	0.519	-.0581078	.1150097
p13	-.4795147	.0708287	-6.77	0.000	-.6183374	-.340692
z1	-.0065164	.0000751	-86.82	0.000	-.0066635	-.0063693
z2	.0000935	3.39e-06	27.54	0.000	.0000868	.0001001
z3	-.0073339	.001908	-3.84	0.000	-.0110735	-.0035943
z4	.002194	.0040367	0.54	0.587	-.0057179	.0101058
z5	.0014288	.000196	7.29	0.000	.0010446	.0018129
z6	.0794922	.0200525	3.96	0.000	.0401896	.1187947
z7	.0071868	.0020723	3.47	0.001	.0031251	.0112486
z8	.0948562	.003082	30.78	0.000	.0888155	.1008969
z13	-.0408637	.0137139	-2.98	0.003	-.0677427	-.0139847
z14	.0984101	.0145855	6.75	0.000	.0698228	.1269974
z15	.0517597	.0139958	3.70	0.000	.0243282	.0791911
z16	.0786231	.0141981	5.54	0.000	.0507951	.1064511
z17	.066218	.0143897	4.60	0.000	.0380145	.0944214
z18	.085232	.014752	5.78	0.000	.0563183	.1141457
z19	.0976285	.0156266	6.25	0.000	.0670007	.1282563
z20	-.0025549	.0025507	-1.00	0.317	-.0075541	.0024444
z21	.0030458	.0047638	0.64	0.523	-.0062911	.0123826
z22	-.0956527	.0047608	-20.09	0.000	-.1049838	-.0863217
totcucon	.0802352	.0056915	14.10	0.000	.0690801	.0913904
totcucon2	.0120164	.0040115	3.00	0.003	.004154	.0198788
totexpen	-.086619	.0051452	-16.83	0.000	-.0967035	-.0765345
totexpen2	-.0473276	.0033326	-14.20	0.000	-.0538594	-.0407959
_Iyearbuip_2	-.0080039	.003412	-2.35	0.019	-.0146913	-.0013164
_Iyearbuip_3	.020008	.0034361	5.82	0.000	.0132732	.0267428
_Iyearbuip_4	-.0205382	.0031744	-6.47	0.000	-.02676	-.0143164
_Iyearbuip_5	-.0531177	.0034381	-15.45	0.000	-.0598563	-.0463792
_Iyearbuip_6	-.0672618	.003535	-19.03	0.000	-.0741904	-.0603333
_Irepairs_2	.0125471	.0037751	3.32	0.001	.0051479	.0199462
_Irepairs_3	.013889	.0032909	4.22	0.000	.0074389	.0203391
_Inumrmp_2	-.0344876	.01627	-2.12	0.034	-.0663765	-.0025987
_Inumrmp_3	-.040378	.0176436	-2.29	0.022	-.0749591	-.0057969
_Inumrmp_4	-.0884664	.0180487	-4.90	0.000	-.1238415	-.0530913
_Inumrmp_5	-.2303025	.0182371	-12.63	0.000	-.2660468	-.1945583
_Inumrmp_6	-.3180163	.018358	-17.32	0.000	-.3539977	-.282035
_Inumrmp_7	-.3579739	.0184514	-19.40	0.000	-.3941383	-.3218095
_Inumrmp_8	-.374259	.0185453	-20.18	0.000	-.4106074	-.3379105
_Inumrmp_9	-.3810745	.0187062	-20.37	0.000	-.4177384	-.3444107
_Inumrmp_10	-.3815659	.0189693	-20.11	0.000	-.4187453	-.3443864
_Inumrmp_11	-.3806187	.0191042	-19.92	0.000	-.4180624	-.3431749
_Inumbedrp_1	.0258885	.0137893	1.88	0.060	-.0011382	.0529151
_Inumbedrp_2	-.0560486	.0146689	-3.82	0.000	-.0847993	-.027298
_Inumbedrp_3	-.1445476	.014921	-9.69	0.000	-.1737924	-.1153027
_Inumbedrp_4	-.1447876	.0151487	-9.56	0.000	-.1744786	-.1150965
_Inumbedrp_5	-.1360985	.0156154	-8.72	0.000	-.1667043	-.1054926
Married	-.0840302	.0142953	-5.88	0.000	-.1120487	-.0560118
Married_w_children	-.1145185	.0144266	-7.94	0.000	-.1427943	-.0862427
Married_non_children	-.0610549	.0151267	-4.04	0.000	-.090703	-.0314068
Single_parent	-.0191209	.0142697	-1.34	0.180	-.0470892	.0088475
pricerural	-.2895449	.0263577	-10.99	0.000	-.3412055	-.2378842
pricebigcity	.2278674	.0233883	9.74	0.000	.1820269	.2737079
mortgage_rate	.0029473	.0027494	1.07	0.284	-.0024416	.0083361
unemployment	.0076227	.0010662	7.15	0.000	.005533	.0097125
_cons	.5751035	.0145626	39.49	0.000	.546561	.603646

. *NOTE: the regression above is an OLS regression

6a) Why are there only 155710 observations in this regression, when the first table shows 170,009 observations in the data file?
---some regressors have missing values, and these observations are excluded from the regression

6b) Why is the coefficient on p3 (the logged price of rental shelter) positive? Is this a problem?
---it says that the probability of renting is higher if rent prices are higher. this is weird. but, remember that it is conditional on all the other regressors. also, it is a small effect: if the price of renting goes up 10%, then the probability of renting goes up 0.7 percentage points.
---any noncrazy answer is okay here.

6c) The estimated constant term is .5751035. What is the meaning of this coefficient?
--- this is the estimated probability of renting for a person whose other characteristics are 0, and facing the prices in ON 2002.

6d) What is the estimated difference in the probability of renting for a person aged 50 in comparison to a person aged 40?
--- age is normed to 40, so z1=z2=0 for the 40 year old. they equal 10 and 100 for a 50 year old.
10*-.0065164 + 100* .0000935 =-0.0652+0.0093 = -0.0559.
they are 5.59 percentage points less likely to rent.

```
. xi: probit renter $v1 $v2
i.yearbuip        _Iyearbuip_1-6      (naturally coded; _Iyearbuip_1 omitted)
i.repairs         _Irepairs_1-3       (naturally coded; _Irepairs_1 omitted)
i.numrmp          _Inumrmp_1-11       (naturally coded; _Inumrmp_1 omitted)
i.numbedrp        _Inumbedrp_0-5      (naturally coded; _Inumbedrp_0 omitted)

Iteration 0:   log likelihood = -90442.243
Iteration 1:   log likelihood = -63510.537
Iteration 2:   log likelihood = -62898.881
Iteration 3:   log likelihood = -62898.142
Iteration 4:   log likelihood = -62898.142

Probit regression                                 Number of obs   =     155710
                                                  LR chi2(65)     =   55088.20
                                                  Prob > chi2     =     0.0000
Log likelihood = -62898.142                       Pseudo R2       =     0.3045

------------------------------------------------------------------------------------
              renter |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
                  p1 |  -1.861741   .3077107    -6.05   0.000    -2.464843   -1.258639
                  p2 |   .2738712   .3339234     0.82   0.412    -.3806067     .928349
                  p3 |   .1917399   .0852199     2.25   0.024      .024712    .3587678
                  p4 |  -.3168121   .0196926   -16.09   0.000    -.3554089   -.2782153
                  p5 |   .0687257   .0413857     1.66   0.097    -.0123887    .1498402
                  p6 |  -1.553845   .2924411    -5.31   0.000    -2.127019    -.980671
                  p7 |   .7689789   .3229336     2.38   0.017     .1360407    1.401917
                  p8 |   .5703419   .1909369     2.99   0.003     .1961124    .9445714
                  p9 |   .7037904   .1171386     6.01   0.000      .474203    .9333777
                 p10 |  -.3159757   .6949886    -0.45   0.649    -1.678128    1.046177
                 p11 |   .7171048   .5359026     1.34   0.181     -.333245    1.767455
                 p12 |   .0994901   .1940013     0.51   0.608    -.2807455    .4797256
                 p13 |  -1.964674   .3105833    -6.33   0.000    -2.573406   -1.355942
                  z1 |  -.0231984   .0003062   -75.76   0.000    -.0237985   -.0225982
                  z2 |    .000304   .0000142    21.47   0.000     .0002762    .0003317
                  z3 |  -.0164492   .0084363    -1.95   0.051     -.032984    .0000856
                  z4 |   .0075948   .0178984     0.42   0.671    -.0274855    .0426751
                  z5 |   .0060866   .0008587     7.09   0.000     .0044036    .0077696
```

z6	.3701751	.0883862	4.19	0.000	.1969413	.5434089
z7	.0333142	.0090923	3.66	0.000	.0154936	.0511349
z8	.3026101	.0122407	24.72	0.000	.2786189	.3266014
z13	-.1503483	.0594889	-2.53	0.011	-.2669445	-.0337521
z14	.2683004	.0622179	4.31	0.000	.1463555	.3902453
z15	.1756516	.0602853	2.91	0.004	.0574947	.2938085
z16	.3044273	.061042	4.99	0.000	.1847873	.4240674
z17	.290972	.061983	4.69	0.000	.1694876	.4124564
z18	.3989357	.0637042	6.26	0.000	.2740778	.5237936
z19	.4868377	.0674775	7.21	0.000	.3545842	.6190912
z20	-.0287499	.0110956	-2.59	0.010	-.0504969	-.0070029
z21	.0145866	.0210307	0.69	0.488	-.0266328	.055806
z22	-.4478686	.021627	-20.71	0.000	-.4902567	-.4054806
totcucon	.4330331	.0265588	16.30	0.000	.3809788	.4850873
totcucon2	.1634092	.0192527	8.49	0.000	.1256747	.2011438
totexpen	-.4786799	.0242555	-19.73	0.000	-.5262198	-.43114
totexpen2	-.2865881	.0164655	-17.41	0.000	-.3188598	-.2543163
_Iyearbuip_2	-.021304	.0147519	-1.44	0.149	-.0502172	.0076091
_Iyearbuip_3	.0783734	.0147899	5.30	0.000	.0493857	.1073611
_Iyearbuip_4	-.0689525	.0138735	-4.97	0.000	-.0961441	-.0417609
_Iyearbuip_5	-.2021379	.0152961	-13.21	0.000	-.2321178	-.1721581
_Iyearbuip_6	-.2329569	.0154037	-15.12	0.000	-.2631476	-.2027663
_Irepairs_2	.0312077	.0162036	1.93	0.054	-.0005509	.0629662
_Irepairs_3	.0266158	.0141833	1.88	0.061	-.0011829	.0544145
_Inumrmp_2	-.0996906	.0626949	-1.59	0.112	-.2225703	.023189
_Inumrmp_3	-.1209528	.0679517	-1.78	0.075	-.2541357	.0122301
_Inumrmp_4	-.2612263	.069403	-3.76	0.000	-.3972538	-.1251989
_Inumrmp_5	-.6171201	.0701324	-8.80	0.000	-.7545771	-.479663
_Inumrmp_6	-.926903	.0707678	-13.10	0.000	-1.065605	-.7882007
_Inumrmp_7	-1.139851	.0714445	-15.95	0.000	-1.279879	-.999822
_Inumrmp_8	-1.256591	.0722209	-17.40	0.000	-1.398141	-1.115041
_Inumrmp_9	-1.329916	.0736204	-18.06	0.000	-1.474209	-1.185623
_Inumrmp_10	-1.349727	.0759006	-17.78	0.000	-1.49849	-1.200965
_Inumrmp_11	-1.399929	.0773143	-18.11	0.000	-1.551462	-1.248395
_Inumbedrp_1	.0953498	.0530575	1.80	0.072	-.0086409	.1993405
_Inumbedrp_2	-.1366849	.0562902	-2.43	0.015	-.2470118	-.026358
_Inumbedrp_3	-.4813989	.0574812	-8.37	0.000	-.59406	-.3687378
_Inumbedrp_4	-.5233803	.059133	-8.85	0.000	-.6392788	-.4074817
_Inumbedrp_5	-.4639827	.0626128	-7.41	0.000	-.5867016	-.3412639
Married	-.3529803	.061529	-5.74	0.000	-.4735749	-.2323858
Married_w_children	-.4322799	.061947	-6.98	0.000	-.5536938	-.3108659
Married_non_children	-.2106902	.0651661	-3.23	0.001	-.3384134	-.0829671
Single_parent	-.0996535	.0611975	-1.63	0.103	-.2195985	.0202915
pricerural	-1.152308	.1155756	-9.97	0.000	-1.378832	-.9257841
pricebigcity	1.020807	.1043466	9.78	0.000	.8162918	1.225323
mortgage_rate	.0163281	.0120703	1.35	0.176	-.0073292	.0399855
unemployment	.0348059	.0047158	7.38	0.000	.0255631	.0440488
_cons	.1403869	.0594509	2.36	0.018	.0238654	.2569085

NOTE: the regression above is a PROBIT regression. The probit regression model is given in question 5 above.

6e) The constant term `_cons` is only .1403869 in this PROBIT regression. Why is it so much lower than in the linear regression?

--- the constant term in the probit is not equal to the probability for that person with all characteristics equal to 0. the probability in a probit is the probability that a normal variate is larger than the negative of that value: 1-Phi(-0.14), which is a little over 50%.

```
. test $v2

 ( 1)  [renter]Married = 0
 ( 2)  [renter]Married_w_children = 0
 ( 3)  [renter]Married_non_children = 0
 ( 4)  [renter]Single_parent = 0
 ( 5)  [renter]pricerural = 0
 ( 6)  [renter]pricebigcity = 0
```

```
 ( 7)  [renter]mortgage_rate = 0
 ( 8)  [renter]unemployment = 0

           chi2(  8) =  713.09
         Prob > chi2 =    0.0000
```

6f) The Wald test statistic for the joint hypothesis that all elements of `v2` may be excluded from the probit is 713. What is the (approximate) probability that a chi-squared with 8 df is at least this large?
--- the p-value is given there: about zero.

```
.
. predict ystar, xb
(14299 missing values generated)
. quietly g invmills=0
. quietly replace invmills=normalden(ystar)/normal(ystar) if renter==1
. quietly replace invmills=normalden(-ystar)/normal(-ystar) if renter==0

. summ real_rent invmills if invmills~=.

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   real_rent |    155710    2531.807    4198.979          0   66235.34
    invmills |    155710    .4498801     .443652   .0000214   3.159213

. xi: regress real_rent $v1 invmills if renter==1
i.yearbuip        _Iyearbuip_1-6      (naturally coded; _Iyearbuip_1 omitted)
i.repairs         _Irepairs_1-3       (naturally coded; _Irepairs_1 omitted)
i.numrmp          _Inumrmp_1-11       (naturally coded; _Inumrmp_1 omitted)
i.numbedrp        _Inumbedrp_0-5      (naturally coded; _Inumbedrp_0 omitted)

      Source |       SS       df       MS              Number of obs =   41665
-------------+------------------------------           F( 58, 41606) =  227.21
       Model |  1.4016e+11    58  2.4166e+09           Prob > F      =  0.0000
    Residual |  4.4252e+11 41606  10636035.8           R-squared     =  0.2405
-------------+------------------------------           Adj R-squared =  0.2395
       Total |  5.8269e+11 41664  13985347.8           Root MSE      =  3261.3

------------------------------------------------------------------------------
   real_rent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          p1 |   6412.758   1156.416     5.55   0.000     4146.159    8679.358
          p2 |   5865.057   1130.947     5.19   0.000     3648.378    8081.736
          p3 |  -1495.396    278.081    -5.38   0.000     -2040.44   -950.3509
          p4 |    450.834   81.73091     5.52   0.000     290.6397    611.0283
          p5 |  -203.7347   135.8644    -1.50   0.134    -470.0318    62.56234
          p6 |  -1160.553   1055.021    -1.10   0.271    -3228.416    907.3108
          p7 |  -3860.289   917.3173    -4.21   0.000     -5658.25   -2062.328
          p8 |   808.8198   637.5209     1.27   0.205    -440.7345    2058.374
          p9 |  -3189.231   399.1439    -7.99   0.000    -3971.562   -2406.901
         p10 |   12447.22   2593.313     4.80   0.000     7364.272    17530.17
         p11 |   -7956.06   1999.014    -3.98   0.000    -11874.17   -4037.951
         p12 |  -4518.113   704.4444    -6.41   0.000    -5898.839   -3137.387
         p13 |    1090.39   1076.483     1.01   0.311    -1019.538    3200.319
          z1 |   51.34882   2.412988    21.28   0.000     46.61931    56.07833
          z2 |   .8913327   .0572563    15.57   0.000     .7791091    1.003556
          z3 |   228.3541   33.19848     6.88   0.000     163.2844    293.4238
          z4 |   37.39072   34.31766     1.09   0.276    -29.87261    104.6541
          z5 |  -17.69415   3.401313    -5.20   0.000     -24.3608   -11.02751
          z6 |  -621.2246   335.2545    -1.85   0.064     -1278.33    35.88124
          z7 |  -194.4275   36.25851    -5.36   0.000    -265.4949   -123.3601
          z8 |    598.695   49.27983    12.15   0.000     502.1056    695.2845
         z13 |   68.89311    85.2827     0.81   0.419    -98.26277     236.049
         z14 |   63.04256   84.64613     0.74   0.456    -102.8656    228.9508
         z15 |   13.31058   46.96055     0.28   0.777    -78.73309    105.3543
         z16 |   115.0996   60.25836     1.91   0.056    -3.008082    233.2072
         z17 |   480.1888   72.24274     6.65   0.000     338.5915    621.7861
         z18 |   579.4274   100.7638     5.75   0.000     381.9282    776.9266
```

```
         z19 |   626.5881   145.2798     4.31   0.000     341.8366    911.3395
         z20 |    214.469   43.82614     4.89   0.000     128.5688    300.3692
         z21 |  -83.17109   74.21071    -1.12   0.262    -228.6256    62.28347
         z22 |  -858.2166   100.6331    -8.53   0.000     -1055.46   -660.9736
    totcucon |   986.7603    119.918     8.23   0.000     751.7185    1221.802
   totcucon2 |  -886.7463   83.44161   -10.63   0.000    -1050.294    -723.199
    totexpen |   1615.724   113.8331    14.19   0.000     1392.609    1838.839
   totexpen2 |   740.1769   77.37537     9.57   0.000     588.5195    891.8342
_Iyearbuip_2 |   72.55305   57.55168     1.26   0.207    -40.24945    185.3556
_Iyearbuip_3 |   310.5412   56.56562     5.49   0.000     199.6714     421.411
_Iyearbuip_4 |   370.9285   55.73619     6.66   0.000     261.6844    480.1726
_Iyearbuip_5 |   536.6758   66.30353     8.09   0.000     406.7195    666.6321
_Iyearbuip_6 |   790.7886   66.24002    11.94   0.000     660.9568    920.6204
 _Irepairs_2 |   167.8792   65.67489     2.56   0.011     39.15503    296.6033
 _Irepairs_3 |    346.155   58.26232     5.94   0.000     231.9596    460.3504
  _Inumrmp_2 |  -272.1859   175.0648    -1.55   0.120    -615.3166    70.94473
  _Inumrmp_3 |  -306.7219   190.3589    -1.61   0.107    -679.8294    66.38557
  _Inumrmp_4 |  -95.58319   197.1444    -0.48   0.628    -481.9903    290.8239
  _Inumrmp_5 |   399.8653   207.7898     1.92   0.054    -7.407089    807.1377
  _Inumrmp_6 |   750.1666   225.5383     3.33   0.001     308.1069    1192.226
  _Inumrmp_7 |   850.8006   244.4692     3.48   0.001     371.6358    1329.965
  _Inumrmp_8 |   644.5254   261.0922     2.47   0.014     132.7793    1156.272
  _Inumrmp_9 |    180.525    282.964     0.64   0.523    -374.0904    735.1404
 _Inumrmp_10 |  -320.6706   308.3378    -1.04   0.298    -925.0193     283.678
 _Inumrmp_11 |   254.7544   323.5224     0.79   0.431    -379.3562     888.865
_Inumbedrp_1 |   695.5946   150.0051     4.64   0.000     401.5813    989.6078
_Inumbedrp_2 |   1117.543   162.9194     6.86   0.000     798.2175    1436.868
_Inumbedrp_3 |   1046.668     177.96     5.88   0.000     697.8627    1395.473
_Inumbedrp_4 |   787.7501   196.3819     4.01   0.000     402.8376    1172.663
_Inumbedrp_5 |   1333.201   227.6538     5.86   0.000      886.995    1779.407
    invmills |  -1473.395   159.3335    -9.25   0.000    -1785.692   -1161.098
       _cons |   7993.169    169.671    47.11   0.000      7660.61    8325.727
------------------------------------------------------------------------------
```

6g) The coefficient on the inverse mills ratio (`invmills`) is -1473. How does this relate to the expectation of the error term in the real rent equation?
--- the expectation of the error term is -1473*invmills. that is why the selection correction works. once it is there, the error has an expectation of zero.
---any reasonable answer is okay here.

6h) How much of the variation in real rent is driven by variation in the inverse mills ratio?
the std dev of real rent is $4200. The std dev of the inverse mills ratio is 0.45, so the std dev of that term is 0.45*1473 ~ $700. About one-sixth of the variation in real rents is driven by the selection term.

```
. predict imp_corrected
(option xb assumed; fitted values)
(14299 missing values generated)

. quietly xi: regress real_rent $v1 if renter==1
. predict imp_straight
(option xb assumed; fitted values)
(14299 missing values generated)

. bys renter: summ imp_straight imp_corrected

-----------------------------------------------------------------------------------------------
-------------------------
-> renter = 0

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
imp_straight |    114045    8548.056    1692.305  -15941.08   15460.49
imp_correc~d |    114045    10447.12    2162.089  -11193.64    19094.9
```

```
-----------------------------------------------------------------------------------------
-------------------------
-> renter = 1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
imp_straight |     41665    8090.797    1828.193   -7052.43   15453.17
imp_correc~d |     41665    8090.797    1834.153  -7519.211   15600.79
```

6h) Why are the means of `imp_straight` and `imp_corrected` the same for renters but different for owners?
--- both the straight regression and the selection corrected regression go through the mean rent of renters. but the imputation for owners does not go through the mean of anything. thus, the selection correction can drive it up for owners.
---any non-wrong answer is okay here.

6i) The R-squared for the first OLS regression is 0.3334, but for the OLS regression of `real_rent` on `v1` and `invmills`, it is only 0.2405. Why is it so much lower in the latter regression?
--- these regressions have different LHS variables and different samples. there is no reason to expect their R2s to be similar.

Some probabilities for the standard normal distribution, z.
Prob[-1.65<z<1.65]=0.90, Prob[-1.96<z<1.96]=0.95, Prob[-2.56<z<2.56]=0.99

Standard normal probability density function: $\phi(u) = \exp\left(\frac{-u^2}{2}\right) / \sqrt{2\pi}$

Some probabilities for the chi-square distribution with 1 degree of freedom, x, and 2 degrees of freedom, y:
Prob[x<3.84]=0.95, Prob[x<6.63]=0.99, Prob[y<5.99]=0.95, Prob[x<9.21]=0.99

Mean and Variance Rules for Linear Combinations:
If for some vector *w*: $E(w) = \overline{w}, V(w) = \Sigma$, then for a linear combination *Aw+b*,

we have that $E(Aw+b) = A\overline{w}+b, V(Aw+b) = A\Sigma A'$.

Inverse of a 2x2 matrix:

$$inv\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$