ECONOMETRICS II, Fall, 2018
Bent E. Sørensen

Midterm Exam - October 1, 2018

Each sub-question in the following carries equal weight except when otherwise noted.

1. (26%)
Assume that you have data for consumption c_i and income y_i for a sample $i = 1, ..., N$ and that you want to estimate the relation

$$c_i = \alpha + \beta y_i + \epsilon_i \ ,$$

where ϵ_i is normally distributed with mean 0 and variance σ^2.
a) (3%) If observations with $c_i > 100$ were excluded from the sample, what is this pattern labeled?
b) (10%) And what would be the log likelihood function? (Write it out explicitly.)
c) (3%) If instead observations $c_i > 100$ were replaced with $c_i = 100$, what is the term (label) for the model that takes this into account?
d) (10%) What would be the log likelihood function? (Write it out explicitly.)

2. (24%) On the next pages, I have reproduced a piece of Matlab code that you used for a homework.
a) 8%) Explain what the code is supposed to do.

b) 8%) Explain what the two lines for z(i)==1 are capturing?

c) 8%) Explain intuitively what happens if rho==0.

3) 25%) Consider the stationary AR(2) model

$$x_t = a x_{t-1} + b x_{t-2} + u_t \ ,$$

where u_t is normally distributed with variance σ^2. Assume the variance-covariance matrix for x_t, x_{t-1}, and x_{t-2} is

$$\Sigma_x \ = \ \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 \\ \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_2 & \gamma_1 & \gamma_0 \end{pmatrix} \ .$$

Assume you have a sample of observations x_t, $t = 1, ..., T$. Write down the first three terms of the unconditional likelihood function; i.e., $f(x_1)$, $f(x_2|x_1)$, and $f(x_3|x_1, x_2)$. (You can write the *log*-likelihood, if you want, the main point is the unconditional and conditional distributions appearing in these three terms. Do not try to invert any matrices, I need to see that you know/can derive the formulas.)

4) (25%)
Prove that, in a system of equations (think of 2 equations and a sample of years for simplicity), where the two error terms are correlated,if the right hand side regressors are the same for each dependent variable, you can estimate the equations one-by-one and obtain the same estimator as if you used GLS on the full system. (Assume that observations are independent across years.)

Code

```
function [L] = logl_ss(b)

global x w y z N

b0 = b(1);
b1 = b(2);
sigmau = b(3);
g0 = b(4);
g1 = b(5);
rho = b(6);

XB = b0*ones(size(x,1),1) + b1*x ;
WG = g0*ones(size(w,1),1) + g1*w ;

L=0 ;

for i = 1:N
  if z(i) == 1
    L = L + log((1/sigmau)*normpdf((y(i) - XB(i))/sigmau))...
            + log(normcdf((1/sqrt(abs(1-rho^2)))*(WG(i) + (rho/sigmau)*(y(i) - XB(i)))));
  elseif z(i) == 0
    L = L + log(normcdf(-WG(i)));
  end
end

L = -L;

end
```