Some notes to Seminar 1.

Question 1

Exercise 1 in Forty exercises

To *"specify a complete econometric model"* usually means "*to specify the properties of the disturbance term*". In the case of stochastic right hand side variables the specification is conditional on the explanatory variables. This means that the "classical assumptions" about zero expectation, homoscedasticity and no autocorrelation (as well as the normality assumption, if that is included) are *conditional on the explanatory variables*—as discussed in the seminar.

Because the explanatory variables are stochastic we also include an assumption about exogeneity: the explanatory variable x_{ki} is uncorrelated with all the disturbances:

$$Cov(u_i, x_{kj}) = 0, \text{ for } k = 1, 2 \text{ and for all } i \text{ and } j.$$

Next, in the question, we are asked to re-write:

$$m_{yk} = M\left[y, x_k\right] = \frac{1}{n}\sum\nolimits_i (y_i - \bar{y})(x_i - \bar{x}_k) \; k = 1, 2.$$

$$\begin{aligned} m_{y1} &= \frac{1}{n}\sum\nolimits_i y_i(x_i - \bar{x}_k) = \frac{1}{n}\sum\nolimits_i (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i)(x_i - \bar{x}_1) \\ &= \frac{1}{n}\sum\nolimits_i (\beta_1 x_{1i} + \beta_2 x_{2i} + u_i)(x_i - \bar{x}_1) = \beta_1 m_{11} + \beta_2 m_{21} + m_{u1} \end{aligned}$$

and, by the same procedure:

$$m_{y2} = \beta_1 m_{12} + \beta_2 m_{22} + m_{u2}.$$

Note that we can obtain the OLS estimators β_1 and β_2 from the two *normal equations*:

$$\begin{aligned} \hat{\beta}_1 m_{11} + \hat{\beta}_2 m_{21} &= m_{y1}, \\ \hat{\beta}_1 m_{12} + \hat{\beta}_2 m_{22} &= m_{y2}. \end{aligned}$$

Solving these equations gives:

$$\begin{aligned} \hat{\beta}_1 &= \frac{m_{22} m_{y1} - m_{y2} m_{12}}{m_{11} m_{22} - m_{12}^2} \\ \hat{\beta}_2 &= \frac{m_{11} m_{y2} - m_{y1} m_{12}}{m_{11} m_{22} - m_{12}^2} \end{aligned}$$

Define

$$A = m_{11} m_{22} - m_{12}^2 = m_{11} m_{22}(1 - r_{12}^2)$$

in order to save notation.

Next, subtract β_1 from both siden of the equation for $\hat{\beta}_1$, and substitute the expressions for m_{y1} and m_{y2} above:

$$\begin{aligned}
\hat{\beta}_1 - \beta_1 &= \frac{1}{A}(m_{22}\left[\beta_1 m_{11} + \beta_2 m_{21} + m_{u1}\right] - \left[\beta_1 m_{12} + \beta_2 m_{22} + m_{u2}\right] m_{12}) - \beta_1 \\
&= \frac{1}{A}\left\{\beta_1(m_{22}m_{11} - m_{12}^2) + \beta_2(m_{22}m_{21} - m_{22}m_{12}) + m_{22}m_{u1} + m_{u2}m_{12}\right\} - \beta_1 \\
&= \beta_1 + \frac{1}{A}\left\{m_{22}m_{u1} + m_{u2}m_{12}\right\} - \beta_1 = \frac{1}{A}\left\{m_{22}m_{u1} + m_{u2}m_{12}\right\}.
\end{aligned}$$

Since we condition on all x_{1i} and x_{2i}, the expectations operator "goes trough" $m_{22}m_{u1}$ (i.e. $E(m_{22}m_{u1}) = m_{22}E(m_{u1})$) and $m_{u2}m_{12}$ to give:

$$E(\hat{\beta}_1 - \beta_1 \mid x_{1i}, x_{2i}]) = 0$$

Same unbiasedness result for $\hat{\beta}_2$.

Unbiasedness also holds unconditionally. The proof is by iterated expectations.

$$E\left\{E(\hat{\beta}_1 - \beta_1 \mid x_{1i}, x_{2i}])\right\} = 0.$$

where the outher E "operates on" x_{1i} and x_{2i}.

It is also possible to show unconditional unbiasedness in one step by use of the exogeneity assumption. *Consistency* is most easily shown by probability limits:

$$\text{plim}(\hat{\beta}_1 - \beta_1) = \frac{\text{plim}\left\{m_{22}m_{u1} + m_{u2}m_{12}\right\}}{\text{plim } A} = 0$$

since all moments converge to their sample counterparts when n grows towards infinity. Exogeneity is sufficient for consistency if the disturbances have the classical properties, while predeterminedness of the regressors may be seen a necessary requirement. If the disturbances are autocorrelated, exogeneity is a necessary requirement. plim $A \neq 0$ is necessary always, But this is a weak requirement, as noted.

BLUE. Then need homoscedasticity and no autocorrelation as well.

Exercise 2 in Forty exercises

(a) Exogeneity.

The most common notion of exogeneity is that the explanatory variables in the model is uncorrelated with the disturbances. This definition is often written as

$$Cov(u_i, x_i) = 0$$

where u_i denotes the disturbance term in the regression model and x_i is an explanatory variable. However, as noted above, the more precise definition would be

$$Cov(u_i, x_j) = 0, \text{ for all } i \text{ and } j$$

so that it becomes clear that the explanatory variable is uncorrelated with *all* disturbances.

With time series and panels, the data have a natural ordering: Past, present and future!

A simple model which makes it relevant to introduce other concepts of exogeneity is:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + u_t,\ t = 1, 2, \ldots, T.$$

In this model, we say that x_t is *strictly exogenous* if

$$Cov(u_s, x_t) = 0, \text{ for all } s \text{ and } t.$$

We say that x_t is *predetermined* if

$$Cov(u_{t+s}, x_t) = 0, \text{for all } s = 0, 1, \ldots$$

We can use these concepts to show that the explanatory variable y_{t-1} is *predetermined but not strictly exogenous*, as discussed in the seminar.

There are other exogeneity concepts as well which are closely linked to the exact definition of a regression model, along the lines that we have seen above.

Weak exogeneity. We have WE if a parameter of interest can be efficiently estimated from the condition model, i.e. from a regression model. More precisely this means that if the parameter of interest is a coefficient in the conditional expectation derived for a the probability function (for y_i and $x_{1i}, x_{2i}, \ldots$) then the explanatory variables of the regression model are WE.

If we do not have WE, the implication is to consider another econometric model than the conditional expectation. Estimation methods associated with these econometric models are 2SLS and other instrumental variables based estimators.

Strong exogeneity. A variable x_t is SE if we have WE and x_t is not Granger caused by y_t.We can then forecast from the conditional expectation.

Super exogeneity. A variable x_t is SuE if we have WE and the coefficient of x_t is invariant to changes in the marginal model for x_t. Without SuE the Lucas critique applies. A change in x_t represents a structural change in the economy, and if the parameters of the regression model are not invariant to that change, the estimated effect of a change in x_i cannot be trusted. Lucas aimed his critique at OLS estimated models where expectations variables are wrongly replaced by observed ("actual") variables. If the structural break is in expectations, then these models will not be invariant. But the point is more general—and in some places it is popular to say that the Lucas-critique is a special case of the Haavelmo-critique, since Haavelmo was explicit about the danger of lack of invariance in econometric relationships in his 1944 "Probability approach".

Remark on testing for exogeneity: WE is in a way the most elusive property to test, but tests exists and will be discussed in the course. SE can be readily tested, since we can always formulate a model for x_t and test for Granger non-causality. SuE is also easy to test—there are usually no lack of regimes-shifts and structural changes in a sample of economic data!!!!!