Motivating Demeaning Approach

Ragip Gurlek

11/2/2020

Contents

This document analytically motivates demeaning of variables in the Yelp dataset that I use for my *Price Placebo Effect* project. The project investigates if consumers' experience with a more expensive restaurant is better even if the restaurant is similar to its competitors in other dimensions. To address the inherent endogeneity in the problem, I use Tax as an instrumental variable (IV). I also demean the variables by their *local means*[1] to eliminate the bias introduces by location. I assume the Directed Acyclic Graph (DAG) in Figure 1 captures the causal relationship.

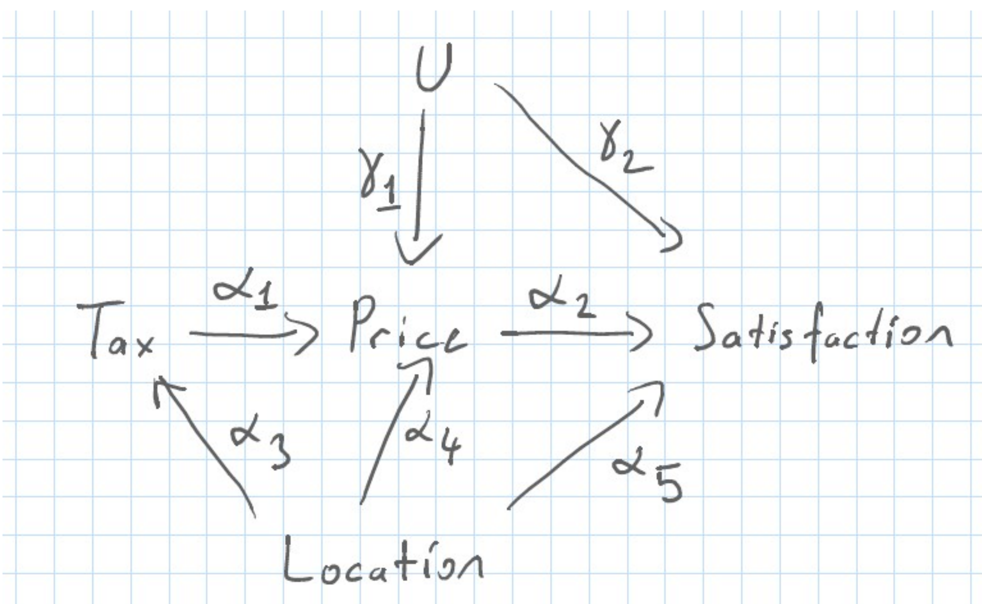

Figure 1: Causal Relationship of Variables

This graph implies the following structural equation model

[1]Local in geographical proximity.

$$
\begin{aligned}
\text{Tax } &= f(\text{ Location }) \\
U &= m(\text{ Location }) + \varepsilon_1 \\
\text{Price } &= g(\text{ Location }) + \alpha_1 \text{ Tax } + \gamma_1 U + \varepsilon_2 \\
&= g(\text{ Location }) + \alpha_1 f(\text{ Location }) + \gamma_1 m(\text{ Location }) + \gamma_1 \varepsilon_1 + \varepsilon_2 \\
\text{Satistaction } &= \alpha_2 \text{ price } + h(\text{ Location }) + \gamma_2 U + \varepsilon_3 \\
&= \alpha_2 g(\text{ Location }) + \alpha_2 \alpha_1 f(\text{ Location }) + \alpha_2 \gamma_1 m(\text{ Location }) + \alpha_2 \gamma_1 \varepsilon_1 + \alpha_2 \varepsilon_2 \\
&+ h(\text{ Location }) + \gamma_2 m(\text{ Location }) + \gamma_2 \varepsilon_1 + \varepsilon_3
\end{aligned}
$$

Let

$$
\begin{aligned}
\overline{\text{Satistaction}}_i &= E[\text{Satistaction}|\text{Location} \in \ell_i(\epsilon)] \\
&= (\alpha_2 \alpha_1 \alpha_3 + \alpha_2 \alpha_4 + \alpha_5)\, E[\text{Location}|\text{Location} \in \ell_i(\epsilon)] + (\alpha_2 \gamma_1 + \gamma_2)\, E[U|\text{Location} \in \ell_i(\epsilon)] \\
\widetilde{\text{Satistaction}}_i &= \text{Satistaction}_i - \overline{\text{Satistaction}_i} \\
&\approx \alpha_2 \alpha_1 \varepsilon_1 + (\alpha_2 \gamma_1 + \gamma_2)\, \tilde{U} + \alpha_2 \varepsilon_2 + \varepsilon_3
\end{aligned}
$$

where ℓ_i is the ϵ-ball around the actual location of the observation. Therefore, the approximation gets better for smaller ϵ.

Similarly,

$$
\begin{aligned}
\widetilde{\text{Tax}}_i &= \text{Tax}_i - \overline{\text{Tax}_i} \\
&\approx \varepsilon_1 \\
\widetilde{\text{Price}}_i &= \text{Price}_i - \overline{\text{Price}_i} \\
&\approx \alpha_1 \varepsilon_1 + \gamma_1 \tilde{U} + \varepsilon_2
\end{aligned}
$$

Therefore, we can approximate the DAG in Figure 1 with the one in Figure 2.

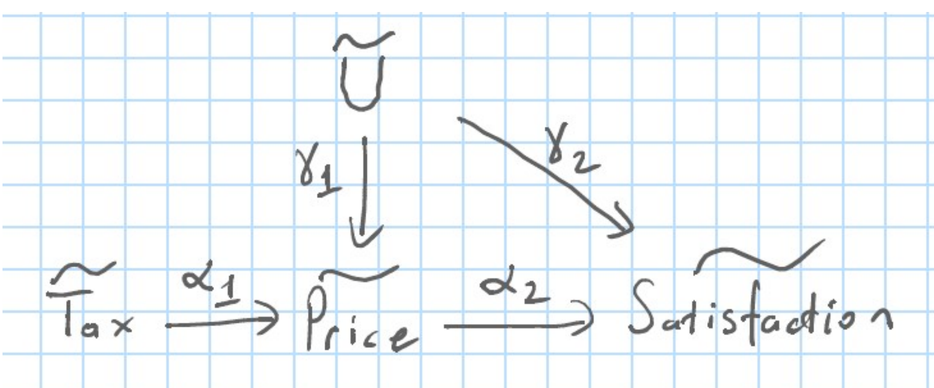

Figure 2: Alternative Causal Relationship

This analysis implies we can employ two approaches. First, we can use the DAG in Figure 2 and estimate a weighted local average for each variable to demean it. Then, we can estimate the effect with 2SLS.

Alternatively, we can directly estimate a 2SLS where we use cluster fixed-effects as an approximation to Location to control it. The DAG in Figure 1 shows that Location is the sufficient set to identify the effects of Tax on both Price and Satisfaction.