RCR: Bounding a linear causal effect using relative correlation restrictions

Brian Krauth, Simon Fraser University
Stata User’s Group Meeting, July 2018
RCR

• rcr is a package that implements the estimator from “Bounding a linear causal effect using relative correlation restrictions,” Journal of Econometric Methods 2016.
The problem

• You want to measure the effect ($\beta_x$) of some variable of interest ($x_i$) on some outcome ($y_i$).

• You estimate a linear regression model of the form:

$$y_i = x_i \beta_x + c_i \beta_c + v_i$$

where $c_i$ is a vector of control variables.

• For this to work, you need:

$$\text{corr}(x_i, v_i) = 0$$

but you aren’t quite sure that’s true.
RCR analysis: Model

Ordinary regression analysis

- Model is:
  \[ y_i = x_i \beta_x + c_i \beta_c + v_i \]
- We assume:
  \[ \text{corr}(x_i, v_i) = 0 \]

RCR analysis

- Model is:
  \[ y_i = x_i \beta_x + c_i \beta_c + v_i \]
- We assume:
  \[ \text{corr}(x_i, v_i) = \lambda \text{corr}(x_i, c_i \beta_c) \]
  where \( \lambda \in [\lambda_L, \lambda_H] \)
- Ordinary regression is a special case in which \( \lambda_L = \lambda_H = 0 \).
RCR analysis: Identification

Ordinary regression analysis

- For $\beta_x$ we can obtain
  - A point estimate $\hat{\beta}_x$
  - A standard error for $\hat{\beta}_x$
  - A confidence interval $[\hat{\beta}_x^L, \hat{\beta}_x^H]$
  - Hypothesis tests

RCR analysis

- For $\beta_x$ we can obtain
  - Bounds $[\hat{\beta}_x^L, \hat{\beta}_x^H]$
  - Standard errors for $\hat{\beta}_x^L$ and $\hat{\beta}_x^H$
  - A confidence interval $[\hat{\beta}_x^L, \hat{\beta}_x^H]$
  - Hypothesis tests
RCR analysis: Implementation

Ordinary regression analysis

• Implementation in Stata:

\[ \text{reg } y \ x \ c \]

RCR analysis

• Implementation in Stata:

\[ \text{rcr } y \ x \ c, \ \text{lambda}(0 \ 1) \]

if you assume \( \lambda \in [0,1] \)
What should $\lambda$ be?

- Strong assumptions (small $\lambda_H - \lambda_L$) mean strong identification (small $\hat{\beta}_x^H - \hat{\beta}_x^L$).
  - If $[\lambda_L, \lambda_H]$ is wide enough, then $[\hat{\beta}_x^L, \hat{\beta}_x^H] = (-\infty, \infty)$
  - If $[\lambda_L, \lambda_H]$ is narrow, then $[\hat{\beta}_x^L, \hat{\beta}_x^H]$ is narrow too.

- Natural points of interest:
  - If $\lambda = 0$, then $corr(x_i, v_i) = 0$. Usual assumption.
  - If $\lambda = 1$, then $corr(x_i, v_i) = corr(x_i, c_i \beta_c)$. (Unmeasurable) correlation with unobservables same as (measurable) correlation with control variables.
  - So $[\lambda_L, \lambda_H] = [0,1]$ is a reasonable starting point benchmark.
  - Can also ask how small $\lambda$ needs to be to preserve OLS results.
RCR options

- **lambda(numlist)**: Allows you to specify the range for the $\lambda$ parameter (default = `lambda(0 1)`)
- **cluster(varname)**: cluster-robust standard errors (default = no clustering)
- **vceadj(#)**: the covariance matrix is multiplied by this number, allows you to implement degrees-of-freedom corrections if you are using transformed data (default = `vceadj(1)`)
- **citype(string)**: Confidence interval method (default = `citype(“conservative”)`)
- **level(#)**: confidence level (default = `level(95)`)
- Can handle all *if, in* and *weight* options.
Applications

• Application #1: Project STAR
  – Krueger (1999) study of effect of class size on test scores in Tennessee elementary schools
  – Imperfectly implemented experimental design

• Application #2: Malaria
  – Bleakley (2010) study of effect of malaria on labor productivity in US states
  – Observational design with lots of control variables
Project STAR: Implementation

**Ordinary regression analysis**
- Implementation in Stata:

```stata
reg SAT Small_Class White_Asian Girl Free_Lunch White_Teacher Teacher_Experience Masters_Degree if Grade == 0, cluster(TCHID) absorb(SCHID)
```

**RCR analysis**
- Implementation in Stata:

```stata
rcr SAT Small_Class White_Asian Girl Free_Lunch White_Teacher Teacher_Experience Masters_Degree if Grade == 0, cluster(TCHID) vceadj(1.013558143577264) lambda(0 1) citype("imbens-manski")
```
## Project STAR: Results

### Ordinary regression analysis

**Linear regression**

- Number of obs = 5839
- \( F(7, \ 322) = 62.24 \)
- Prob > F = 0.0000
- R-squared = 0.0927
- Root MSE = 22.202

### RCR analysis

**RCR model**

- Number of obs = 5839
- Lower bound on lambda = 0
- Upper bound on lambda = 1

### SAT Coefficients (Robust)

|                | SAT | Coef.  | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------------|-----|--------|-----------|-------|-----|----------------------|
| Small Class    | 5.201503 | 1.031215 | 5.04 | 0.000 | 3.172733 | 7.230272 |
| White Asian    | 8.388708 | 1.35401 | 6.20 | 0.000 | 5.724884 | 11.05253 |
| Girl           | 4.382026 | .625326 | 7.01 | 0.000 | 3.151785 | 5.612266 |
| Free Lunch     | -13.07747 | .7688515 | -17.01 | 0.000 | -14.59008 | -11.56486 |
| White Teacher  | -1.08975 | .620766 | -1.77 | 0.077 | -2.361187 | .221762 |
| Teacher_Expe   | -1.08975 | .620766 | -1.77 | 0.077 | -2.361187 | .221762 |
| Masters Degree | -1.08975 | .620766 | -1.77 | 0.077 | -2.361187 | .221762 |
| Cons           | 47.09615 | 2.405294 | 19.58 | 0.000 | 42.36407 | 51.82822 |

### Lambda Coefficients (Robust)

|                | SAT | Coef.  | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|----------------|-----|--------|-----------|-----|-----|----------------------|
| lambdaInf      | 12.3106 | 8.372929 | 1.47 | 0.141 | -4.10004 | 28.72124 |
| beta_xInf      | 8.16971 | 43.9407 | 0.19 | 0.853 | -77.95248 | 94.2919 |
| lambda0        | 28.93549 | 146.1582 | 0.20 | 0.843 | -257.5293 | 315.403 |
| beta_xH        | 5.201503 | 1.03756 | 5.01 | 0.000 | 3.167923 | 7.235082 |
| betax          | 2.4865 | 7.210477 | 1.43 | 0.152 | -14.59008 | -11.56486 |

**Treatment Variable:** Small Class
**Control Variables:** White_Asian Girl Free_Lunch White_Teacher Teacher_Experience Masters_Degree
Project STAR: Interpretation

Ordinary regression analysis

• If class size is uncorrelated with unobservables, then
  – Being in a small class increases test scores by 5.2 percentile points
• 95% CI = (3.2, 7.2)

RCR analysis

• If class size is no more correlated with unobservables than it is with the control variables,
  – Being in a small class increases test scores by 5.1 to 5.2 percentile points
• 95% CI = (2.5, 7.2)

• If class size is less than 12.3 times as correlated with unobservables
  – The class size effect is positive.
Malaria: Implementation

Ordinary regression analysis

• Implementation in Stata:

```stata
reg docc malmort1890 south lebergott99 [aw=wtbpl]
```

RCR analysis

• Implementation in Stata:

```stata
rcr docc malmort1890 south lebergott99 [aw=wtbpl], lambda(0 1)
```
Malaria: Results

Ordinary regression analysis

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 48</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>.196184286</td>
<td>3</td>
<td>.065394762</td>
<td>F( 3, 44) = 15.52</td>
</tr>
<tr>
<td>Residual</td>
<td>.185372534</td>
<td>44</td>
<td>.004213012</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>.381556819</td>
<td>47</td>
<td>.00811823</td>
<td>R-squared = 0.5142</td>
</tr>
</tbody>
</table>

| Coefficient | Std. Err. | t   | P>|t| | [95% Conf. Interval] |
|-------------|-----------|-----|-------|---------------------|
| malmort1890 | .1124381  | .0476992 | 2.36 | 0.023 | [.0163066, .2085695] |
| south       | .04293    | .0370928 | 1.16 | 0.253 | [-.0318257, .1176857] |
| lebergott99 | -.0288005 | .0227818 | -1.26 | 0.213 | [-.0747141, .0171131] |
| _cons       | .2702744  | .0131139 | 20.61 | 0.000 | [.2438451, .2967038] |

RCR analysis

| Coefficient | Std. Err. | t   | P>|t| | [95% Conf. Interval] |
|-------------|-----------|-----|-------|---------------------|
| lambdaInf   | .9345488  | .1745427 | 5.35 | 0.000 | [.5924514, 1.276646] |
| betaxInf    | .2822161  | .0554518 | 5.09 | 0.000 | [.1735326, .3908997] |
| lambda0     | .3165399  | .1179838 | 2.68 | 0.007 | [.0853274, .5477523] |
| betaxL      | 5.0e+306  | . | . | . | . |
| betaxH      | 5.0e+306  | . | . | . | . |

| Coefficient | Std. Err. | t   | P>|t| | [95% Conf. Interval] |
|-------------|-----------|-----|-------|---------------------|
| betax       | . | . | . | . | . |

Treatment Variable: malmort1890
Control Variables: south lebergott99
Malaria: Interpretation

**Ordinary regression analysis**

- If malaria is uncorrelated with unobservables, then:
  - Malaria eradication increases the occupational index score by 0.11 points per unit of pre-eradication malaria mortality
- 95% CI = (0.03,0.19)

**RCR analysis**

- If malaria is no more correlated with unobservables than it is with the control variables, then:
  - We cannot put nontrivial bounds on the malaria eradication effect
- If class size is less than 0.32 times as correlated with unobservables, then:
  - The malaria eradication effect is positive.
Limitations

• Estimator is partly implemented in Fortran
  – Ready-to-run for Windows
  – Tyler Ransom has code and libraries for Linux
    (https://github.com/tylerransom/rcr)
Availability

• Full package available with
  – Ado-files for
    • `rcr` command
    • Postestimation commands for hypothesis testing, graphical tools, etc.
  – Help for all commands
  – Example code and data
• Go to [http://www.sfu.ca/~bkrauth/code.htm](http://www.sfu.ca/~bkrauth/code.htm)
  – or just Google “rcr Krauth”
• I’m happy to provide technical support, just email me.