PP346 Program Evaluation Final Exam

UCID 12174174

Dec 8, 2018

1. In Harding et al: there are three conceptual and methodological problems in estimate the effect of incarceration. The first is specifying a proper comparison group of individuals who could have been sentenced to prison but were not. The second is unobserved confounding. Individuals who are sentenced to prison tend to have very poor labor market prospects even before imprisonment, so it is difficult to convincingly separate individual characteristics from the effect of imprisonment itself. The third is the difficulty of specifying for whom imprisonment affects labor market experiences and how such effects arise.

2. (a)

   2.1. **Instrument relevance**: $Cov(D, Z^I) \neq 0$
   The instrument must be correlated with the endogenous explanatory variable, *independent of the other explanatory variables*, where Z^I denotes the independent part of the instrument and, naturally, can be thought of as the residuals from an auxiliary regression of the instrument on the other exogenous explanatory variables, $Z_i = \delta_1 + \delta_3 X_{3i} + ... + \delta_K X_{Ki} + v_i$.
   The stronger the correlation, the better.

   2.2. **Instrument validity:** $Cov(Z, u) = 0$ and $Z \notin X_k$
   If Z is a valid instrument for the effect of treatment D on outcome Y, then Z has no causal connection with Y except through D. The instrument Z should be randomly assigned.

   2.3. **Monotonicity**: $D(z = 1) - D(z = 0) \geq 0 \ \forall \ i$
   It requires that all people who react to the instrument change their treatment status the same way. The monotonicity assumption therefore rules out defiers.

   (b) Haring et al want to estimate the causal effect of being sentenced to prison on labor market outcomes. The assumptions require are: the causal effect of interest D (sentence to prison) should be correlated with the IV they are going to use, namely judge identifiers. Additionally, in order for judge identifiers to be valid IV, it can only affect employment through the sentence the judge imposes and there are no other unobserved variables that create an association between the instrument and the outcome. Last but not least, it requires that the instrument only affects the treatment in one direction -- a judge who imposes more punitive sentences than her colleagues to some individuals does not also impose more lenient sentences than her colleagues to others.

3. The relevance condition is based on the idea that judges have considerable discretion in sentencing and that different judges systematically sentence more or less harshly than other others. Haring et al examine the relevance condition by examining how the probability of sentencing to prison varies by judge within county and also calculated Shea's partial R-squared, which shows the proportion of the variation in each treatment variable independently explained by the instruments. Those values show that the first stage explains a substantial portion of the variation in treatment across all treatments and all subgroups, meaning that the IVs are sufficiently strong. They also report the relevant F-statistics, which were large for the overall sample for the prison, linear prison length, and linear probation length treatments when they used only the judge identifiers as IV. Thus, I am convinced the assumption is satisfied.

4. All felony cases in Michigan are handled by circuit courts, and all circuit courts have computerized case management systems that assign cases at filing to judges using a random number generator. Thus, I am convinced judges are randomly assigned in Michigan.
   As for the assumption the judge to whom one is assigned only affects employment through the sentence the judge imposes, I am not completely convinced that the assumption is satisfied. Haring et al discussed several scenarios that would violated the assumption one by one, instead of considering them simultaneously. Haring et al should have done the multiple hypothesis test on those scenarios, because those scenarios are correlated and may affect the instrument validity together. In that case, it will raise multiple testing issue.

5. The authors relaxed the "monotonicity" assumption by interacting judge dummies with presentencing individual characteristics and also treating those interactions as instruments. I would like to see whether the impact of judge assignment on the probability of punitive sentences/lenient sentences is monotonic across defendants of the same race.

6. The one in terms of time since sentencing is more likely to be valid. Starting the risk period at sentencing will capture and quantify the direct impact of incarceration on employment via incapacitation. This approach also provides the cleanest counterfactual comparison between those who receive different types of sentences.

   Starting the risk period for prisoners at release have several problems; Individuals in the prison and probation groups who were sentenced in the same year will start their risk periods in different years; treatment effects could be biased because people in the prison group are likely to be older, on average, at the start of the risk period; Release dates are endogenous because they are determined in part by post sentencing behavior in prison, which introduces bias; Finally, starting the risk period for the treated at their release date will shrink their follow-up period.

7. Before-after estimation require zero trends assumption, since the Before-After estimator is just the difference between the outcome variable at two time periods: one period from before the intervention and one period from after. Zero trends assumption requires the outcome remained constant in the absence of the intervention. It is a strong assumption.

   It does not appear to be satisfied in the authors' data, because once an individual got sentenced, lots of things will change: his emotional and physical health, his family relationship and so on. Using a B-A estimation will get the counterfactual wrong.

8. The authors could have done the difference-in-difference analysis by estimate the effect of imprisonment on employment on the prison group and use the probation group as the control.

   The DD estimator requires the counterfactual trend in the prison group equals the observed trend in the probation group. This is referred to as parallel trends. It appears to be satisfied in the authors' data(Fig .1), the employment trends for *Prison, Black* and *Probation, Black before* seem pretty parallel over the window of time beginning 36 quarters before the sentence to 3 quarters before the sentence. Parallel trends also hold for *Prison, White* and *Probation, White* over the window of time beginning 36 quarters before the sentence to 3 quarters before the sentence.
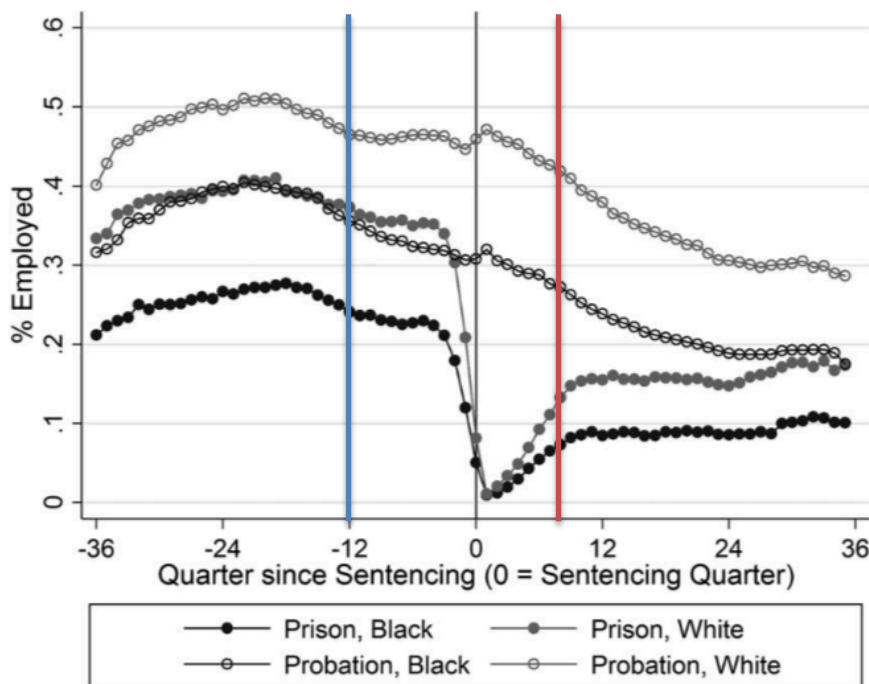

FIG. 1.—Employment relative to sentence date, by sentence type and race.

24 months = 2 years = 8 quarters.
Let t= -12 as before-treatment period, t = 8 is the after-treatment period.

For Black:
$\Delta^{DD} = [Y^1_{\text{Prison,8}} - Y^0_{\text{Prison, -12}}] - [Y^0_{\text{Probation,8}} - Y^0_{\text{Probation, -12}}] = 0.08 - 0.24 - (0.28 - 0.35) = -0.09$

For white:
$\Delta^{DD} = [Y^1_{\text{Prison,8}} - Y^0_{\text{Prison, -12}}] - [Y^0_{\text{Probation,8}} - Y^0_{\text{Probation, -12}}] = 0.14 - 0.38 - (0.42 - 0.47) = -0.19$

9. There are four assumptions specific to the RDD that are at least partially empirically verifiable and must be assessed prior to analysis. They are as follows:
    i.   There is a discontinuity in the probability of exposure at the cut-off.
    ii.   Individuals' value of the forcing variable was not manipulated.
    iii.   Exposure groups are exchangeable around the cut-off.
    iv.   The outcome probability is continuous at the cut-off in the absence of the intervention.

10. (1) Figures 2–6 and 7–11 provide evidence that discontinuities in treatment exist within rows and columns of the grid, respectively. Supporting the treatment (incarceration) varies discontinuously at this cutoff.

    (2) The author does not find any systematic variation in a large set of observable characteristics that is correlated with recidivism and that varies discontinuously around the cutoffs supports this assumption. In fact, the estimated treatment effect does not change when controlling for a wide array of variables.

    (3) In order to perfectly follow the assignment mechanism. The author dropped the 738 youths whose sentences do not appear from the sample. Treatment in the remaining sample is perfectly assigned by the rules of the grid.

    Last but not least, the prosecutors could not use discretion to systematically charge juveniles with crimes that place them on either side of a cutoff. Also, individuals cannot control the value of their score.

11. The author's analysis is a sharp RDD. There is complete compliance. No one below the cutoff gets treatment, and everyone above takes it up. *D_above_cutoff* is the treatment variable, indicating whether an individual fall in a cell of the grid that prescribes state incarceration.

12. I would make the author's regression model to be a fuzzy RDD. Clearly there is non-compliance, in both directions. According to the author, there are 424 youths who are not sentenced to DJR but whom the author placed in a cell prescribing incarceration, while 314 youths are sentenced to DJR but are not placed in such a cell. Thus, if I could make one change I would make it a fuzzy RDD.

Submitted by Tianchu Shu