



Empirical Economics

Case-based seminar I - Differen in Differences



Teacher: Andrew Proctor
andrew.proctor@phdstudent.hhs.se

September 14, 2017



Outline

① Card and Krueger (AER, 1993)

Introduction

Empirical Strategy

Descriptive Statistics

Results

② Fortson (RESTAT, 2011)

Introduction

Descriptive Statistics

Main Regression

Robustness Checks



Today's main paper

The assigned paper for today is:

“Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania”

David Card and Alan Krueger

American Economic Review (1994)



Introduction

What is the general topic that Card and Krueger are interested in?



Introduction

What is the general topic that Card and Krueger are interested in?

- Card and Krueger are interested in the effect of minimum wages on employment.

What is the specific research question addressed in this analysis?



Introduction

What is the general topic that Card and Krueger are interested in?

- Card and Krueger are interested in the effect of minimum wages on employment.

What is the specific research question addressed in this analysis?

- How did the minimum wage increase from (\$4.25 to \$5.05) in New Jersey during 1992 affect employment in fast-food restaurants.

What data is used for this research?



Introduction

What is the general topic that Card and Krueger are interested in?

- Card and Krueger are interested in the effect of minimum wages on employment.

What is the specific research question addressed in this analysis?

- How did the minimum wage increase from (\$4.25 to \$5.05) in New Jersey during 1992 affect employment in fast-food restaurants.

What data is used for this research?

- Interview data from 410 fast-food restaurants across two waves (time periods).



Source of 'Identification'

What are the basic features of Card and Krueger's dataset that they claim allows them to estimate the effect of the minimum wage increase on employment



Source of 'Identification'

What are the basic features of Card and Krueger's dataset that they claim allows them to estimate the effect of the minimum wage increase on employment

- 1 Using the difference-in-difference strategy, they are able to control for anything fixed within New Jersey and Eastern Pennsylvania or that varies in common across the period of change.
- 2 Card and Krueger claim that because New Jersey is small and it's economy is closely related to eastern Pennsylvania, the only significant thing affecting the difference in outcomes for fast food workers between New Jersey and Pennsylvania is the minimum wage law



Source of 'Identification' ctd

Why do Card and Krueger use the fast food industry in particular?



Source of 'Identification' ctd

Why do Card and Krueger use the fast food industry in particular?

- Card and Krueger claims that the fast food industry is a good subject because it is both one of the largest employers of low-wage workers and that its relatively homogeneous business structure allows for straightforward comparisons.



About Differences-in-Differences in Card and Krueger

- The typical difference-in-difference strategy controls for time- and group-invariant characteristics via the following specification:

$$y_{it} = \alpha + \beta_0 \text{After}_t + \beta_1 \text{Treated}_i + \delta \text{After}_t \times \text{Treated}_i + \text{other factors},$$

Where *treated* is an indicator for a group affected by some reform and "after" is an indicator indicating whether or not the treatment has happened in a given period.



About Differences-in-Differences in Card and Krueger

- Card and Krueger essentially combine this with a first-differencing approach:

$$y_{i1} = \alpha + \beta_0 \text{After}_1 + \beta_1 \text{Treated}_i + \delta \text{After}_1 \times \text{Treated}_i + \text{other factors}$$

$$-y_{i0} = \alpha + \beta_0 \text{After}_0 + \beta_1 \text{Treated}_i + \delta \text{After}_0 \times \text{Treated}_i + \text{other factors}$$

$$\Delta y_i = \beta_0 + \delta \text{Treated}_i,$$

Only including some time varying controls that aren't differenced out in the final regression.



TABLE 1—SAMPLE DESIGN AND RESPONSE RATES

	All	Stores in:	
		NJ	PA
<i>Wave 1, February 15–March 4, 1992:</i>			
Number of stores in sample frame: ^a	473	364	109
Number of refusals:	63	33	30
Number interviewed:	410	331	79
Response rate (percentage):	86.7	90.9	72.5
<i>Wave 2, November 5–December 31, 1992:</i>			
Number of stores in sample frame:	410	331	79
Number closed:	6	5	1
Number under renovation:	2	2	0
Number temporarily closed: ^b	2	2	0
Number of refusals:	1	1	0
Number interviewed: ^c	399	321	78

^aStores with working phone numbers only; 29 stores in original sample frame had disconnected phone numbers.

^bIncludes one store closed because of highway construction and one store closed because of a fire.

^cIncludes 371 phone interviews and 28 personal interviews of stores that refused an initial request for a phone interview.





TABLE 2—MEANS OF KEY VARIABLES

Variable	Stores in:		t^a
	NJ	PA	
1. <i>Distribution of Store Types (percentages):</i>			
a. Burger King	41.1	44.3	-
b. KFC	20.5	15.2	1.2
c. Roy Rogers	24.8	21.5	0.6
d. Wendy's	13.6	19.0	-1.1
e. Company-owned	34.1	35.4	-0.2





TABLE 2—MEANS OF KEY VARIABLES

Variable	Stores in:		t^a
	NJ	PA	
<i>2. Means in Wave 1:</i>			
a. FTE employment	20.4 (0.51)	23.3 (1.35)	-2.0
b. Percentage full-time employees	32.8 (1.3)	35.0 (2.7)	-0.7
c. Starting wage	4.61 (0.02)	4.63 (0.04)	-0.4
d. Wage = \$4.25 (percentage)	30.5 (2.5)	32.9 (5.3)	-0.4
e. Price of full meal	3.35 (0.04)	3.04 (0.07)	4.0
f. Hours open (weekday)	14.4 (0.2)	14.5 (0.3)	-0.3
g. Recruiting bonus	23.6	29.1	-1.0





TABLE 2—MEANS OF KEY VARIABLES

3. Means in Wave 2:

a. FTE employment	21.0 (0.52)	21.2 (0.94)	-0.2
b. Percentage full-time employees	35.9 (1.4)	30.4 (2.8)	1.8
c. Starting wage	5.08 (0.01)	4.62 (0.04)	0.8
d. Wage = \$4.25 (percentage)	0.0	25.3 (4.9)	—
e. Wage = \$5.05 (percentage)	85.2 (2.0)	1.3 (1.3)	36.1
f. Price of full meal	3.41 (0.04)	3.03 (0.07)	5.0
g. Hours open (weekday)	14.4 (0.2)	14.7 (0.3)	-0.8
h. Recruiting bonus	20.3 (2.3)	23.4 (4.9)	-0.6



TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE
 IN NEW JERSEY MINIMUM WAGE

Variable	Stores by state			Stores in New Jersey ^a			Differences within NJ ^b	
	PA (i)	NJ (ii)	Difference, NJ – PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low– high (vii)	Midrange– high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	–2.89 (1.44)	9.56 (0.77)	20.08 (0.84)	22.25 (1.14)	–2.69 (1.37)	–2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	–0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	–2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	–2.04 (1.14)	3.36 (1.48)	2.91 (1.41)
4. Change in mean FTE employment, balanced sample of stores ^c	–2.28 (1.25)	0.47 (0.48)	2.75 (1.34)	1.21 (0.82)	0.71 (0.69)	–2.16 (1.01)	3.36 (1.30)	2.87 (1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0 ^d	–2.28 (1.25)	0.23 (0.49)	2.51 (1.35)	0.90 (0.87)	0.49 (0.69)	–2.39 (1.02)	3.29 (1.34)	2.88 (1.23)




$$(1a) \quad \Delta E_i = a + \mathbf{b}\mathbf{X}_i + c\text{NJ}_i + \varepsilon_i$$

or

$$(1b) \quad \Delta E_i = a' + \mathbf{b}'\mathbf{X}_i + c'\text{GAP}_i + \varepsilon'_i$$

where ΔE_i is the change in employment from wave 1 to wave 2 at store i , \mathbf{X}_i is a set of characteristics of store i , and NJ_i is a dummy variable that equals 1 for stores in New Jersey. GAP_i is an alternative measure of the impact of the minimum wage at store i based on the initial wage at that store (W_{1i}):



$$(1a) \quad \Delta E_i = a + \mathbf{b}\mathbf{X}_i + c\text{NJ}_i + \varepsilon_i$$


or

$$(1b) \quad \Delta E_i = a' + \mathbf{b}'\mathbf{X}_i + c'\text{GAP}_i + \varepsilon'_i$$

where ΔE_i is the change in employment from wave 1 to wave 2 at store i , \mathbf{X}_i is a set of characteristics of store i , and NJ_i is a dummy variable that equals 1 for stores in New Jersey. GAP_i is an alternative measure of the impact of the minimum wage at store i based on the initial wage at that store (W_{1i}):

TABLE 4—REDUCED-FORM MODELS FOR CHANGE IN EMPLOYMENT

Independent variable	Model				
	(i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	2.33 (1.19)	2.30 (1.20)	—	—	—
2. Initial wage gap ^a	—	—	15.65 (6.08)	14.92 (6.21)	11.91 (7.39)
3. Controls for chain and ownership ^b	no	yes	no	yes	yes
4. Controls for region ^c	no	no	no	no	yes
5. Standard error of regression	8.79	8.78	8.76	8.76	8.75
6. Probability value for controls ^d	—	0.34	—	0.44	0.40



TABLE 5—SPECIFICATION TESTS OF REDUCED-FORM EMPLOYMENT MODELS





Specification	Change in employment		Proportional change in employment	
	NJ dummy (i)	Gap measure (ii)	NJ dummy (iii)	Gap measure (iv)
1. Base specification	2.30 (1.19)	14.92 (6.21)	0.05 (0.05)	0.34 (0.26)
2. Treat four temporarily closed stores as permanently closed ^a	2.20 (1.21)	14.42 (6.31)	0.04 (0.05)	0.34 (0.27)
3. Exclude managers in employment count ^b	2.34 (1.17)	14.69 (6.05)	0.05 (0.07)	0.28 (0.34)
4. Weight part-time as 0.4 × full-time ^c 	2.34 (1.20)	15.23 (6.23)	0.06 (0.06)	0.30 (0.33)
5. Weight part-time as 0.6 × full-time ^d	2.27 (1.21)	14.60 (6.26)	0.04 (0.06)	0.17 (0.29)
6. Exclude stores in NJ shore area ^e 	2.58 (1.19)	16.88 (6.36)	0.06 (0.05)	0.42 (0.27)
7. Add controls for wave-2 interview date ^f	2.27 (1.20)	15.79 (6.24)	0.05 (0.05)	0.40 (0.26)
8. Exclude stores called more than twice in wave 1 ^g 	2.41 (1.28)	14.08 (7.11)	0.05 (0.05)	0.31 (0.29)
9. Weight by initial employment ^h	—	—	0.13 (0.05)	0.81 (0.26)
10. Stores in towns around Newark ⁱ 	—	33.75 (16.75)	—	0.90 (0.74)
11. Stores in towns around Camden ^j	—	10.91 (14.09)	—	0.21 (0.70)
12. Pennsylvania stores only ^k	—	-0.30 (22.00)	—	-0.33 (0.74)





TABLE 6—EFFECTS OF MINIMUM-WAGE INCREASE ON OTHER OUTCOMES

Outcome measure	Mean change in outcome			Regression of change in outcome variable on:		
	NJ (i)	PA (ii)	NJ – PA (iii)	NJ dummy (iv)	Wage gap ^a (v)	Wage gap ^b (vi)
<i>Store Characteristics:</i>						
1. Fraction full-time workers ^c (percentage)	2.64 (1.71)	-4.65 (3.80)	7.29 (4.17)	7.30 (3.96)	33.64 (20.95)	20.28 (24.34)
2. Number of hours open per weekday	-0.00 (0.06)	0.11 (0.08)	-0.11 (0.10)	-0.11 (0.12)	-0.24 (0.65)	0.04 (0.76)
3. Number of cash registers	-0.04 (0.04)	0.13 (0.10)	-0.17 (0.11)	-0.18 (0.10)	-0.31 (0.53)	0.29 (0.62)
4. Number of cash registers open at 11:00 A.M.	-0.03 (0.05)	-0.20 (0.08)	0.17 (0.10)	0.17 (0.12)	0.15 (0.62)	-0.47 (0.74)
<i>Employee Meal Programs:</i>						
5. Low-price meal program (percentage)	-4.67 (2.65)	-1.28 (3.86)	-3.39 (4.68)	2.01 (5.63)	-30.31 (29.80)	-33.15 (35.04)
6. Free meal program (percentage)	8.41 (2.17)	6.41 (3.33)	2.00 (3.97)	0.49 (4.50)	29.90 (23.75)	36.91 (27.90)
7. Combination of low-price and free meals (percentage)	-4.04 (1.98)	-5.13 (3.11)	1.09 (3.69)	1.20 (4.32)	-11.87 (22.87)	-19.19 (26.81)

TABLE 7—REDUCED-FORM MODELS FOR CHANGE IN THE PRICE OF A FULL MEAL


Independent variable	Dependent variable: change in the log price of a full meal				
	(i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	0.033 (0.014)	0.037 (0.014)	 —	—	—
2. Initial wage gap ^a	—	—	0.077 (0.075)	0.146 (0.074)	0.063 (0.089)
3. Controls for chain and ownership ^b	no	yes	no	yes	yes
4. Controls for region ^c	no	no	no	no	yes
5. Standard error of regression	0.101	0.097	0.102	0.098	0.097



TABLE 8—ESTIMATED EFFECT OF MINIMUM WAGES ON NUMBERS OF MCDONALD'S RESTAURANTS, 1986–1991

Independent variable	Dependent variable: proportional increase in number of stores				Dependent variable: (number of newly opened stores) ÷ (number in 1986)			
	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
<i>Minimum-Wage Variable:</i>								
1. Fraction of retail workers in affected wage range 1986 ^a	0.33 (0.20)	—	0.13 (0.19)	—	0.37 (0.22)	—	0.16 (0.21)	—
2. (State minimum wage in 1991) ÷ (average retail wage in 1986) ^b	—	0.38 (0.22)	—	0.47 (0.22)	—	0.47 (0.23)	—	0.56 (0.24)
<i>Other Control Variables:</i>								
3. Proportional growth in population, 1986–1991	—	—	0.88 (0.23)	1.03 (0.23)	—	—	0.86 (0.25)	1.04 (0.25)
4. Change in unemployment rates, 1986–1991	—	—	-1.78 (0.62)	-1.40 (0.61)	—	—	-1.85 (0.68)	-1.40 (0.65)
5. Standard error of regression	0.083	0.083	0.071	0.068	0.088	0.088	0.077	0.073



General Thoughts

- What are the assumptions necessary for the identification strategy to work?



General Thoughts

- What are the assumptions necessary for the identification strategy to work?
- Condition on controls used in a given specific, there should be nothing differently happening in New Jersey compared to eastern Pennsylvania that affects the changes in employment from period 1 to period 2.
- Is this believable?



General Thoughts

- What are the assumptions necessary for the identification strategy to work?
- Condition on controls used in a given specific, there should be nothing differently happening in New Jersey compared to eastern Pennsylvania that affects the changes in employment from period 1 to period 2.
- Is this believable?
- Not completely, but maybe. I would certainly be concerned though that even though the recession going on at that time can have largely varying affects. The fact that the regional indicators were significant implies there were different regional trends (even within states). But I am not convinced that just these indicators captures the heterogeneity between regions across time.



General Thoughts

- Another implicit assumption is that the minimum wage change in New Jersey doesn't affect employment in Pennsylvania. Is this a safe assumption?



General Thoughts

- Another implicit assumption is that the minimum wage change in New Jersey doesn't affect employment in Pennsylvania. Is this a safe assumption?
- Maybe not. To the extent that Card and Krueger's point is true that New Jersey is near and closely related market to eastern Pennsylvania, some workers might choose to work in NJ rather than PA. And the minimum wage may affect labor supply at a given price in Pennsylvania as well.



Introduction

“Mortality Risk and Human Capital Investment: The Impact of HIV/AIDS in Sub-Saharan Africa”

Jane G. Fortson

Review of Economics and Statistics (2011)



Introduction

What is the general topic that Fortson is interested in?



Introduction

What is the general topic that Fortson is interested in?

- Forston is ultimately interested in the effect of HIV/AIDs on economic growth, but here looks at the possible channel of schooling.

What is the specific research question addressed in this analysis?



Introduction

What is the general topic that Fortson is interested in?

- Fortson is ultimately interested in the effect of HIV/AIDS on economic growth, but here looks at the possible channel of schooling.

What is the specific research question addressed in this analysis?

- How does HIV prevalence affect the household choice to invest in schooling?



Introduction ctd

What is the reason Fortson proposes in her model for why HIV will affect the level of educational investment?



Introduction ctd

What is the reason Fortson proposes in her model for why HIV will affect the level of educational investment?

- Fortson notices that by reducing expected longevity, HIV/AIDS should reduce the lifetime expected returns to educational investment.

What data is used for this research?



Introduction ctd

What is the reason Fortson proposes in her model for why HIV will affect the level of educational investment?

- Fortson notices that by reducing expected longevity, HIV/AIDS should reduce the lifetime expected returns to educational investment.

What data is used for this research?

- Representative cross-sections from Demographic and Health Surveys in 15 Sub-Saharan countries, capturing adults aged 15 to 49 over birth cohorts 1952 to 1991.



Source of 'Identification'

What are the basic features of Fortson's dataset that she argues allows her to estimate the effect of HIV/AIDS prevalence on educational investment?



Source of 'Identification'

What are the basic features of Fortson's dataset that she argues allows her to estimate the effect of HIV/AIDS prevalence on educational investment?

- 1 HIV/AIDS became much more prevalent after roughly 1980. Using the difference-in-difference strategy, Fortson is able to control for anything that affects educational investment across cohorts, but is fixed within countries. Moreover, differencing further allows Fortson to control for anything that varies across time, but is common across countries.
- 2 As Fortson notes, 'This approach amounts to assuming that HIV had no effect on the educational outcomes of cohorts born before 1980 and a constant effect on cohorts born in or after 1980.'

Descriptive Statistics

TABLE 1.—REGIONAL HIV PREVALENCE IN SURVEY YEAR: DETAILED SUMMARY STATISTICS

Regional HIV Prevalence	BF	CM	CI	ET	GH	GN	KE	LS
Mean	1.75	5.27	4.09	1.42	2.13	1.51	6.74	23.35
Standard deviation	1.13	2.67	1.31	0.97	0.90	0.53	4.14	3.67
25th percentile	1.13	1.97	3.37	0.73	1.12	1.16	4.87	20.70
Median	1.66	4.70	3.73	1.41	2.26	1.63	5.26	23.20
75th percentile	2.26	7.97	5.46	1.74	2.67	1.76	5.99	25.47
Observations	14	12	11	11	10	8	8	10
Regional HIV Prevalence	MW	ML	NI	RW	SN	TZ	ZM	Total
Mean	11.64	1.72	0.69	2.98	0.71	6.85	15.33	4.94
Standard deviation	6.57	0.55	0.41	1.46	0.57	3.26	4.94	4.31
25th percentile	6.44	1.55	0.45	2.17	0.40	4.88	11.21	1.74
Median	8.04	1.74	0.52	2.75	0.64	6.51	15.36	3.72
75th percentile	17.56	2.13	1.04	3.31	0.70	7.34	19.93	6.52
Observations	3	9	8	12	11	21	9	157

Results are from the DHS for Burkina Faso (2003, BF), Cameroon (2004, CM), Côte d'Ivoire (2005, CI), Ethiopia (2005, ET), Ghana (2003, GH), Guinea (2005, GN), Kenya (2003, KE), Lesotho (2004, LS), Malawi (2004, MW), Mali (2001, ML), Niger (2006, NI), Rwanda (2005, RW), Senegal (2005, SN), Tanzania (2003, TZ), and Zambia (2001/2002, ZM). The table shows detailed summary statistics for the regional HIV rate in the survey year, expressed in percentage points. The unit of observation is a region. These HIV rates are calculated from the DHS HIV data using a sample that includes men and women aged 15 to 49, weighted using appropriate HIV sample weights. In calculating summary statistics for a given country, region observations are weighted by the sum of the household sample weights. In calculating summary statistics overall, these weights are adjusted by population.



TABLE 2.—SAMPLE CHILD CHARACTERISTICS: SUMMARY STATISTICS

	<i>N</i>	Mean	Standard Deviation
Regional HIV prevalence in survey year	165,660	4.178	4.658
Years behind grade for age	163,616	2.712	2.065
Year of birth	165,660	1994	2.603
Female	165,645	0.490	0.500
Rural	165,660	0.789	0.408
Burkina Faso	165,660	0.048	0.214
Cameroon	165,660	0.057	0.231
Côte d'Ivoire	165,660	0.023	0.151
Ethiopia	165,660	0.270	0.444
Ghana	165,660	0.076	0.264
Guinea	165,660	0.035	0.184
Kenya	165,660	0.117	0.322
Lesotho	165,660	0.006	0.079
Malawi	165,660	0.046	0.210
Mali	165,660	0.040	0.197
Niger	165,660	0.046	0.210
Rwanda	165,660	0.032	0.175
Senegal	165,660	0.038	0.191
Tanzania	165,660	0.127	0.333
Zambia	165,660	0.038	0.192



TABLE 3.—SAMPLE ADULT CHARACTERISTICS: SUMMARY STATISTICS

	<i>N</i>	Mean	Standard Deviation
Regional HIV prevalence in survey year	304,084	4.479	4.658
Years of schooling	302,495	4.426	4.382
Years of schooling > 0	302,495	0.617	0.486
Completed primary school	302,747	0.385	0.487
Year of birth	304,084	1976	9.639
Female	304,082	0.525	0.499
Rural	304,084	0.697	0.459
Burkina Faso	304,084	0.044	0.205
Cameroon	304,084	0.060	0.237
Côte d'Ivoire	304,084	0.078	0.268
Ethiopia	304,084	0.237	0.425
Ghana	304,084	0.074	0.262
Guinea	304,084	0.029	0.168
Kenya	304,084	0.124	0.330
Lesotho	304,084	0.008	0.087
Malawi	304,084	0.041	0.199
Mali	304,084	0.035	0.183
Niger	304,084	0.034	0.181
Rwanda	304,084	0.033	0.178
Senegal	304,084	0.041	0.198
Tanzania	304,084	0.125	0.331
Zambia	304,084	0.037	0.189



$$S_{icr} = \beta_0 + \beta_1 HIV_r \times I(c \geq 1980) + \beta_2 F_{icr} + \beta_3 rural_{icr} + \gamma_c + \alpha_r + \varepsilon_{icr}, \quad (4)$$

where S_{icr} is an educational outcome for respondent i in cohort c and region r , HIV_r is HIV prevalence in region r (survey year), $I(c \geq 1980)$ is an indicator for whether cohort c was born in or after 1980, F_{icr} is an indicator for whether respondent i in cohort c in region r is female, $rural_{icr}$ is an indicator for whether respondent i in cohort c in region r is in a rural area, γ_c is a fixed effect for birth cohort c , and α_r is a fixed effect for region r . γ_c allows for a flexible trend in educational outcomes over time, and α_r accounts for underlying differences in educational outcomes across regions of residence. (Note that the level effects of current HIV rates will be absorbed by the region indicators.)



Main Regression

FIGURE 2.—DIFFERENCE-IN-DIFFERENCES SCATTER PLOT

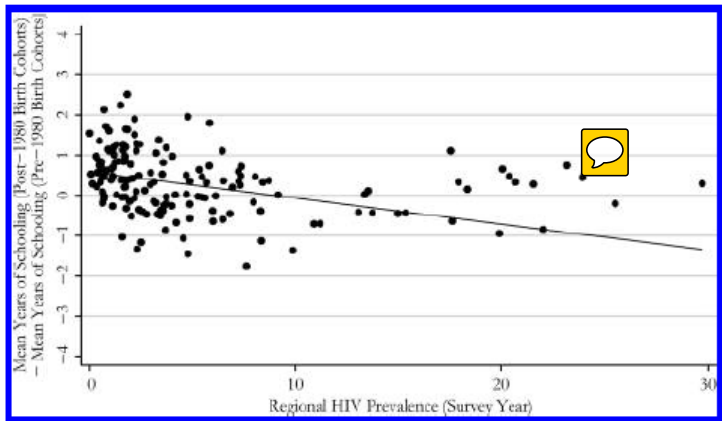




TABLE 4.—DIFFERENCE-IN-DIFFERENCES REGRESSION: POOLED

	(1) Years	(2) Years > 0	(3) Primary
Regional HIV prevalence × Post-1980 Cohort	-0.053* (0.021)	-0.006* (0.002)	-0.008* (0.002)
Rural	-2.665* (0.261)	-0.194* (0.025)	-0.270* (0.026)
Female	-1.285* (0.079)	-0.142* (0.018)	-0.112* (0.007)
Additional controls	Region FEs, birth year FEs		
Sample	Ages 15–49, most recent wave		
Observations	302,494	302,494	302,745

Robustness Checks

TABLE 5.—ROBUSTNESS CHECK: ALTERNATE TIME PATHS



Years of Schooling	(1) Exclusion 	(2) Oster 	(3) UNAIDS
Regional HIV Prevalence × Post-1985 Cohort	-0.084* (0.030)		
Childhood regional HIV prevalence (Oster Imputation)		-0.099* (0.046)	
Childhood regional HIV prevalence (UNAIDS Imputation)			-0.067* (0.023)
Rural	-2.571* (0.247)	-3.051* (0.350)	-2.738* (0.306)
Female	-1.328* (0.078)	-0.976* (0.124)	-0.982* (0.107)
Additional controls Sample	Region FEs, birth year FEs Ages 15–49 and (year of birth ≤ 1970 or year of birth ≥ 1985), most recent wave	Region FEs, birth year FEs Ages 15–49, most recent wave	Region FEs, birth year FEs Ages 15–49, most recent wave
Observations	163,917	130,716	125,654



TABLE 6.—ROBUSTNESS CHECK: DIFFERENCES PRIOR TO AFFECTED TIME PERIOD

	(1) Years	(2) Years > 0	(3) Primary
Regional HIV prevalence × Post-1970 Cohort	0.021 (0.012)	-0.002 (0.001)	0.001 (0.001)
Rural	-2.636* (0.249)	-0.186* (0.023)	-0.240* (0.023)
Female	-1.584* (0.080)	-0.157* (0.018)	-0.141* (0.007)
Additional controls	Region FEs, birth year FEs		
Sample	Ages 15–49 and year of birth < 1980, most recent wave		
Observations	169,897	169,897	170,095



TABLE 7.—SENSITIVITY ANALYSIS: MIGRATION AND MORTALITY

	(1) Years	(2) Years > 0	(3) Primary
Regional HIV Prevalence × Post-1980 Cohort	−0.032* (0.016)	−0.001 (0.002)	−0.008* (0.002)
Rural	−2.578* (0.303)	−0.193* (0.037)	−0.298* (0.035)
Female	−0.879* (0.112)	−0.117* (0.021)	−0.075* (0.011)
Additional controls	Region FEs, birth year FEs, wave FEs		
Sample	Ages 15–25, multiple waves		
Observations	234,812	234,812	234,908

TABLE 8.—CHANNELS: ORPHANHOOD

Years Behind Grade-for-Age	(1) Full Sample	(2) No Orphans
Regional HIV Prevalence × Post-1992 Cohort	0.070* (0.012)	0.075* (0.013)
Rural	0.916* (0.071)	0.929* (0.074)
Female	0.033 (0.034)	0.035 (0.033)
Additional controls	Region FEs, age FEs, birth year FEs	Region FEs, age FEs, birth year FEs
Sample	Ages 7–14, most recent wave	Nonorphans, ages 7–14, most recent wave
Observations	161,250	136,550



TABLE 9.—CHANNELS: DIFFERENCES BY SEX

	(1) Years	(2) Years > 0	(3) Primary
Regional HIV Prevalence × Post-1980 Cohort × Male	-0.103* (0.015)	-0.010* (0.002)	-0.011* (0.002)
Regional HIV Prevalence × Post-1980 Cohort	-0.003 (0.017)	-0.001 (0.002)	-0.003 (0.001)
Rural	-2.661* (0.262)	-0.193* (0.025)	-0.270* (0.026)
Female	-1.479* (0.078)	-0.161* (0.019)	-0.133* (0.007)
Additional controls	Region FEs, birth year FEs		
Sample	Ages 15–49, most recent wave		
Observations	302,494	302,494	302,745