ECONOMETRICS II, Fall 2020
Bent E. Sørensen

Midterm Exam II—October 26, 2019

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

1. (20%) Assume you have T observations ($t = 1, ..., T$) for 2 individuals ($i = 1, 2$) for two equations

$$y_{it} = \mu + e_{it} \, ,$$

and

$$w_{it} = \kappa + u_{it} \, ,$$

where y and w are some variables. μ and κ are constants, that we want to estimate by the best linear estimator and e and u are error terms. Assume the variance of e is σ_e^2, the variance of u is σ_u^2, and the covariance of e and u is σ_{eu}.

a) Explain how we estimate μ and κ.
b) Explain how we estimate σ_e^2, σ_u^2, and σ_{eu}.
c) How would you test $\hat{\mu} = \hat{\kappa}$?

2. (20%) Consider the model :

$$v_i = b\, w_i + c\, x_i + e_i,$$

$$w_i = d\, y_i + f\, q_i + g\, x_i + u_i \, ,$$

and

$$y_i = h\, w_i + k\, v_i + p_i,$$

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

a) Explain which (if any) of these equations one can estimate without bias.
b) Suggest a typical consistent estimator for the equation(s) that you can estimate. (You do not need to write down any formula, but you do need to explain which instruments, if any, a valid for each equation.)

3. (20%)

This code estimates the simultaneous equation model

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 y_2 + \beta_2 x_1 + u_1 \ , \\
y_2 &= \beta_4 + \beta_5 y_1 + \beta_7 x_2 + u_2 \ ,
\end{aligned}
$$

with $U \sim NID(0, \Sigma)$, using three stage least squares. Complete the code. (You will get most points by explaining in words, but even if you use words, you have to say what will do in terms of the output generated by the program. If you do not understand the code, you will not get points.)

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);
```

Three Stage Least Squares.

Estimate the model using three stage least squares.

```
% Stage 1: OLS.

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.

% Stage 2.
```

```
X1_2sls = [ones(N,1) Y2_hat x1];
X1X1 = X1_2sls'*X1_2sls;
X1Y1 = X1_2sls'*Y1;

b1_2sls = (X1X1)\X1Y1;                % 2SLS (eq. 1).

X2_2sls = [ones(N,1) Y1_hat x2];
X2X2 = X2_2sls'*X2_2sls;
X2Y2 = X2_2sls'*Y2;

b2_2sls = (X2X2)\X2Y2;                % 2SLS (eq. 2).

u1_2sls = Y1-X1_2sls*b1_2sls;         % Residuals, u1.
u2_2sls = Y2-X2_2sls*b2_2sls;         % Residuals, u2.


% Stage 3.

umat = [u1_2sls u2_2sls];             % Residuals, U.

%%% FILL IN HERE. %%%
```

4. (20%)

Midterm 2 Code: Panel Data.

This code uses OLS to estimate a fixed effects panel regression

$$y_{it} = \mu_t + \alpha_i + \beta x_{it} + u_{it},$$

where $u_{it} \sim N(0, \sigma^2 I)$. There are individual fixed effects in the current esitmation. Modify the code so that there are individual and time fixed effects.

Set the parameters.

There are 500 individuals observed over 35 years. 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. Each column is for an individual and each row is for a year.

```
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.
```

Estimate using OLS.

Esitmate the model using OLS with individual fixed effects. Each column is for an individual and each row is for a year.

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

Y_cfx = mean(Y,1);                  % Y is T by N.
```

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

Y_cfx = Y-Y_cfx;
X_cfx = X-X_cfx;

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

Xvec = reshape(X_cfx,[],1);
Yvec = reshape(Y_cfx,[],1);

bols = inv(Xvec'*Xvec)*Xvec'*Yvec;
```

5. (20%) Consider this expression:

$$\hat{\beta}_k = (X'(I - kM_Z)X)^{-1}X'(I - kM_Z)Y \, .$$

(5%) a) What is this estimator called and what does it do? (You should also explain what the variables are.)
(5%) b) Show that OLS is a special case. (Be explicit using formulas.)
(10%) c) Show that 2SLS is a special case. (Be explicit using formulas.)