ECONOMETRICS II PROJECT (ECO-305)

Project Motivation:

Understanding the determinants of healthcare expenditure is crucial for policymaking and for ensuring the efficient allocation of resources within the healthcare sector. However, estimating the impact of variables like health insurance on medical expenses is complicated due to the potential for endogeneity; for instance, those with higher health risks may be more likely to have insurance. Endogeneity can bias Ordinary Least Squares (OLS) estimates, leading to inaccurate conclusions.

This project aims to address the endogeneity issue by employing Two-Stage Least Squares (2SLS) estimation, an instrumental variable technique that provides a method for obtaining consistent estimates when dealing with endogenous regressors. By selecting appropriate instruments that are correlated with the endogenous variable (health insurance) but not with the error term, we can isolate the causal impact of health insurance on medical expenses.

The paper discusses the application of 2SLS with different instrument sets and compares the results with OLS estimates. This comparison provides insight into the biases that may arise from endogeneity and illustrates the importance of using instrumental variables in econometric analysis. The findings of this project will not only contribute to the academic understanding of healthcare economics but also have practical implications for policy interventions aimed at controlling healthcare costs and improving insurance coverage effectiveness.

Instrumental Variables Example:

- We want to study the factors influencing medical expenses (y1) given the endogenous regressor of having health insurance (y2) and exogenous regressors of illnesses, age, and income (x1). Instruments are the SS income ratio and firm multiple locations (x2).
- Data are from the Medical Expenditure Panel Survey (MEPS).

Single equation:

OLS regression: $y1 = \gamma 2'\beta 1 + x1'\beta 2 + u$
2SLS, first-stage equation: $y2 = x1'\gamma 1 + x2'\gamma 2 + e$
2SLS, second-stage equation: $y1 = \gamma 2'\beta 1 + x1'\beta 2 + u$

Systems of equations:

$y1 = \gamma 2'\beta 1 + z1'\gamma 1 + u1$
$y2 = \gamma 1'\beta 2 + z2'\gamma 2 + u2$

2SLS Estimation, Just-Identified Case (1 endogenous variable, 1 instrument)

	OLS regression for y1(log of med expenses)	2SLS: first stage for y2(health insurance)	2SLS:second stage for y1(log med expenses)
Have health insurance	0.075*	-	-0.852*
Illnesses (x1)	0.441*	0.011*	0.449*
Age (x1)	0.017*	0.054*	0
Log income			
SS income ratio	-	-0.200*	
constant	5.780*	0.959	6.590*

- Interpretation of coefficient on the endogenous variable in OLS model: For individuals with health insurance, the medical expenses are 7.5% higher than those for individuals without health insurance.
- Interpretation of the coefficient of the endogenous variable in 2SLS: After instrumentation, for individuals with health insurance, their medical expenses are 85.2% lower than those for individuals without health insurance.
- 2SLS coefficient turned out quite different from the OLS coefficient.

2SLS Estimation, Just-Identified Case (1 endogenous variable, 2 instrument)

	OLS regression for y1	2SLS: first stage for y2	2SLS: second stage for y1
Have health insurance	0.075*	-	-0.97
Illnesses	0.441*	0.012*	0.450*
Age	-0.003	-0.008	-0.013
Log income	0.017*	0.051	0.108*
SS income ratio	-	-0.191	-
Firms location	-	0.116*	-
Constant	5.780*	0.912*	6.692*

- With two instruments instead of one, the estimates changed only slightly from -0.852 to -0.970 for the coefficient on have health insurance.

Tests

- The Durbin-Wu-Hausman test compares OLS 2SLS model coefficients. The null hypothesis is that the regressors are exogenous is rejected. Therefore, the health insurance is an endogenous regressor and we need to use instrumental variable approach.
- The test for overidentifying restriction shows all instruments are valid.
- There is low correlation among instruments and endogenous variable, of about 0.1-0.25 in absolute value. The correlation is low but not an indication weak instruments.
- The test for weak instruments looks at the F statistic for joint significance of instruments. The number is 69 from the model with 1 instrument and 59 from the model with 2 instruments, which is larger than the rule of thumb of 10. Therefore, the instruments are not weak.

System of equations, 2SLS and 3SLS

	2SLS estimation for y1	2SLS: first stage for y2	3SLS: second stage for y1	3SLS estimation for y2
Have health insurance	-1.673	-	-1.599	-
Illnesses	0.458*	-0.1	0.456*	-0.1
Age	-0.019	-	-0.018	-
Log income	0.142*	-	0.136*	-
SS income ratio	-0.164	_	-0.134	-
Firms location	-	-0.283	-	-0.283
Constant	7.377*	-283	7.237*	-0.972
Log of med expenses	_	-0.972	_	0.235

- Interpretation of coefficients: for individuals with health insurance, medical expenses are 167.3% or 159.9% lower from the 2SLS or 3SLS models.
- The 3SLS results are different (even though only slightly) from the 2SLS because of different regressors and instruments.
- We will get identical results if the system is just identified, 1 instrumental variable for each endogenous variable, and the same exogenous variables in both equations.