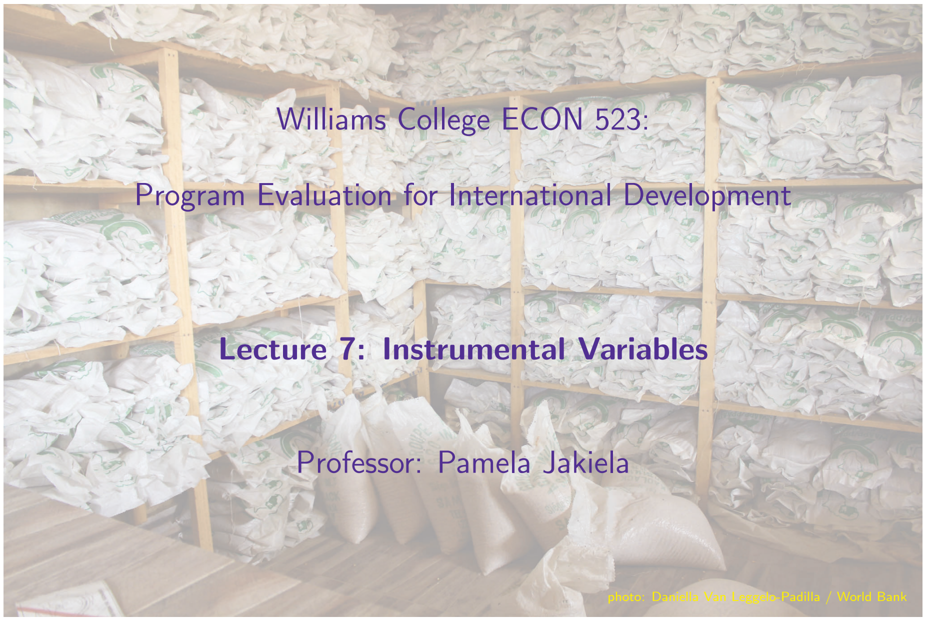


Williams College ECON 523:

Program Evaluation for International Development

Lecture 7: Instrumental Variables

Professor: Pamela Jakiela

Instrumental Variables in the Wild

We want to estimate the impact of program P on outcome Y: $Y_i = \alpha + \beta P_i + \varepsilon_i$

- Program participation (P) is not randomly assigned, leading to selection bias

$$E[Y_{0i}|P_i = 1] \neq E[Y_{0i}|P_i = 0] \quad (1)$$

- Some determinants of outcomes (V) are unobservable and correlated with participation

$$Y_i = \alpha + \beta P_i + \phi W_i + \psi V_i + \epsilon_i \quad (2)$$

- Instrumental variables relies on a source of exogenous variation in participation (Z)

$$P_i = \zeta + \lambda Z_i + \theta W_i + \gamma V_i + \xi_i \quad (3)$$

Requirements: $\lambda \neq 0$, Z is independent of V, Z doesn't enter Equation (2)

Instrumental Variables in the Wild

A valid instrument must move P but have no direct effect on outcome of interest Y

- Selection bias reflects differences in potential outcomes between treated, untreated

$$E[Y_{0i}|P_i = 1] \neq E[Y_{0i}|P_i = 0]$$

- Individual partition decision does not only depend on Y, difference in potential outcomes
  - Y may or may not be salient, relevant to decision-maker
  - Y often reflects future costs and benefits of participation
  - Decision to participate also depends on immediate costs and benefits
- In **natural experiments**, "decision" to participate is made exogenously (e.g. twins)

Instrumental Variables in the Wild: Rainfall

Paxson (1992) uses rainfall as a source of exogenous variation in income (in rural Thailand)

- Rainfall has been widely used as an instrumental variable to estimate impact of (negative) income (shocks) and poverty on (eg) remittances and consumption (Yang and Choi 2007), crime (Iyer and Topolova 2014), conflict (Miguel, Satyanath, and Sargenti 2004)
- Sarsons (2015) finds evidence against the exclusion restriction in crime/conflict studies: in India, agricultural output is less reliant on rainfall in districts downstream from dams; but communal violence responds to rainfall in the same way in dam and non-dam districts

Instrumental Variables in the Wild: Rainfall

Rainfall has a direct effect on the costs of participations in outdoor protests, rallies, riots, etc.

- Collins and Margo (2007): rainfall on the day of MLK assassination (4 April 1968) reduced rioting, and (instrumented) race-based rioting reduced property values
- Madestam, Shoag, Veuger, and Yanagizawa-Drott (2013): Tea Party movement stronger, Republican ↑ vote share higher, with good weather on April 15, 2009 (day of first rally)
- Larreboure and Gonzalez (2021): 2017 Women's March, 2018 House elections

Instrumental Variables in the Wild: The Calendar

Dates of religious holidays based on lunar calendar, not aligned with Gregorian calendar

- Iyer and Shrivastava (2018): odds of Hindu-Muslim rioting ↑ when Hindu festival falls on a Friday (first stage); religious riots increase BJP (Hindu nationalist party) vote share
- Intensity of Ramadan fasting depends on latitude and seasonal timing (first stage):
    - Campante and Yanagizawa-Drott (2013): increased fasting reduces output growth in predominantly Muslim countries, but increases subjective wellbeing among Muslims
    - Aksoy, Gambetta et al. (2022): longer fasting increases Islamist vote share
    - Mehmood, Seror, and Chen (2023): Muslim judges in India and Pakistan more lenient when they fast more, no effect for non-Muslim judges trying similar cases, no impact on recidivism

Instrumental Variables in the Wild: Distance

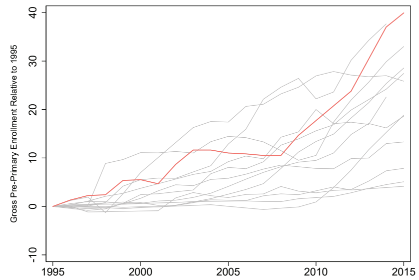
Trade literature suggests that trade costs are proportional to distance

- Nunn (2008): distance to New World slave-trading ports (Brazil, Caribbean) as an instrument for historical exposure to African slave trades (which impacts GDP per capita)
- McClellan and Newhouse (1994): distance to hospitals that perform specific procedures to instrument for intensive treatment for acute myocardial infarction (heart attacks)
- Card (1995), Dee (2004): distance to college as an instrument for enrollment

Whether distance instruments satisfy exogeneity, exclusion restriction likely to vary with context

Instrumental Variables in Practice

Early Childhood Education: Access and Quality



SDG 4.2: "...access to quality ECD care and pre-primary education..."

In Kenya, enrollment went from 47 percent in 2005 to 76 percent in 2015

Early Childhood Education: Access and Quality

Does access to preprimary education improve child development outcomes?

Access to public ECD education does not always translate into use

- Take-up rates can be low (Naudeau et al. 2017, Bouguen et al. 2018)
- Free preprimary is often not offered all day (Piper et al. 2018)

Quality is an issue (Black et al. 2017, Özler et al. 2018)

- Parents and teachers may have different objectives (Wolf et al. 2019)

Low-quality early childhood education may be worse than no ECE (Fort, Inchino, Zanella 2020)

- Existing evaluations typically estimate impacts of new schools (Naudeau et al. 2017, Bouguen et al. 2018), which may differ from impacts of existing ECE programs

Early Childhood Education: Access and Quality

Does access to preprimary education improve child development outcomes?

Access to public ECD education does not always translate into use

- Take-up rates can be low (Naudeau et al. 2017, Bouguen et al. 2018)
- Free preprimary is often not offered all day (Piper et al. 2018)

Quality is an issue (Black et al. 2017, Özler et al. 2018)

- Parents and teachers may have different objectives (Wolf et al. 2019)

Low-quality early childhood education may be worse than no ECE (Fort, Inchino, Zanella 2020)

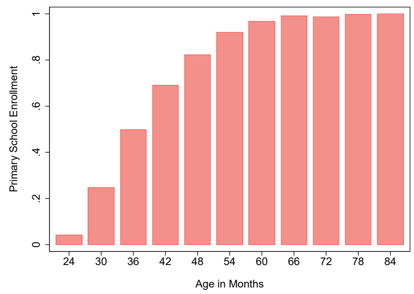
- Existing evaluations typically estimate impacts of new schools (Naudeau et al. 2017, Bouguen et al. 2018), which may differ from impacts of existing ECE programs

Study Context: Early Childhood Education in Kenya

In 2017, we conducted a baseline survey prior to implementation of an early literacy intervention

- Until 2018, Kenyan government schools followed 8-4-4 educational system implemented
- Most schools also offered three levels of preprimary: Baby Class, Nursery, and Pre-Unit

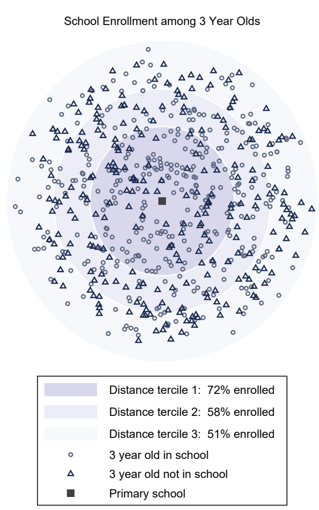
School Enrollment Increases with Age in Kenya

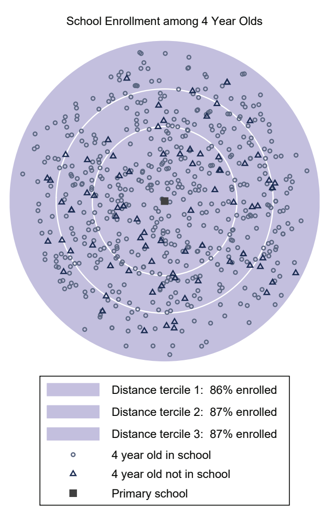


We estimate the impact of Kenya's public pre-primary education classes on child development

- Sample of three year olds: probability of starting school at age three varies with distance

Distance from School Predicts Age of Entry in Kenya





First Stage: Predicting Enrollment

Our first-stage regression is:

$$P_{ih} = \alpha_1 + \delta D_h + \lambda_1 X_{ih} + \varepsilon_{ih}$$

where

- P_{ih} is an indicator equal to one if child i in household h is enrolled in preprimary
- D_h is the distance from household h to the primary school
- X_{ih} is a vector of household characteristics
- ε_{ih} is a conditionally-mean-zero error term

First Stage: Predicting Enrollment

Age in Years:	**3 Years**	**4 Years**	**5 Years**	**6 Years**
	OLS	**OLS**	**OLS**	**OLS**
	(1)	(2)	(3)	(4)
Panel A: without covariates				
Distance to school (km)	-0.496	0.003	-0.041	0.009
	(0.116)	(0.075)	(0.025)	(0.012)
	[p<0.001]	[0.969]	[0.099]	[0.469]
Panel B: covariate-adjusted				
Distance to school (km)	-0.463	-0.052	-0.057	0.002
	(0.111)	(0.073)	(0.028)	(0.010)
	[p<0.001]	[0.475]	[0.047]	[0.828]
Obs.	634	610	669	590

Standard errors in parentheses, p-values in brackets. Dependent variable is indicator for being enrolled in school. Covariates included in Panel B: child age in month (fixed effects), child gender (indicator for male), child height-for-age z-score, an indicator equal to one if a child's mother is their primary caregiver, mother's education, an indicator for having a Luo mother, household size, the number of older siblings in the household, and a household wealth index.

Is the Instrument Plausibly Exogenous?

	Coefficient	S.E.	p-value
Child age (in months)	-0.511	0.793	0.519
Height-for-age z-score	-0.432	0.339	0.203
Child is male	0.025	0.121	0.836
Mother is child's primary caregiver	0.000	0.080	0.995
Mother's education in years	0.347	0.619	0.575
Mother is Luo	-0.044	0.054	0.420
Father absent from household	-0.000	0.084	0.996
Father's education in years	1.137	0.666	0.088
Father is Luo	0.074	0.040	0.062
Household size	0.419	0.510	0.412
Older siblings in household	0.528	0.301	0.080
Asset index (out of 10)	0.541	0.355	0.128

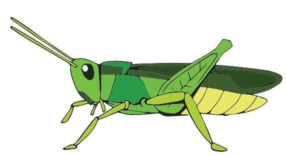
Coefficients from OLS regressions of outcome variables on distance from the school (in km).

Outcome Variables

"Show me the DOG"

Receptive Vocabulary

"What is this?"



Expressive Vocabulary

Outcome Variables

Outcome variables from baseline survey of early childhood development:

- **Luo receptive vocabulary:** language of instruction in preschool
  - New test developed by adapting British Picture Vocabulary Scale (Knauer et al. 2019)
- **English receptive vocabulary:** not language of instruction in ECD
  - Adapted British Picture Vocabulary Scale, a widely used vocabulary test
- **Expressive vocabulary:**
  - New test developed for intervention (Knauer et al. 2019)
- **Fine motor index:**
  - Selected items from Malawi Developmental Assessment Tool (draw a cross, block tower)

Estimation Strategy

Impacts of enrollment in ECD estimated via two-stage least squares (2SLS):

$$P_{ih} = \alpha_1 + \beta Distance_i + \lambda_1 X_{ih} + \varepsilon_{ih} \quad \text{[first stage]}$$

$$Y_{ih} = \alpha_2 + \delta \hat{P}_i + \lambda_2 X_{ih} + \xi_{ih} \quad \text{[second stage]}$$

where

- Y_{ih} is the outcome of interest, measured at the child level
- P_{ih} is an enrollment dummy, and $\hat{P}_{ih}$ is predicted enrollment
- X_{ih} is a vector of controls (gender, HAZ, mother's eduction, etc.)
- ε_{ih} is a conditionally mean-zero error term, clustered by household

IV Estimates of the Impact of Pre-Primary

	Vocabulary				
	Luo	**English**	**Expressive**	**Fine Motor**	**ECD Index**
	(1)	(2)	(3)	(4)	(5)
Panel A: without covariates					
Enrolled in ECD	1.160	-0.131	0.701	0.581	0.780
	(0.520)	(0.483)	(0.475)	(0.444)	(0.471)
	[0.026]	[0.786]	[0.140]	[0.191]	[0.098]
Panel B: covariate-adjusted					
Enrolled in ECD	1.528	-0.151	0.954	0.722	1.030
	(0.607)	(0.500)	(0.486)	(0.443)	(0.477)
	[0.012]	[0.762]	[0.050]	[0.103]	[0.031]
Obs.	634	634	634	634	634

All specifications estimated via 2-stage least squares (2SLS). First-stage F-statistics: 18.28 (Panel A) and 17.25 (Panel B). Standard errors in parentheses, p-values in brackets. Covariates included in Panel B: child age in month (fixed effects), child gender (indicator for male), child height-for-age z-score, an indicator equal to one if a child's mother is their primary caregiver, mother's education, an indicator for having a Luo mother, household size, the number of older siblings in the household, and a household wealth index.

Does pre-primary impact child development?

Coefficient estimates suggest that pre-primary has a big* impact

*Big but imprecisely estimated

- IV coefficients 2–4 times larger than OLS coefficients
- We can never reject equality (IV confidence intervals are large)

Do these short-term impacts persist? Do they increase over time?

- If we assume distance also predicts age of school entry in earlier cohorts, we can estimate reduced form regressions for all children
- EMERGE endline survey will collect retrospective data on enrollment

Reduced Form "Impacts" on Luo Vocabulary

Age in Years:	**3 Years**	**4 Years**	**5 Years**	**6 Years**
	OLS	**OLS**	**OLS**	**OLS**
	(1)	(2)	(3)	(4)
Distance to school (km)	-0.575	-0.256	-0.052	-0.075
	(0.246)	(0.248)	(0.241)	(0.249)
	[0.020]	[0.302]	[0.828]	[0.763]
Obs.	618	600	665	574

Standard errors in parentheses, p-values in brackets.