London School of Economics
Department of Economics

Vassilis Hajivassiliou
Autumn Term 2024

M.Sc. Econometrics (Ec402)
2024–2025
Answers for Problem Set #9

Question 1

Consider a linear regression model with time-series data (y, X) of sample size T. The k regressors are grouped in two parts, X_A and x_B, of dimensions $T \times (k-1)$ and $T \times 1$ respectively. In other words, the second regressor group consists of a single regressor.

Suppose that the model satisfies the following assumptions:

$$
\begin{array}{cc}
A1: & rank(X) = k < T \\
A2: & y = X\beta + \epsilon = X_A\beta_A + x_B\beta_B + \epsilon \quad \text{with } E\epsilon = 0 \\
A3Rmi.X_A: & E(\epsilon|X_A) = E\epsilon \\
A3.x_B & \epsilon \text{ and } x_B \text{ correlated for all periods } t \\
A4\Omega: & E(\epsilon\epsilon'|X) = c^2\Omega \\
A5G: & \epsilon_t|X \sim N(0, \sigma^2)
\end{array}
$$

In other words, regressor x_B is an endogenous regressor. We are particularly interested in the true coefficients of the X_A variables, β_A.

1. (a) Suppose that the $T \times T$ matrix Ω is fully known. Define the Ordinary Least Squares (OLS) and Ideal Generalized Least Squares (IGLS) estimators of the whole β vector in this case. Explain carefully whether or not the OLS and/or IGLS can be unbiased and consistent for the true β.

ANSWER:

The sampling error vectors for the two estimators will be given by:

$$
\left(\begin{array}{c} \hat{\beta}_A^{ols} \\ \hat{\beta}_B^{ols} \end{array} \right) - \left(\begin{array}{c} \beta_A \\ \beta_B \end{array} \right) = \left(\begin{array}{cc} X_A'X_A & X_A'x_B \\ x_B'X_A & x_B'x_B \end{array} \right)^{-1} \left(\begin{array}{c} X_A'\epsilon \\ x_B'\epsilon \end{array} \right)
$$

$$
\left(\begin{array}{c} \hat{\beta}_A^{igls} \\ \hat{\beta}_B^{igls} \end{array} \right) - \left(\begin{array}{c} \beta_A \\ \beta_B \end{array} \right) = \left(\begin{array}{cc} X_A'\Omega^{-1}X_A & X_A'\Omega^{-1}x_B \\ x_B'\Omega^{-1}X_A & x_B'\Omega^{-1}x_B \end{array} \right)^{-1} \left(\begin{array}{c} X_A'\Omega^{-1}\epsilon \\ x_B'\Omega^{-1}\epsilon \end{array} \right)
$$

Given these formulae, OLS and IGLS will both be Biased and Inconsistent for *both* parts A and B: this is because

(i) the endogeneity of regressor x_B implies that the terms $x_B'\epsilon$ and $x_B'\Omega^{-1}\epsilon$ will not be zero in expectation, nor will $x_B'\epsilon/T$ and $x_B'\Omega^{-1}\epsilon/T$ converge to zero asymptotically. And

(ii) the bias/inconsistency carries over even to the part A estimators because in general neither $X_A'x_B$ nor $X_A'\Omega^{-1}x_B$ will vanish.

(b) Now suppose that the matrix Ω is known to equal the identity matrix I_S. In view of the endogeneity of x_B, a colleague proposes Instrumental Variables (IV) Estimation, defined by:

$$\hat{\beta}_{IV} = (W'X)^{-1}W'y$$

The colleague explains that the matrix W should be of the same dimension as X. She further explains that W should be constructed using only exogenous variables, implying that X_A can be used. Since this disallows the use of the endogenous regressor x_B, the colleague adds that we must find k_z additional "instrument" variables ($k_z \geq 1$) to construct W. Explain this method and describe the properties that all instrument variables used to construct W should possess. Your answer should include the terms "instrument validity" and "instrument relevance."

ANSWER:

Textbook answer: W should consist or solely of *valid* instruments or linear combinations of such valid instruments. An instrument variable z is termed "valid" if it is weakly exogenous w.r.t. to the error term, $E(z_s, error_s) = 0$. An instrument variable z is termed "relevant" if it has a high correlation with the endogenous variables of the model, in this case the endogenous regressor x_B.

Possible W matrices must satisfy: Req1=dimesions $T \times k$ with $rank(W'X) = k$. Req2: all columns of W must be weakly exogenous w.r.t. ϵ, implying that they must be selected from or be linear combinations of "valid" instrument variables. Call the matrix of such instruments Z, dimension $T \times k_z$.

Optimal W^* matrix: all the weakly exogenous variables in the model are (X_A, Z). The optimal combination that maximizes the correlations between W^* and the original regressors (X_A, x_B) is to regress by OLS X_A and x_B on (X_A, Z) and take the OLS fitted values, $\hat{X}_A$ and $\hat{x}_B$. The first $\hat{X}_A = X_A$, of course, since X_A was both on the LHS and the RHS of that regression.

(c) The colleague proposes the following instrument variables:

Variable z_1 is the sum of the first three regressor variables from the X_A group, i.e., $z_{t1} = x_{A,t1} + x_{A,t2} + x_{A,t3}$

ANSWER:

Since z_{t1} is a perfect linear combination of the first three regressor variables of the A group, the W matrix will not have a full rank of k. Hence $W'X$ will not be invertible and the IV method will break down. Note that z_{t1} will be weakly exogenous since each constituent variable is weakly exogenous, so it will be "valid" in the strict sense of the term. But it will be useless as an instrument since it cannot be used for the IVE construction. Note also that it would appear as highly "relevant" if we simply checked its correlation with the endogenous regressor x_B., since the X_A variables are typically going to be correlated with x_B.

Variable z_2 is the square of the fourth regressor variable from the X_A group, i.e., $z_{t2} = x_{A,t5}^2$

ANSWER:

This may be a decent instrument depending on the precise nature of the $x_{A,t5}$ regressor: it will be valid since it can be expected to be weakly exogenous because $x_{A,t5}$ is believed to be weakly exogenous and its square will be also. And it will be relevant to some extent because it can be expected to be correlated with x_B if $x_{A,t5}$ is correlated. And finally, it will not violate the rank condition since the correlation with the original $x_{A,t5}$ will not be perfect.

Variable z_3 is a measure of sunspot activity in period t.

ANSWER:

z_3 will clearly be a "valid" instrument in the sense of being weakly exogenous w.r.t. the error term, since sunspot activity can safely be assumed exogenous w.r.t. basically everything! But it will be completely "irrelevant" in that there is no reason to expect it to be correlated wth the endogenous regressor x_B.