# Discussion of Bacher-Hicks, Billings, and Deming (2020) – The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime

Chris Walters

UC Berkeley and NBER

NBER Summer Institute 2020

# Overview

- Bacher-Hicks, Billings, and Deming (BBD, 2020) study the impacts of school suspensions on crime and other outcomes
  - Estimate regression-adjusted school suspension rates (suspension "value-added")
  - Relate these to crime outcomes using Charlotte-Mecklenburg attendance boundary reform
  - Find substantial adverse effects of attending schools with high suspension rates
- Convincing evidence that high-suspension schools increase crime – an important and provocative new finding
- Suggests "suspension hypothesis:" Strict school discipline policies cause crime
  - Contrasts with lottery evidence for "No Excuses" charter schools (Dobbie and Fryer, 2014; Angrist et al., 2016)
- Discussion agenda: To what extent do BBD's findings support the suspension hypothesis?

# Research Question vs. Research Design

- Research question: BBD are interested in estimating the causal effect of suspensions on crime
  - Title: "Impacts of School Suspensions on Adult Crime"
  - Intro: "In this paper, we study the impact of school discipline on the achievement, educational attainment, and subsequent criminal activity of students"
- Research design: Changes in school attendance boundaries
  - Generates exogenous variation in school attendance, not suspensions
- What assumptions allow us to infer the effects of suspensions from a school attendance experiment?

# Estimation Procedure

- Two-stage estimation procedure:

$$S_i = \sum_j \mu_j D_{ij} + X_i' \delta + \eta_i,$$

$$Y_i = \beta \hat{S}_i + X_i' \gamma + \epsilon_i$$

- $Y_i$ is crime, $D_{ij}$ are school dummies, $S_i$ is suspension, $X_i$ are controls, $\hat{S}_i$ is predicted value from first stage
- $D_{ij}$'s assumed as good as randomly assigned conditional on $X_i$
- Implementation details:
  - Different samples and controls in each stage
  - Shrink $\hat{S}_i$ to deal with finite-sample noise
  - School attendance vs. assignment
- Details matter, but basic interpretation unaffected

# 2SLS Interpretation

$$S_i = \sum_j \mu_j D_{ij} + X_i'\delta + \eta_i,$$

$$Y_i = \beta \hat{S}_i + X_i'\gamma + \epsilon_i$$

- This is a two-stage least squares (2SLS) approach
  - Endogenous variable: Suspensions
  - Instruments: School indicators
- How should we interpret $\beta$?

# Interpretation 1: Standard 2SLS

$$S_i = \sum_j \mu_j D_{ij} + X_i' \delta + \eta_i,$$

$$Y_i = \beta \hat{S}_i + X_i' \gamma + \epsilon_i$$

- In a standard 2SLS model, we are trying to estimate the effect of the endogenous (instrumented) variable
- $\beta$ is interpreted as the causal effect of a suspension on crime – the answer to BBD's research question
- But requires exclusion restriction: instruments (schools) only affect crime through suspensions. This does not seem plausible
  - Deming (2011): schools affect crime through peer networks and human capital

# Interpretation 2: School-Level Regression Coefficient

- Reduced form and first stage for 2SLS system:

$$Y_i = \sum_j \theta_j D_{ij} + X_i' \psi + u_i,$$

$$S_i = \sum_j \mu_j D_{ij} + X_i' \delta + \eta_i$$

- 2SLS produces (weighted) regression of RF on FS:

$$\beta \approx \frac{Cov(\theta_j, \mu_j)}{Var(\mu_j)}$$

- Interpret $\beta$ as the coefficient from a linear projection of a school's crime effect on its suspension effect
- More palatable than defending exclusion restriction, and establishes that suspension effects predict crime effects

# IV Without Exclusion?

- But without the exclusion restriction, $\beta$ doesn't answer BBD's research question
  - Not the effect of suspension on crime for an individual kid
  - Not the effect of changing a school's suspension rate on crime
- Possible solution: Maybe exclusion violations average out
  - "Many invalid instruments" idea (Kolesar et al., 2015)
  - Some evidence that $\mu_j$ is uncorrelated with observables (Fig. 5)
  - But beware the ecological regression!
- Without exclusion, group-level relationship between suspensions and crime could be driven by generic clustering or general impacts on behavior
  - Rose, Schellenberg, and Shem-Tov (2019): teacher effects on future suspensions and crime are correlated due to "non-cognitive value-added"

# Alternative: Focus on the Reduced Form

- With a good research design for school attendance, why not estimate school effects on crime ($\theta_j$'s) directly?
  - Crime effects and suspension effects are identified under similar assumptions
  - No exclusion restriction required!
- $Var(\theta_j)$ speaks to "school to prison pipeline" question
- Many interesting potential predictors of crime effects besides suspension effects: school inputs, characteristics of schools, teachers, students, neighborhoods
- If suspension effects are the dominant predictor, perhaps evidence in favor of the suspension hypothesis

# Thanks

- Thanks, great paper!