

## Midterm Exam II—October 25, 2021

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

1. (20%) Assume you have  $T$  observations ( $t = 1, \dots, T$ ) for two equations

$$y_t = \alpha_{10} + \alpha_{11}y_{t-1} + \alpha_{12}w_{t-1} + e_t,$$

and

$$w_t = \alpha_{20} + \alpha_{21}y_{t-1} + \alpha_{22}w_{t-1} + u_t,$$

where  $y$  and  $w$  are some variables. The  $\alpha$ s are parameters that we want to estimate by the best linear estimator (taking the initial values as given) and  $e$  and  $u$  are error terms. Assume the variance of  $e$  is  $\sigma_e^2$ , the variance of  $u$  is  $\sigma_u^2$ , and the covariance of  $e$  and  $u$  is  $\sigma_{eu}$ .

- Explain how we estimate the  $\alpha$  coefficients using the simplest efficient estimator.
- Explain how we estimate  $\sigma_e^2$ ,  $\sigma_u^2$ , and  $\sigma_{eu}$ .
- Assume that the equation for  $y$  still holds but

$$w_t = \alpha_{20} + \alpha_{21}y_{t-1} + \alpha_{22}w_{t-1} + \beta x_t + u_t,$$

where  $x$  is an exogenous regressor. How would you now estimate the coefficients (the  $\alpha$ s and  $\beta$ ) using an efficient estimator?

2. (20%) Consider the model:

$$h_i = c + e_i,$$

$$w_i = d + f h_i + u_i,$$

and

$$y_i = g + k h_i + l w_i + v_i,$$

and where  $h$ ,  $w$ , and  $y$  are endogenous random variables,  $c, d, f, g, k$  and  $l$  are unknown parameters (assume they are non-zero), and  $e$ ,  $u$ , and  $v$  are independent white noise terms.

a) Explain for which (if any) of these equations one can estimate consistently (i.e., the parameters are identified).

b) Now assume that  $u$  and  $v$  are correlated. Which, if any, of the coefficients in the three equations are identified?

3. (20%) This code simulates the simultaneous equation model

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 x_1 + u_1,$$

$$y_2 = \beta_4 + \beta_5 y_1 + \beta_7 x_2 + u_2,$$

with  $U \sim NID(0, \Sigma)$ . Write down the 2SLS estimators using the terms from the code (the places are indicated in the code).

## Set the parameters.

There are 1000 observations. Set  $\beta_0 = 0.1$ ,  $\beta_1 = 0.5$ ,  $\beta_2 = 0.3$ ,  $\beta_4 = 0.2$ ,  $\beta_5 = 0.3$ ,  $\beta_7 = 1$ ,  $\sigma_1 = 1.2$ ,  $\sigma_2 = 1.1$  and  $\rho = 0.05$ .

```
clear

clc

N = 1000;                                % Number of observations.

beta0 = 0.1;
beta1 = 0.5;
beta2 = 0.3;
beta4 = 0.2;
beta5 = 0.3;
beta7 = 1;

sigma1 = 1.2;
sigma2 = 1.1;
rho = 0.05;

mu = [0 0];                               % Mean vector of U.
smat = [sigma1^2 rho*sigma1*sigma2;       % Variance matrix of U.
        rho*sigma1*sigma2 sigma2^2];
```

```

G_inv = ones(2,2); % G^{-1} matrix.
G_inv(2,1) = beta1;
G_inv(1,2) = beta5;
G_inv = G_inv./(1-beta1*beta5);

B = zeros(3,2); % B matrix.
B(1,1) = beta0;
B(2,1) = beta2;
B(1,2) = beta4;
B(3,2) = beta7;

```

Generate the data.

Generate the data,  $X$ , then draw the error terms,  $U$ , and construct  $Y$ .

```

x1 = 2 + ((1:N)'/N).*normrnd(0,1,N,1);
x2 = 3 + 0.5*x1 + normrnd(0,1,N,1);
X = [ones(N,1) x1 x2]; % X.

u = mvnrnd(mu, smat, N); % U.

```

```

Y = X*B*G_inv + u;           % Y.
Y1 = Y(:,1);
Y2 = Y(:,2);

b1_ols = (X'*X)\X'*Y1;
Y1_hat = X*b1_ols;           % Fitted Y1.

b2_ols = (X'*X)\X'*Y2;
Y2_hat = X*b2_ols;           % Fitted Y2.

X1_2s1s = [ones(N,1) Y2_hat x1];

A:  WRITE THE 2SLS estimator for the $y_1$ equation here.

B:  WRITE THE 2SLS estimator for the $y_2$ equation here
    (you need to explain more for this one than for the previous one).

```

4. (20%) a) Explain what this code does. Write down the model being estimated.
- b) What happens where the code says A: what happens here?
- c) What happens where the code says B: what happens here?
- d) Fill in where it says C: what goes here.
- e) The question in the code, where it says D: What is calculated here?

## Set the parameters.

Set  $\beta = 0.5$  and  $\sigma = 1.5$ .

```
clc
clear

T = 35;                % Number of time periods.
N = 500;               % Number of individuals.

beta = 0.5;           % Beta.
sigma = 1.5;          % Standard deviation.
```

## Generate the data.

Draw the error terms,  $U$ , from the normal distribution and generate the data,  $X$  and  $Y$ .

```
u = normrnd(0,sigma,T,N);    % Error terms.

mu = repmat((0:T-1)'./T,1,N); % Time fixed effects.
alpha = repmat((0:N-1)./N,T,1); % Individual fixed effects.
```

```
X = normrnd(1,sigma,T,N);           % X.
```

```
Y = mu + alpha + beta*X + u;       % Y.
```

```
% Step 1: Regress Y and X on the dummies and get the residuals.
```

```
Y_ff = mean(Y,1);                  A: WHAT HAPPENS HERE? What are the new variables) %
```

```
X_ff = mean(X,1);                  % X is T by N.
```

```
Y_c = Y-Y_ff;
```

```
X_c= X-X_ff;                       B: WHAT HAPPENS HERE? What are the new variables?
```

```
% Step 2: Regress the residuals of Y from step 1 on those of X.
```

```
Xvec = xx                          C: WHAT GOES HERE (and also next line)
```

```
Yvec = xx
```

```
bols = inv(Xvec'*Xvec)*Xvec'*Yvec; D: WHAT IS CALCULATED HERE?
```

5. (20%) a) Explain what this program does. It is not enough to state the name of the procedure, you have to explain when one would use the model in economics (you can give

an example).

b) Fill in the missing code where it says SOMETHING IS MISSING.

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

global x z N

b0 = b(1);
b1 = b(2);
b2 = b(3);

XB = b0*x ;

L=0 ;

for i = 1:N

    if z(i) == 0

        L = L + log(normcdf(b1 - XB(i)));

    elseif z(i) == 1

        SOMETHING IS MISSING

    elseif z(i) == 2

        L = L + log(normcdf(XB(i) - b2));
```

```
end
```

```
end
```

```
L = -L;
```

```
end
```