

Appendix D Control Approach and Fixed Effects

1. Controlling for Confounding Factors

The selection bias arises because individuals in treated and control groups are different from each other. But what if these differences are due to factors that we can observe and measure? For example, piece rate workers and salary workers may differ in their ability, but what if we could compare the piece rate workers and salary workers of *same* ability?

This idea of controlling for observable differences is the basis of strategy known as controlling for confounding factors, or simply the control approach. To illustrate how this approach works, suppose that the potential outcomes are given by:

$$y_i^0 = \alpha + \gamma z_i + \varepsilon_i$$

$$y_i^1 = \alpha + \beta + \gamma z_i + \varepsilon_i$$

where z represents the set of observable and measurable characteristics of individuals.

Given these potential outcomes, the observed outcome can be expressed as:

$$y_i = y_i^0 + (y_i^1 - y_i^0)d_i$$

$$= \alpha + \beta d_i + \gamma z_i + \varepsilon_i$$

The comparison of outcomes for treated and untreated individuals, conditional on observable characteristics z , is then:

$$E[y_i | z_i; d_i = 1] - E[y_i | z_i; d_i = 0]$$

$$= \beta + E[\varepsilon_i | z_i; d_i = 1] - E[\varepsilon_i | z_i; d_i = 0]$$

Therefore, this approach can identify the causal effect of treatment if the following is true:

$$E[\varepsilon_i | z_i; d_i = 1] = E[\varepsilon_i | z_i; d_i = 0]$$

Intuitively, the assumption states that the only reason why the treatment and control groups are different is because of differences in their observable characteristics. However, once we control for these confounding factors, the treatment and control groups are sufficiently similar so that we can infer causal effect by simply comparing average outcomes for the two groups.

If this assumption holds, then the model

$$y_i = \alpha + \beta d_i + \gamma z_i + \varepsilon_i$$

can be estimated by ordinary least squares method and the estimate of β can be interpreted as the causal effect of treatment. (See Appendix E)

The credibility of the control approach relies on the assumption that all relevant differences across individuals are captured by factors included in the model. The difficult question is how do we know what relevant differences are; i.e. what factors should one include in the model?

As a general rule, one should include any factor that satisfies *both* of these criteria:

- (1) It has an impact of outcome, and

- (2) It is correlated with the treatment.

The first criterion is intuitive: even if individuals are different with respect to a factor that has no impact of outcome, this difference is irrelevant. The second criterion simply states that even if a factor has an impact of outcome but its average value is similar for both treatment and control groups, the factor need not be included in the model.

What would happen if we mistakenly included a factor that is neither correlated with the treatment nor has an impact on the outcome? The estimated causal effect would be the same as if we did not include this factor. However, the estimated variance of the causal effect would be larger. In this case, we may conclude that the causal effect is statistically insignificant when in fact this is not the case.

There are also factors that should never be included in the model. Specifically, one should not include factors that are either determined by the outcome or factors that are jointly determined with the outcome.

In conclusion, the control approach relies on the assumption that all relevant differences between the treatment and control groups are differences that could be observed, measured, and accounted for. This is clearly a restrictive assumption, and for this reason the control approach is one of the least credible empirical strategies. However, the control approach may be used in conjunction

with other identification strategies to increase its credibility. For example, one may wish to control for observed differences between individuals in a randomized experiment, even though randomization by itself should make the treatment and control groups similar.

2. Fixed Effects

The control approach is based on the idea that observable differences between treatment and control groups can be accounted for, but it suffers from the criticism that it can't control for the unobserved differences. What cannot be observed cannot be measured.

The fixed-effects approach attempts to improve the control approach by assuming that a part of unobservable differences across individuals is fixed and does not vary over time.

Specifically, suppose that the potential outcomes are given by:

$$y_{it}^0 = u_{it}$$

$$y_{it}^1 = \beta + u_{it}$$

where t indexes time and i indexes individuals. Then, the observed outcome can be expressed as:

$$y_{it} = y_{it}^0 + (y_{it}^1 - y_{it}^0)d_{it}$$

$$= \beta d_{it} + u_{it}$$

As usual, the problem with comparing treatment and control individuals is that they may be different in factors related

to both the outcome and their probability of being treated:

$$\begin{aligned} & E[y_{it} | d_{it} = 1] - E[y_{it} | d_{it} = 0] \\ &= \beta + E[u_{it} | d_{it} = 1] - E[u_{it} | d_{it} = 0] \end{aligned}$$

The fixed effect approach assumes that unobservable differences in u can be decomposed into those that vary over time (ε_{it}) and those that do not vary over time (α_i):

$$u_{it} = \alpha_i + \varepsilon_{it}$$

Further, suppose that factors that vary over time do so in a similar way for both the treatment and control individuals:

$$E[\varepsilon_{it} | d_{it} = 1] = E[\varepsilon_{it} | d_{it} = 0].$$

If we now compare average changes over time for treated and control individuals, we get:

$$\begin{aligned} & \Delta E[y_{it} | d_{it} = 1] - \Delta E[y_{it} | d_{it} = 0] \\ &= \beta + \Delta E[u_{it} | d_{it} = 1] - \Delta E[u_{it} | d_{it} = 0] \\ &= \beta + \Delta E[\varepsilon_{it} | d_{it} = 1] - \Delta E[\varepsilon_{it} | d_{it} = 0] \\ &= \beta \end{aligned}$$

Therefore, by differencing outcome over time for the treated and control individuals, we can obtain an estimate of the causal effect of the treatment.

Notice what the fixed-effects strategy is based on. First, there are some fixed effects specific to individuals, and these effects will difference out if we compare changes rather than levels of outcome. Second, changes in other unobservable factors do not affect the treatment and

control groups differently. Clearly, all relevant factors that we can observe and that change over time can be controlled for, but there still will be factors that change over time that we cannot observe. In the fixed-effects approach, these unobserved factors are irrelevant because they change in the same way for both treatment and control groups.

Notice also the advantages of using the fixed effects approach. If we were just comparing treated and control individuals at a point in time, we could not possibly control for unobservable differences that don't vary over time. Therefore, there is always a risk that such a difference explains the outcome and not the treatment we are interested in. In addition, if we were just comparing treated individuals over time, at times when they received the treatment and at times when they didn't, we can never be sure that some other factors did not change over time that could explain the difference in outcome. The fixed effects approach explores variation in outcome over time to difference out fixed but unobserved factors, and uses the control group to approximate changes in the outcome for the treatment group in the absence of any treatment.