

Econometrics Assignment 4
Eric Tu

(a) $y_{i1} - y_{i0} = \alpha + \beta d_i + \gamma y_{i0} + x_i' \delta + \epsilon_i$

d_i represents the answer to the survey question, “Did you follow this diet in the past year.” This variable may be endogenous as previously healthy individuals may find it easier to follow the diet but experience less weight loss due to their already healthy lifestyle. This would lead to OLS underestimating the true effect of the diet. There are probably additional factors that can lead to overestimation or underestimation of the OLS estimator making it difficult to state with conviction that the OLS will overestimate or underestimate the true effect of the diet.

- (b) The first condition, $\frac{1}{n} \mathbf{Z}' \epsilon \rightarrow 0$ as n grows large can be interpreted as the average effect of living in a region with or without door-to-door advertisement of the diet having no effect on the error in measuring the effect of the diet (no direct effect on effectiveness of the diet).
The next condition, $\frac{1}{n} \mathbf{Z}' \mathbf{X} \rightarrow \mathbf{Q} \neq 0$ as n grows large can be interpreted as the average effect of living in a region with or without door-to-door advertisement of the diet does effect whether an individual follows the diet or not.

- (c) The first condition, $\frac{1}{n} \mathbf{Z}' \epsilon \rightarrow 0$ as n grows large is tested using the Sargan test. The test is performed by comparing the nR^2 value against the significant value of the $\chi^2(m - k)$ distribution, where R^2 is the r-squared measurement of the e_{2SLS} regression on Z , n is the number of observaions, m is the number of instruments in Z , and k is the number of explanatory variables in X . Since $m = 1$ (One z instrument: door-to-door advertising) and $k = 1$ (β transforms d_i : whether the diet was followed), the chi square distribution cannot be evaluated with 0 degrees of freedom, and the first condition cannot be evaluated.

The second condiion, $\frac{1}{n} \mathbf{Z}' \mathbf{X} \rightarrow \mathbf{Q} \neq 0$ as n grows large can be tested simply by regressing the potentially endogenous variables against the instruments and exogenous variables in the regression $X_1 = Z^* \gamma_1 + X_2 \gamma_2 + \eta$ and testing the null hypothesis $H_0 : \gamma_1 = 0$.

(d) $\beta = (Z'X)^{-1} Z'y$

$$\begin{aligned}
 Z &= \begin{bmatrix} 1 & z_1 \\ 1 & z_2 \\ \dots & \dots \\ 1 & z_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & d_1 \\ 1 & d_2 \\ \dots & \dots \\ 1 & d_n \end{bmatrix} \\
 Z'X &= \begin{bmatrix} n & \sum_{i=1}^n d_i \\ \sum_{i=1}^n z_i & \sum_{i=1}^n d_i z_i \end{bmatrix} \quad Z'y = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i z_i \end{bmatrix} \\
 \beta &= \frac{1}{n \sum_{i=1}^n d_i z_i - \sum_{i=1}^n d_i \sum_{i=1}^n z_i} \begin{bmatrix} \sum_{i=1}^n d_i z_i & - \sum_{i=1}^n d_i \\ - \sum_{i=1}^n z_i & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i z_i \end{bmatrix} \\
 \beta &= \frac{1}{n \sum_{i=1}^n d_i z_i - \sum_{i=1}^n d_i \sum_{i=1}^n z_i} \begin{bmatrix} \sum_{i=1}^n d_i z_i \sum_{i=1}^n y_i - \sum_{i=1}^n d_i \sum_{i=1}^n y_i z_i \\ - \sum_{i=1}^n z_i \sum_{i=1}^n y_i + n \sum_{i=1}^n y_i z_i \end{bmatrix} \\
 &= \frac{1}{n \sum_{i=1}^n d_i z_i - \sum_{i=1}^n d_i \sum_{i=1}^n z_i} \begin{bmatrix} 0 \\ n \sum_{i=1}^n y_i z_i - \sum_{i=1}^n z_i \sum_{i=1}^n y_i \end{bmatrix} \\
 &= \frac{n \sum_{i=1}^n y_i z_i - \sum_{i=1}^n z_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n d_i z_i - \sum_{i=1}^n d_i \sum_{i=1}^n z_i} \\
 &= \frac{\sum_{i=1}^n y_i z_i - \frac{1}{n} \sum_{i=1}^n z_i \sum_{i=1}^n y_i}{\sum_{i=1}^n d_i z_i - \frac{1}{n} \sum_{i=1}^n d_i \sum_{i=1}^n z_i} \\
 &= \frac{\frac{1}{\sum_{i=1}^n z_i} \sum_{i=1}^n y_i z_i - \frac{1}{n} \sum_{i=1}^n y_i}{\frac{1}{\sum_{i=1}^n z_i} \sum_{i=1}^n d_i z_i - \frac{1}{n} \sum_{i=1}^n d_i} \\
 &= \frac{\Delta - \frac{1}{d}}{\frac{1}{d} - \frac{1}{d}}
 \end{aligned}$$