

a) people self-select for the diet.

life style decision \rightarrow following the diet or not
 \rightarrow weight change

b) 1. The advertising should not be dependent on health/weight characteristics.
regions with advertising should not be selected based on the average weight of the region's inhabitants.

2. The advertising should be effective: the prob. of following the diet should be higher in regions in which the diet was advertised.

c) 1. No, one would need to have more instruments (Sargan test)

2. Yes, create a model for d_i with z_i , y_i , and x_i as expl. variables and test whether z_i is significant.

$$d) \quad Z = \begin{pmatrix} 1 & z_1 \\ 1 & z_2 \\ \vdots & \vdots \\ 1 & z_n \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} 1 & d_1 \\ 1 & d_2 \\ \vdots & \vdots \\ 1 & d_n \end{pmatrix}$$

$$(Z'X)^{-1} Z'y = \begin{pmatrix} n & \sum d_i \\ \sum z_i & \sum d_i z_i \end{pmatrix}^{-1} \begin{pmatrix} \sum y_i \\ \sum z_i y_i \end{pmatrix}$$

$$= \frac{1}{n \sum d_i z_i - \sum z_i \sum d_i} \begin{pmatrix} \sum d_i z_i - \sum d_i \\ -\sum z_i & n \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum z_i y_i \end{pmatrix}$$

$$\begin{aligned} \hat{\beta} &= \frac{n \sum z_i y_i - \sum y_i \sum z_i}{n \sum d_i z_i - \sum z_i \sum d_i} = \frac{\sum z_i y_i - \frac{1}{n} \sum y_i \sum z_i}{\sum d_i z_i - \frac{1}{n} \sum z_i \sum d_i} \\ &= \frac{\frac{1}{\sum z_i} \sum z_i y_i - \frac{1}{n} \sum y_i}{\frac{1}{\sum z_i} \sum d_i z_i - \frac{1}{n} \sum d_i} = (\Delta' - \Delta) / (\bar{d}' - \bar{d}) \end{aligned}$$