Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Econometria II

Professor: Daniel Domingues dos Santos

Monitor: Diego Menezes

Aula 02

1. Variáveis Instrumentais

O objetivo deste exercício é estimar o efeito de possuir ou não seguro de saúde no log dos gastos com remédios. Como base de dados, será usado o arquivo "iv_health.dta", cedido pelo site Econometrics Academy. Essa base de dados procura retratar o sistema de saúde americano e é composta por 10089 observações de indivíduos com suas diversas características individuais. Algumas dessas relacionadas à interação do indivíduo com o sistema de saúde, como "illness", i.e., quantas doenças ele teve no ano; outras relacionadas apenas às características descritivas dele, como "female", i.e., se é uma mulher ou não. Ao todo constam 29 características diferentes.

Vocês irão usar estimadores MQO e variáveis instrumentais nesse exercício para conseguir capturar os efeitos desejados do seguro de saúde nos gastos com remédios.

1.1

Estamos interessados no efeito da dummy de seguro de saúde ("healthinsu") sobre o log dos gastos com remédios ("logmedexpense"). Realize uma regressão naïve que capte esse efeito e analise o resultado. Por que esse resultado potencialmente não é o efeito causal que você está procurando?

1.2

Explique o método de Variáveis Instrumentais e por que deve ser realizada uma regressão em dois estágios. Em seguida, explique as restrições necessárias para uma variável poder ser considerada uma variável instrumental de outra variável endógena.

1.3

Tendo a variável "healthinsu" como variável endógena, realize uma regressão MQ2E utlizando a variável "SSI/income", i.e., qual proporção do salário mensal do indivíduo depende do seguro americano para inválidos, cegos e idosos de baixa renda (SSI), como instrumento. Teste se a variável instrumental escolhida satisfaz a restrição de inclusão.

Analise os resultados achados, comparando-os com aqueles encontrados na regressão naïve.

1.4

Realize a mesma regressão MQ2E utilizada no exercício anterior, porém agora controlando para as variáveis exógenas “illness”; i.e., quantas doenças o indivíduo teve no ano; “age”; i.e., a idade do indivíduo; e “logincome”; i.e., o log da renda mensal dele. Os resultados variam? Por quê?