Recitation 10
IDENTIFICATION TECHNIQUES

- What are identification techniques used for?

- What are the identification techniques that we discussed in class?
Suppose we want to identify the causal impact of $X$ on $Y$

- OLS: $Y_i = \alpha + \beta X_i + \epsilon_i$

- What are the identification assumptions?
  - $\epsilon$ is assumed to be uncorrelated with $X$. If $X$ is a dummy,
    $E[\epsilon|X = 0] = E[\epsilon|X = 1]$
  - what if we include a control variable $W$?
**Instrumental Variables**

- **When do we use IV?**
  - We want to estimate the causal impact of $X$ on $Y$ but believe that $X$ is endogenous:
    - reverse causality: $Y$ also explains $X$
    - missing variable correlated with $X$
  - We have an instrument $Z$

- **Suppose $Z$ is a dummy variable.**
  - What is the first stage?
    $$E[X_i|Z_i = 1] - E[X_i|Z_i = 0]$$
  - What is the reduced form?
    $$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$$

- **What are the identification assumptions?**
  - There must be a first stage: $Z$ must affect $X$
  - Exclusion restriction: $E[\epsilon|Z = 0] = E[\epsilon|Z = 1]$

- **What is the Wald estimator?**
When do we put fixed effects?

Equation we estimate: \( Y_{i,c} = \alpha_c + \beta X_{i,c} + \epsilon_{i,c} \) where \( \alpha_c \) are (for instance) country fixed effects.

What variation is then used to measure the impact of \( X \) on \( Y \)?
Regression Discontinuity Design

- $T_i = 1$ if $Z_i \geq c$ and $T_i = 0$ if $Z_i < c$

- Identification assumption:
  \[
  \lim_{x \downarrow c} E[Y_i(0)|Z_i = x] = \lim_{x \uparrow c} E[Y_i(0)|Z_i = x]. \text{ What’s } Y_i(0)?
  \]

- What equation do we estimate?
  \[
  Y_i = \alpha + \beta 1_{Z_i \geq c} + \epsilon
  \]
Suppose we come up with a identification strategy to regress growth on a dummy equal to 1 if the country is democratic. We add controls for disease environment and whether the country is landlocked.

### Results

<table>
<thead>
<tr>
<th></th>
<th>Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democracy</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
</tr>
<tr>
<td>Disease environment</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
</tr>
<tr>
<td>Landlocked</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
</tr>
</tbody>
</table>

Test $H_o$: democracy does not affect growth. What do you conclude?
How to read a table?

- Results

<table>
<thead>
<tr>
<th></th>
<th>Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democracy</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
</tr>
<tr>
<td>Disease environment</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
</tr>
<tr>
<td>Landlocked</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
</tr>
</tbody>
</table>

- What is your estimate of the impact of democracy on growth? **MAKE A SENTENCE!!!!** What if the outcome was ln(Growth) instead of Growth?

- What is a confidence interval? What is the confidence interval for the estimate of the impact of democracy on growth?

- Can we conclude that democracy is the only factor affecting growth?
Recitation 10

Review Session for the Final Exam

Theory Problem