Problem set 3 Generalized Roy Model Control Functions

Take the same model (the two choice model) and dataset as in the previous problem set (dataset _2b). We are now going to forget about factor models and MCMC or MLE. We are now going to estimate nonparametric control functions. Remember that you know the right answer for ATE, TT from the last problem set. For this purpose we are going to focus again in the present value of $Y$ as before. The model is the same as in the previous problem set.

Now however we are going to stay non parametric so we are going to make no assumption as to what the distributions of $\varepsilon_1$, $\varepsilon_0$ and $V$ (remember however that even before we were being semiparametric (flexible) by assuming mixtures of normals). We just assume they are not independent and we will forget about the factor structure.

1. How would you define the average treatment effect in this general case?
2. How about the average effect of treatment on the treated?
3. Use Kernel Regression and Local Linear Regression techniques to estimate both these parameters in terms of the propensity score (use a probit model to estimate the propensity score, if you want to, you can try different methods to estimate the propensity score. Your results should not change much). How good are your results? How sensitive are the results to choice of Kernel? To choice of bandwidth?

Now suppose $f_2$ becomes known to you so that $f_2$ is now like another $X$ and a $Z$.

4. Redo question 3. Did your results change much (they should not)?
5. Compare what we were doing in PS2 and PS3. Would you say that what we did in problem set 2 is a control function approach?