1 Empirical Methodology

To estimate the effect of P2E on labor market outcomes and recidivism, we begin by presenting the following model

$$Y_i = \beta_0 + \beta_1 P2E_i + X_i\beta_2 + \varepsilon_i \tag{1}$$

where Y_i is the outcome of interest, i.e., an indicator variable that takes on the value of one if individual i is employed. The variable of interest, $P2E_i$, is another indicator variable that takes on the value of one if the offender i receives $P2E$ treatment services, X_i is a vector of observed characteristics (e.g., individual's gender, race, and age) and ε_i is the error term. The coefficient β_1 represents the effect of $P2E$ treatment services on the likelihood of employment (or reoffending). For the sake of brevity, we use employment as our outcome of interest below, but we extend our analysis to reoffending outcomes as well.

Straightforward estimation of equation (1) via Ordinary Least Squares (OLS) provides an unbiased effect of treatment if participation in $P2E$ services is exogenously determined. However, there are many potential unobserved factors that affect employment, and that are also correlated with participation in treatment services (e.g., individual's ability and motivation). Ignoring these factors in the estimation of equation (1) is likely to yield a biased coefficient estimate of the impact of $P2E$ treatment services. It is also important to note that potential contamination is not limited to OLS. Other estimation techniques such as propensity score and semi-(or non-) parametric estimation techniques, which hinge upon selection on observables assumption, suffer from similar biases.

To address these potentially confounding effects, in the absence of gold-standard randomized control trials or availability of an instrument variable, we rely on estimation strategy proposed by Altonji et al. (2005) (AET method) and extended by Oster (2019). This technique would allow us to investigate and draw conclusions about the sensitivity of $P2E$ estimates to potential omitted variable bias.

1.1 AET Method

The proposed estimation method is based on the notion that the careful selection on the observables provides information about the amount of selection on the unobserved explanatory variables. This strategy is beneficial especially when prior information is unavailable about the exogeneity of the variable of interest. The method allows us to quantitatively assess the degree of omitted variables bias. Under the equality of selection on observables and unobservables, it is possible to estimate the size of the asymptotic bias.

We intend to determine the selection bias based on computing the ratio of selection on unobserved explanatory variables to selection on observables. This would be required if we have to characterize the entire effect of $P2E$ programs to selection bias. The ratio of the coefficient of interest estimate and the implied bias would provide a measure for how strong the selection on unobservables would have to be to explain away the entire treatment effect, relative to selection on observables.

The assumptions required here are that in order to infer about selection on the unobservables from selection on the observables, we need to have sizable number of observed explanatory variables with sufficient explanatory power. Put differently, the observables are likely to be representing the maximum range of factors determining the outcome variable.

1.2 Oster Method

Oster (2019) takes a step further in extending the AET methodology to evaluate the robustness of estimates to omitted variable bias but with less restrictive assumptions of the AET method. As per the test proposed by AET method, it is valid only when the null hypothesis is a zero-treatment effect. It implicitly assumes that inclusion of unobservables would lead to fully explaining the outcome (i.e., R^2 value equal to 1). Note that R^2 of 1 may underestimate the robustness of results if there exists measurement error in the outcome variable.

The extension proposed by Oster allows to calculate a consistent estimate of the bias-adjusted treatment effect. This is done by assuming a value for the relative degree of selection on observed and unobserved variables (δ) and a maximum value of R^2 ($R^2_{\max}$) which could reasonably justify if we could include all unobservables in the estimation process.

The bias adjusted coefficient is defined as

$$\beta = \beta_{long} - (\beta_{short} - \beta_{long}) \frac{R^2_{\max} - R^2_{long}}{R^2_{long} - R^2_{short}} \tag{2}$$

where β is the bias-adjusted coefficient, β_{long} and R^2_{long} are the coefficient estimate and R^2 from the regression including controls (observables). β_{short} and R^2_{short} are the coefficient estimate and R^2 from the regression without controls. $R^2_{\max}$ is the maximum value assumed for R^2. The baseline specification would be the one without controls. The equation considers the movement in the coefficient on $P2E$ participation which is further rescaled by the movement in R^2 values.

The fear of omitted variable bias can be tackled by considering the coefficient stability and considering the importance of the controls in explaining the variance of the outcome. We will explore the sensitivity of the treatment effects to the sequential inclusion of observed controls. The key point here is that the quality of a control variable is detected by how much of the variance in the outcome is explained by its inclusion. Omitted variable bias is proportional to changes in coefficient estimates, if these changes are scaled by the change in R^2 when controls are included. In other word, how much the R^2 moves when the controls are introduced. By employing a tractable strategy, our estimator will be consistent for the bias-adjusted treatment effect.

References

[1] Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113(1), 151-184.

[2] Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business and Economic Statistics*, 37(2), 187-204.