Webinar Series in Applied Quantitative Analysis

Date	Topic
	Session One
February 29	Potential Outcomes and Omitted Variable Bias I (Theory)
March 7	Potential Outcomes and Omitted Variable Bias II (Application)
	Session Two
March 21	Difference-in-differences I (Theory)
March 28	Difference-in-differences II (Application)
	Session Three
April 25	Power analysis, clustering and sample size calculations I (Theory)
May 2	Power analysis, clustering and sample size calculations II (Application)
	Session Four
May 23	Fixed-effects I (Theory)
May 30	Fixed-effects II (Application)
	Session Five
June 20	Instrumental variables I (Theory)
June 27	Instrumental Variables II (Application)
	Session Six
July 25	Lagged dependent variables and the Arellano-Bond Estimator I (Theory)
August 1	Lagged dependent variables and the Arellano-Bond Estimator II (Application)

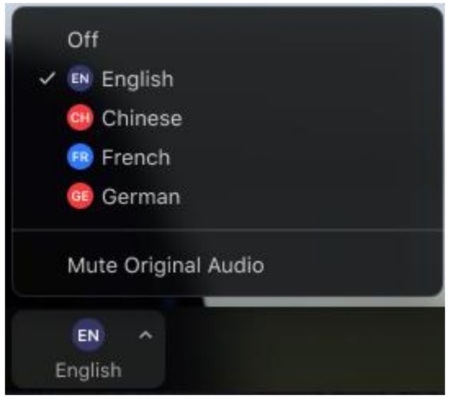
AIR®
Advancing Evidence.
Improving Lives.

https://www.airp3-africa.org/applied-quantitative-analysis-webinar-series

AERC
CREA

Listening to language interpretation
Windows | macOS

1. In your meeting/webinar controls, click Interpretation .
2. Click the language that you would like to hear. (We will have French) No need to choose English, that is the language in the main Zoom room



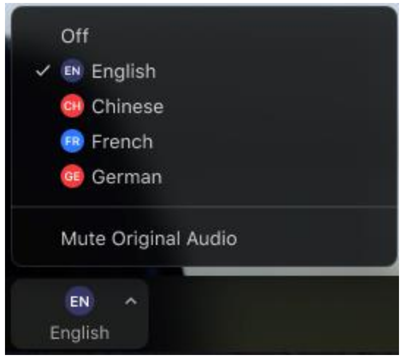
3. (Optional) To hear the interpreted language only, click Mute Original Audio.

Notes:

- You must join the meeting audio through your computer audio/VoIP. You cannot listen to language interpretation if you use the dial-in or call me phone audio features.
- As a participant joining a language channel, you can broadcast back into the main audio channel if you unmute your audio and speak.

Écoute de l'interprétation d'une langue
Windows | macOS

1. Dans les contrôles de votre réunion/webinaire, cliquez sur Interprétation .
2. Cliquez sur la langue que vous souhaitez entendre. (Nous aurons le français) Pas besoin de choisir l'anglais, c'est la langue de la salle Zoom principale



3. (Facultatif) Pour entendre uniquement la langue interprétée, cliquez sur Couper le son original.

Remarques:

- Vous devez rejoindre l’audio de la réunion via l’audio/VoIP de votre ordinateur. Vous ne pouvez pas écouter l’interprétation linguistique si vous utilisez les fonctions audio de connexion ou d’appel téléphonique.
- En tant que participant rejoignant une chaîne linguistique, vous pouvez retransmettre sur le canal audio principal canal si vous réactivez votre audio et parlez.

Difference-in-Differences II

Ashu Handa
Institute Fellow – AIR
Kenan Eminent Professor of Public Policy – UNC-CH
March 28, 2024

[Some slides are courtesy of Prof. Gustavo Angeles]

Topics for today

- Review the 2x2 DiD model and answer lingering questions
- Extend the model to additional time periods
- <u>Brief</u> intro to DiD with staggered treatment (advanced topic)
- STATA demonstrations (data and scripts are on the website)

$$Y = B_0 + B_1*(Treat) + B_2*(Post) + B_3*(Treat*Post) + e_i$$

Treat=1 if treated group, 0 if control group
Post=1 if post-intervention period (after), 0 if pre-intervention

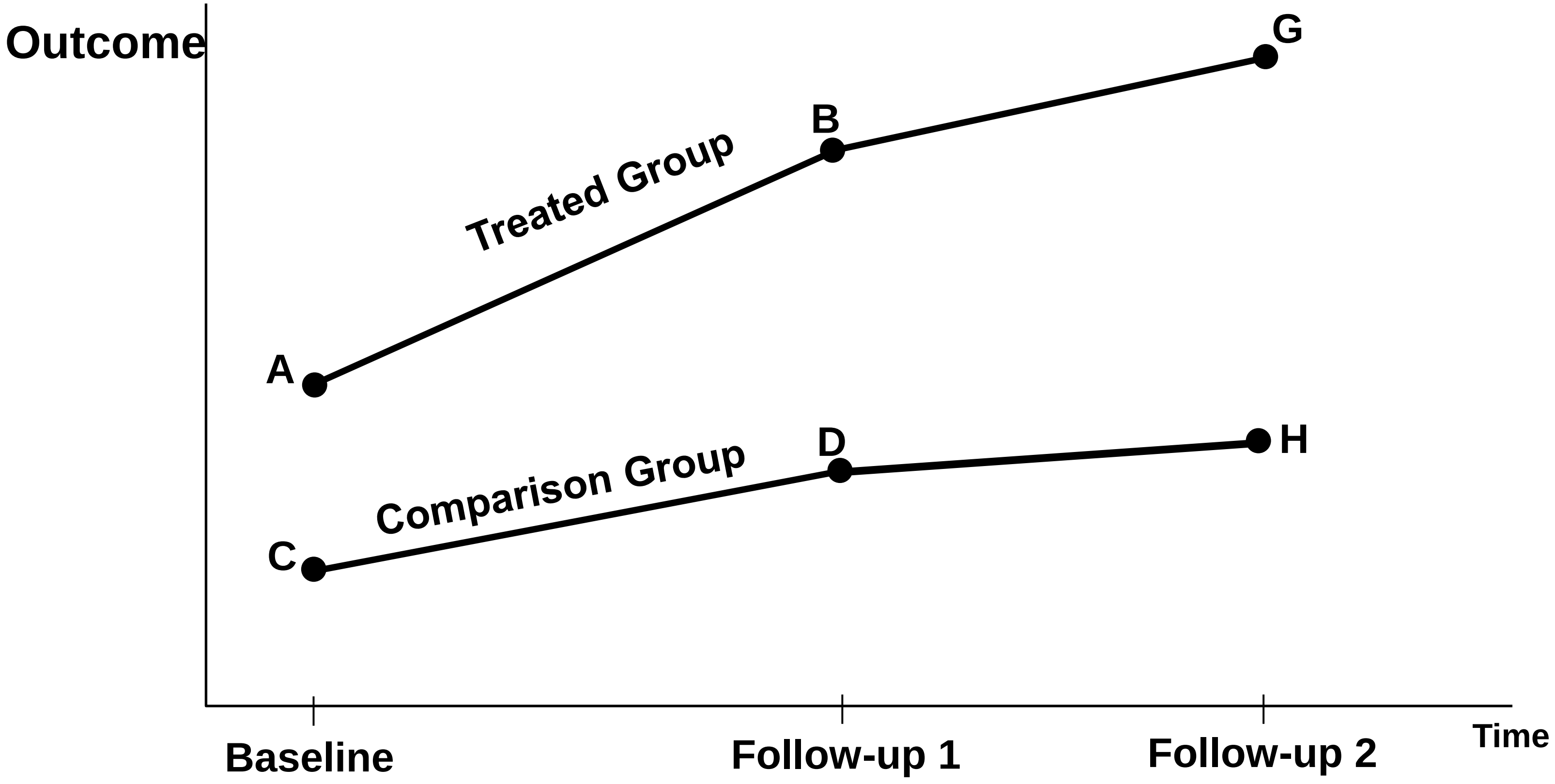
	Before (Post=0)	After (Post=1)	Difference
Comparison (Treat=0)	Y = B0	Y = B0 + B2	B2
Treated (Treat=1)	Y = B0 + B1	Y = B0 + B1 + B2 + B3	B2+B3
Difference	B1	B1 + B3	B3

DID Impact

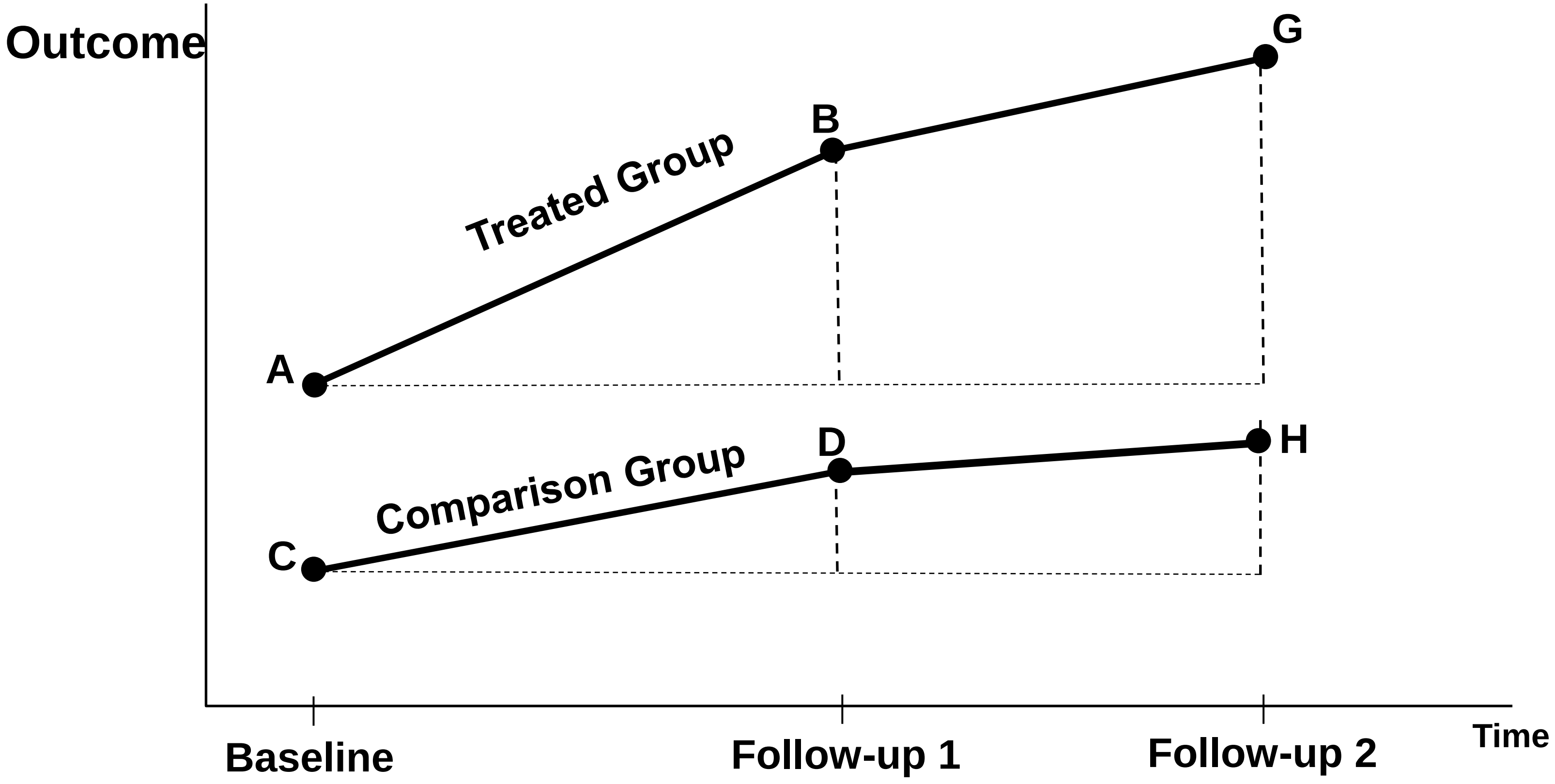
Coefficient interpretation

1. B_2 is the time trend
2. B_1 gives the difference between the two groups before the intervention
3. B_0 is the mean of Y in the comparison group before the intervention
4. B_3 is the DiD estimate

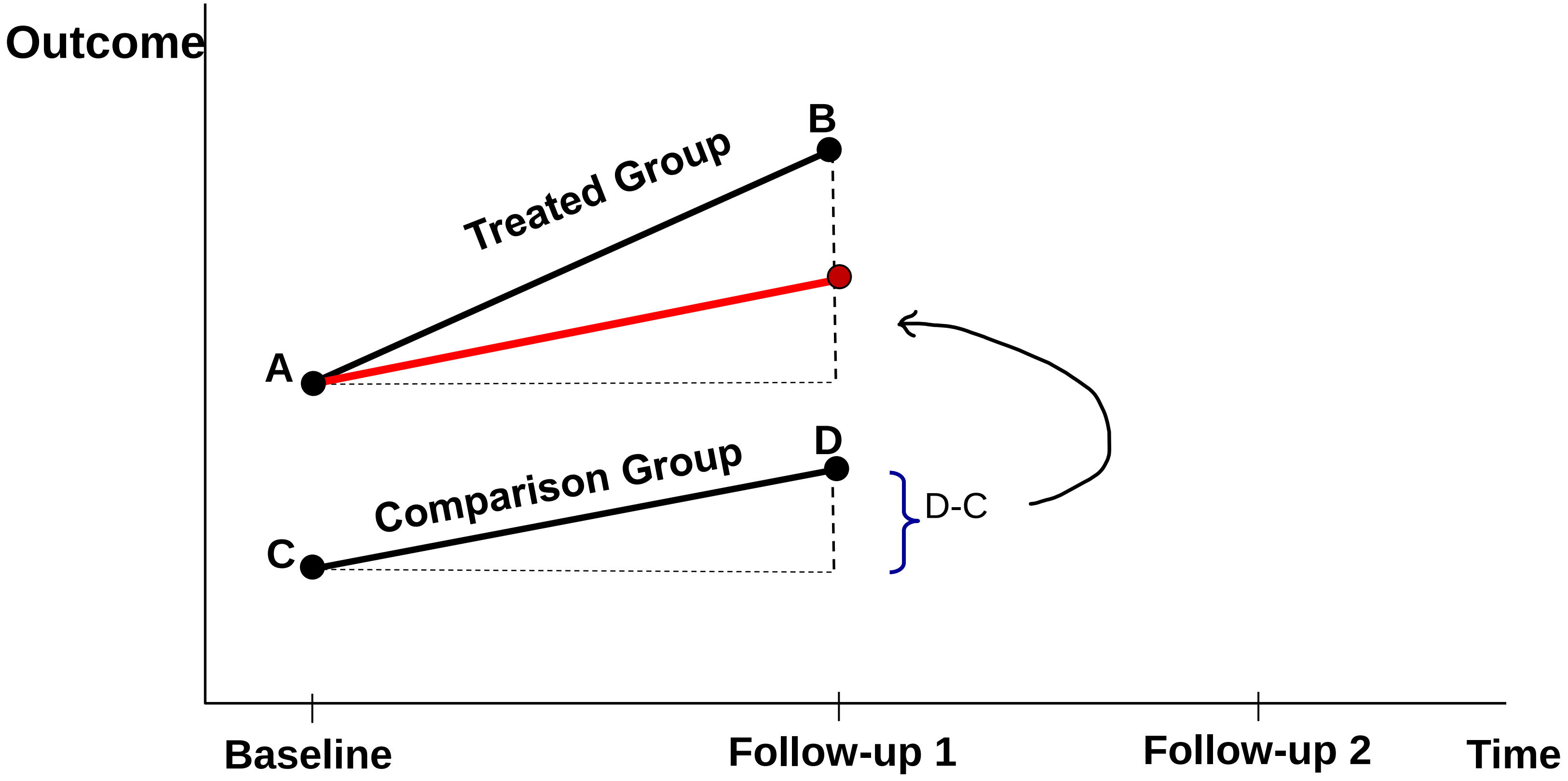
DID Extensions: Multiple times – Case: 3 points in time



DiD Extensions: Multiple times – Case: 3 points in time



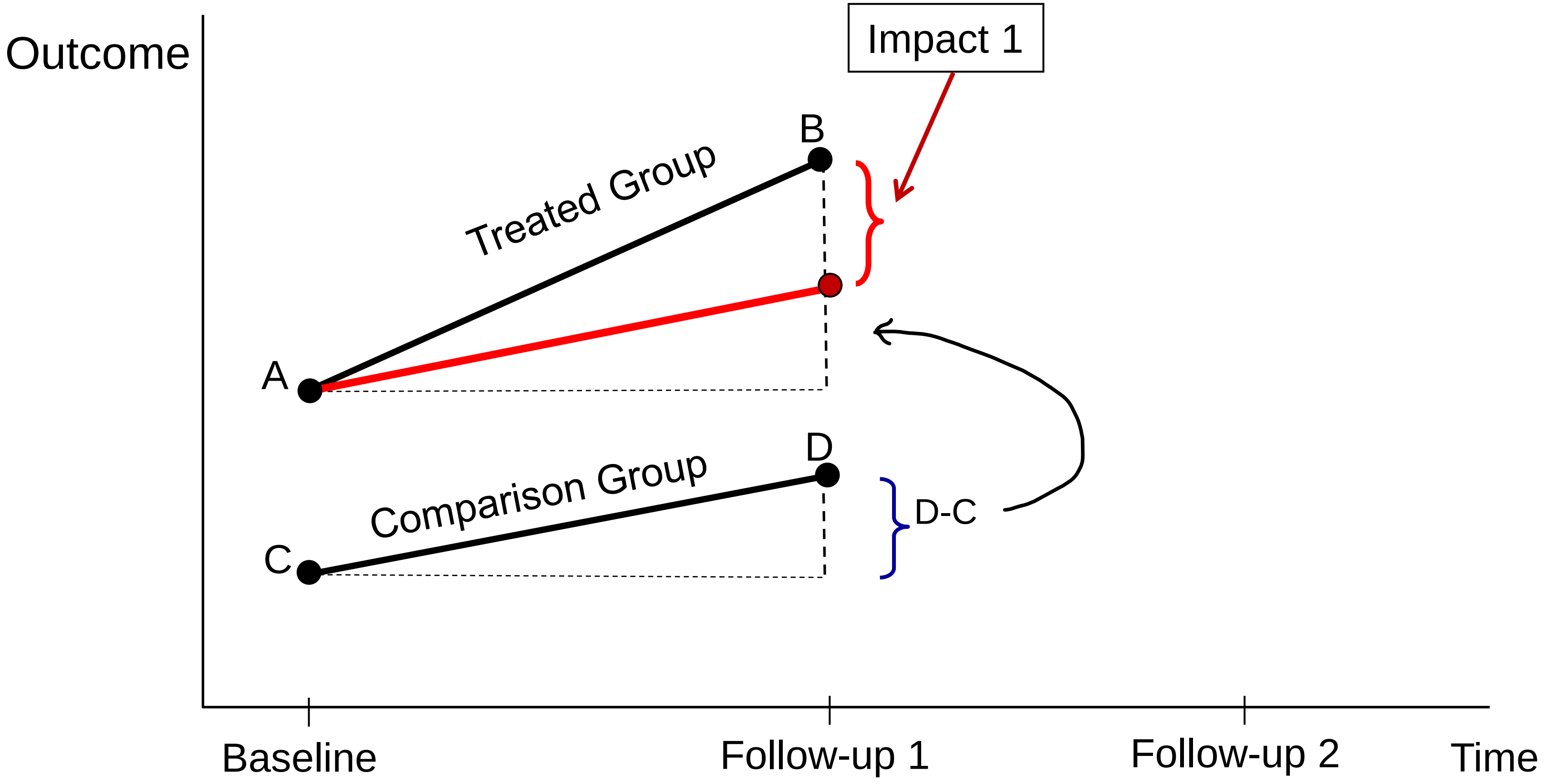
DiD Extensions - 3 points in time



<u>Key assumption 1:</u> Parallel trends assumption holds for each time period.

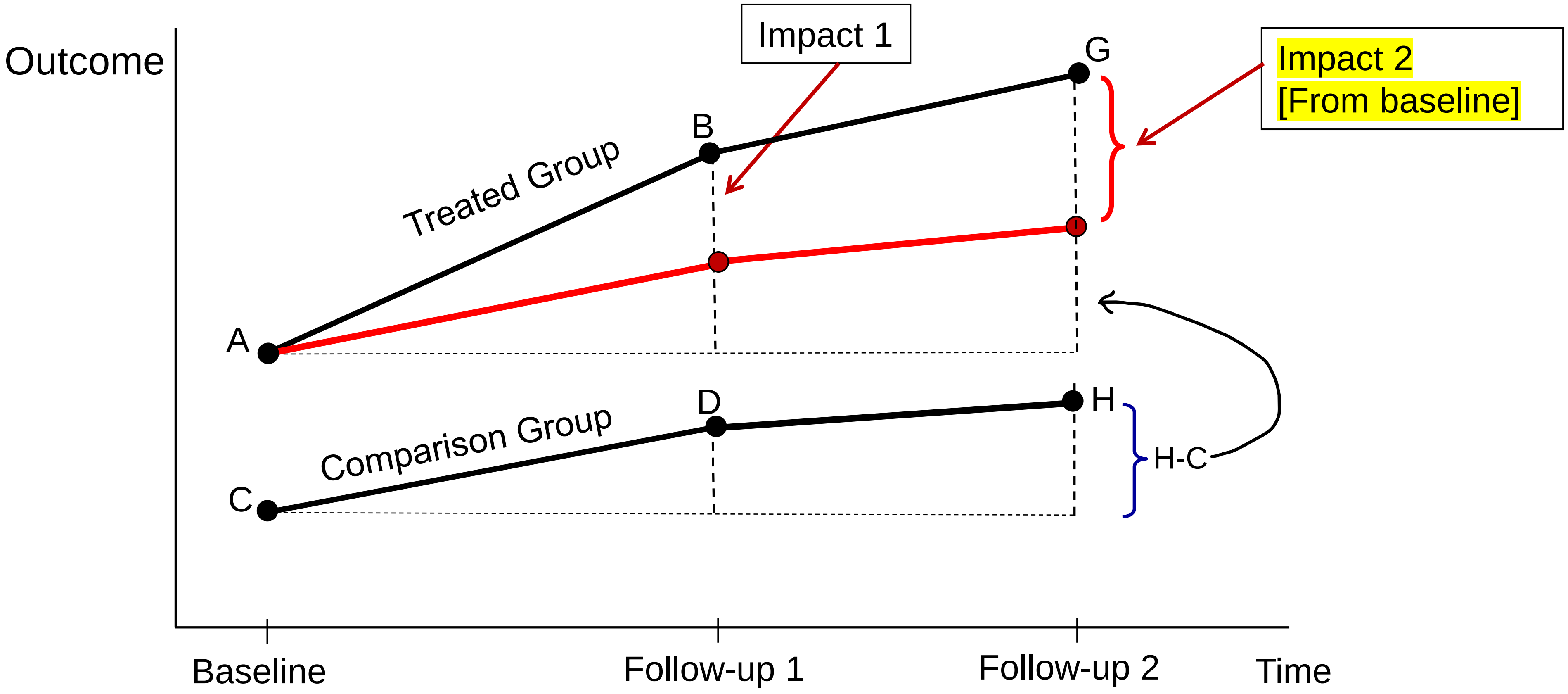
<u>Key assumption 2</u>: No anticipation effects

DiD Extensions - 3 points in time



<u>Key assumption 1:</u> Parallel trends assumption holds for each time period.

DiD Extensions - 3 points in time



Key assumption 1: Parallel trends assumption holds for each time period.

DID Extensions - 3 points in time: The empirical model

The model:

$$Y_{ijt} = \beta_0 + \beta_1 Treat_j + \beta_2 T1_t + \beta_3 T2_t + \beta_4 Treat_j * T1_t + \beta_5 Treat_j * T2_t + \beta_6 X_{ijt} + \varepsilon_{ijt}$$

Where,`

Y_{ijt} : $Outcome\ of\ interest$

$$Treat_j = \begin{cases} 1, if\ community\ j\ is\ in\ the\ treated\ area \\ 0, if\ community\ j\ is\ in\ the\ comparison\ area \end{cases}$$

$$T1_t = \begin{cases} 1, if\ observation\ is\ from\ Follow-up\ 1\ survey \\ 0, if\ observation\ is\ from\ other\ time \end{cases}$$

$$T2_t = \begin{cases} 1, if\ observation\ is\ from\ Follow-up\ 2\ survey \\ 0, if\ observation\ is\ from\ other\ time \end{cases}$$

X_{ijt}: $Individual\ characteristics$

ε_{ijt}: $Random\ error$

DID Extensions - 3 points in time: The empirical model

The model:

$$Y_{ijt} = \beta_0 + \boxed{\beta_1 Treat_j} + \beta_2 T1_t + \beta_3 T2_t + \beta_4 Treat_j * T1_t + \beta_5 Treat_j * T2_t + \beta_6 X_{ijt} + \varepsilon_{ijt}$$

Where,

Y_{ijt} : *Outcome of interest*

$$\boxed{Treat_j} = \begin{cases} 1, if\ community\ j\ is\ in\ the\ treated\ area \\ 0, if\ community\ j\ is\ in\ the\ comparison\ area \end{cases}$$

$$T1_t = \begin{cases} 1, if\ observation\ is\ from\ Follow-up\ 1\ survey \\ 0, if\ observation\ is\ from\ other\ time \end{cases}$$

$$T2_t = \begin{cases} 1, if\ observation\ is\ from\ Follow-up\ 2\ survey \\ 0, if\ observation\ is\ from\ other\ time \end{cases}$$

X_{ijt}: *Individual characteristics*

ε_{ijt}: *Random error*

DID Extensions - 3 points in time: The empirical model

The model:

$$Y_{ijt} = \beta_0 + \boxed{\beta_1 Treat_j} + \beta_2 T1_t + \beta_3 T2_t + \beta_4 Treat_j * T1_t + \beta_5 Treat_j * T2_t + \beta_6 X_{ijt} + \varepsilon_{ijt}$$

Where,`

Y_{ijt} : $Outcome\ of\ interest$

$$\boxed{Treat_j} = \begin{cases} 1, if\ community\ j\ is\ in\ the\ treated\ area \\ 0, if\ community\ j\ is\ in\ the\ comparison\ area \end{cases}$$

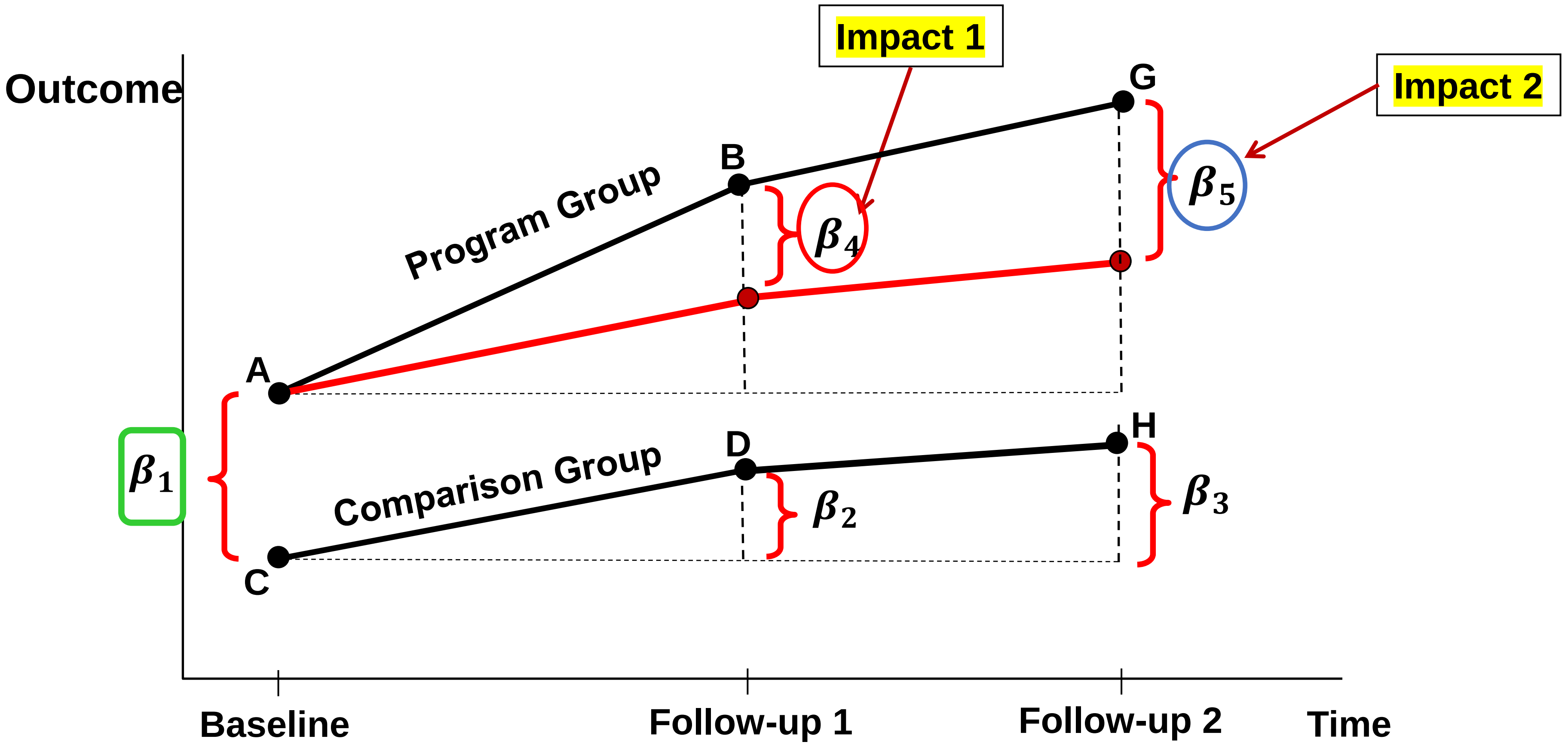
$$T1_t = \begin{cases} 1, if\ observation\ is\ from\ Follow-up\ 1\ survey \\ 0, if\ observation\ is\ from\ other\ time \end{cases}$$

$$T2_t = \begin{cases} 1, if\ observation\ is\ from\ Follow-up\ 2\ survey \\ 0, if\ observation\ is\ from\ other\ time \end{cases}$$

X_{ijt}: $Individual\ characteristics$

ε_{ijt}: $Random\ error$

$$Y_{ijt} = \beta_0 + \boxed{\beta_1} Treat_j + \beta_2 T1_t + \beta_3 T2_t + \beta_4 Treat_j * T1_t + \beta_5 Treat_j * T2_t + \beta_6 X_{ijt} + \varepsilon_{ijt}$$

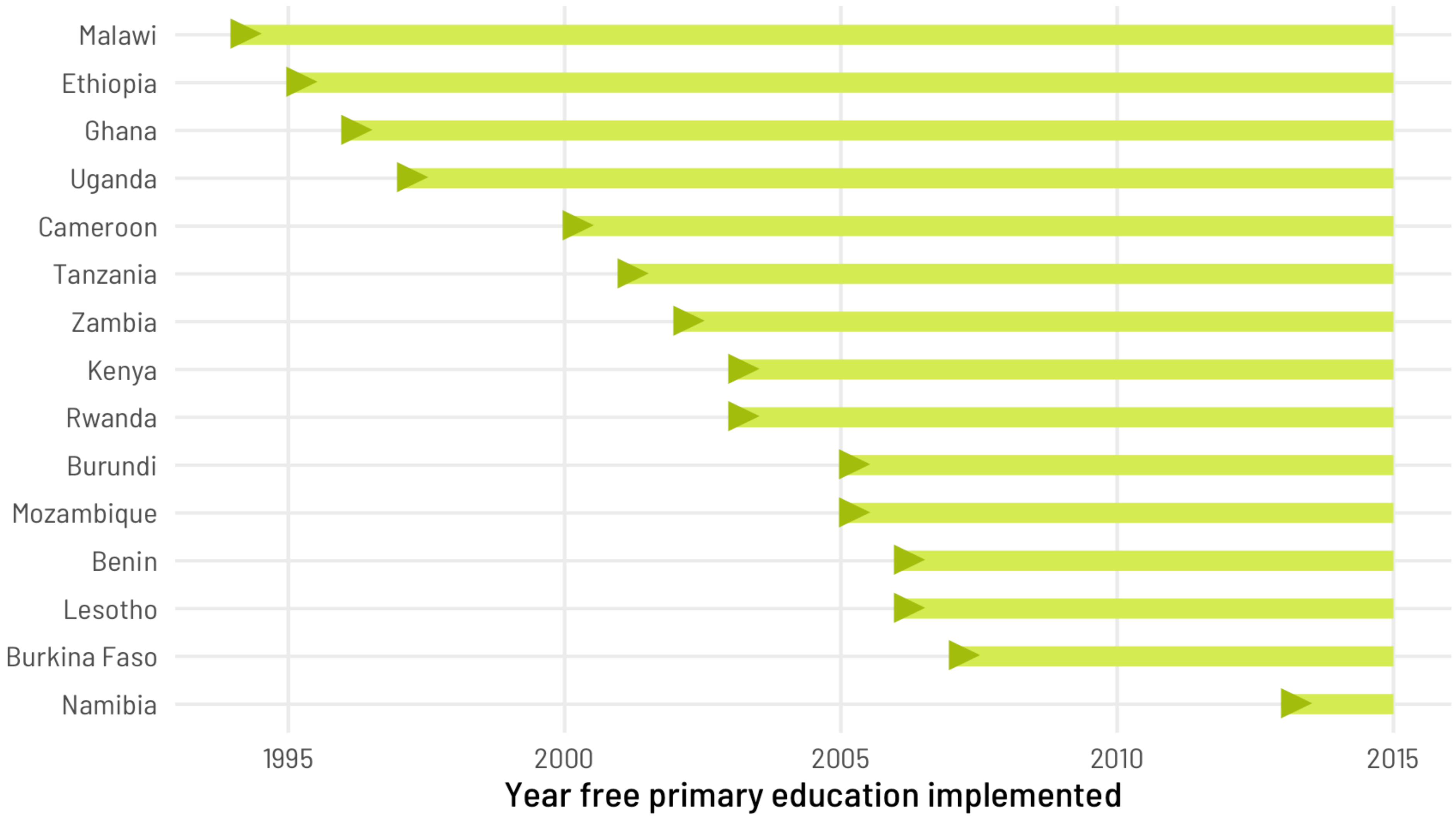


<u>Assumption 1:</u> Parallel trends assumption holds for each time period.

New developments in DiD that address deviations from the standard 2x2 model (brief overview)

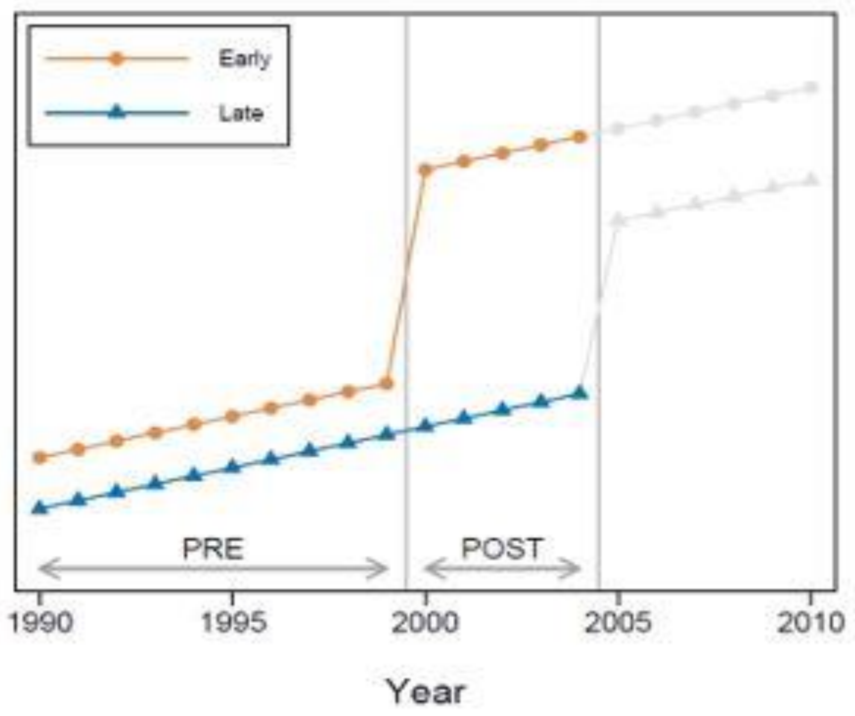
1. Multiple post-intervention periods with staggered treatment;
2. Multiple post-intervention periods with heterogeneous treatment effects over time;
3. Treatment heterogeneity across units (where units refer to the level of treatment—clusters, countries, districts);
4. A combination of #2 and #3.

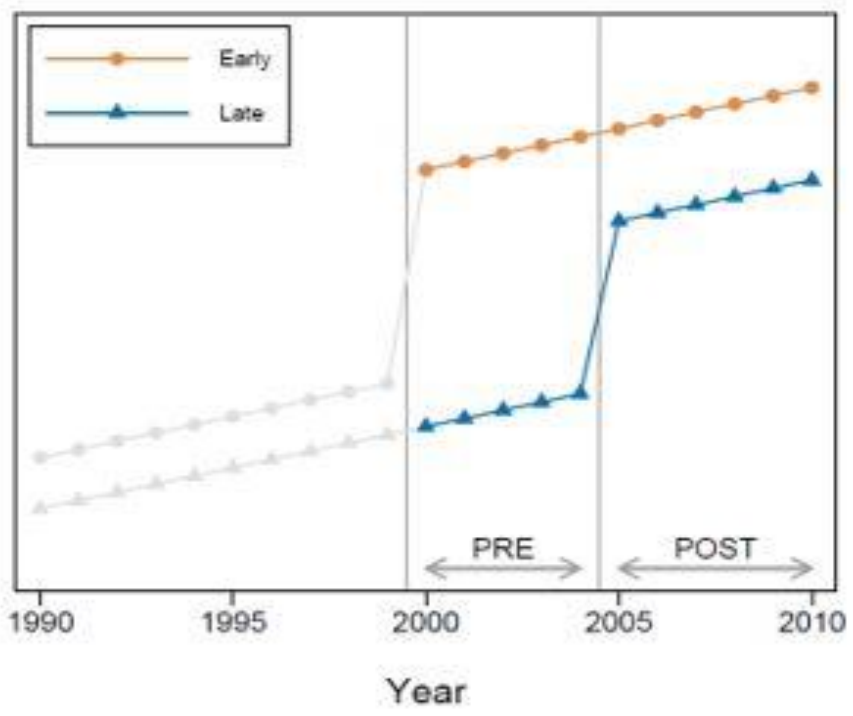
Effect of free primary on schooling: staggered treatment



Intuition: Effect of free primary education on schooling

- **Two types of countries:** one group that made primary education free in 2000 (early), and another group that made it free in 2005 (late)
- **Data:** Primary school completion rates every year from 1990 to 2010.
- **Two distinct DiD comparisons are possible:**
  - First, we could focus on the period from 1990 to 2004 (**the left panel in the figure below**) - late adopter countries are “never treated”– so they can be used as a comparison group to estimate the impact of free primary schooling on test scores in the early adopter countries
  - Second, we can also construct a second DD estimate of the treatment effect of free primary school by focusing on the years 2001 to 2010 (**the right panel in the figure below**). Treatment status of the early adopters never changes – they remain treated throughout – so they can be used as a comparison group to estimate the impact of free primary schooling on test scores in the late adopting countries





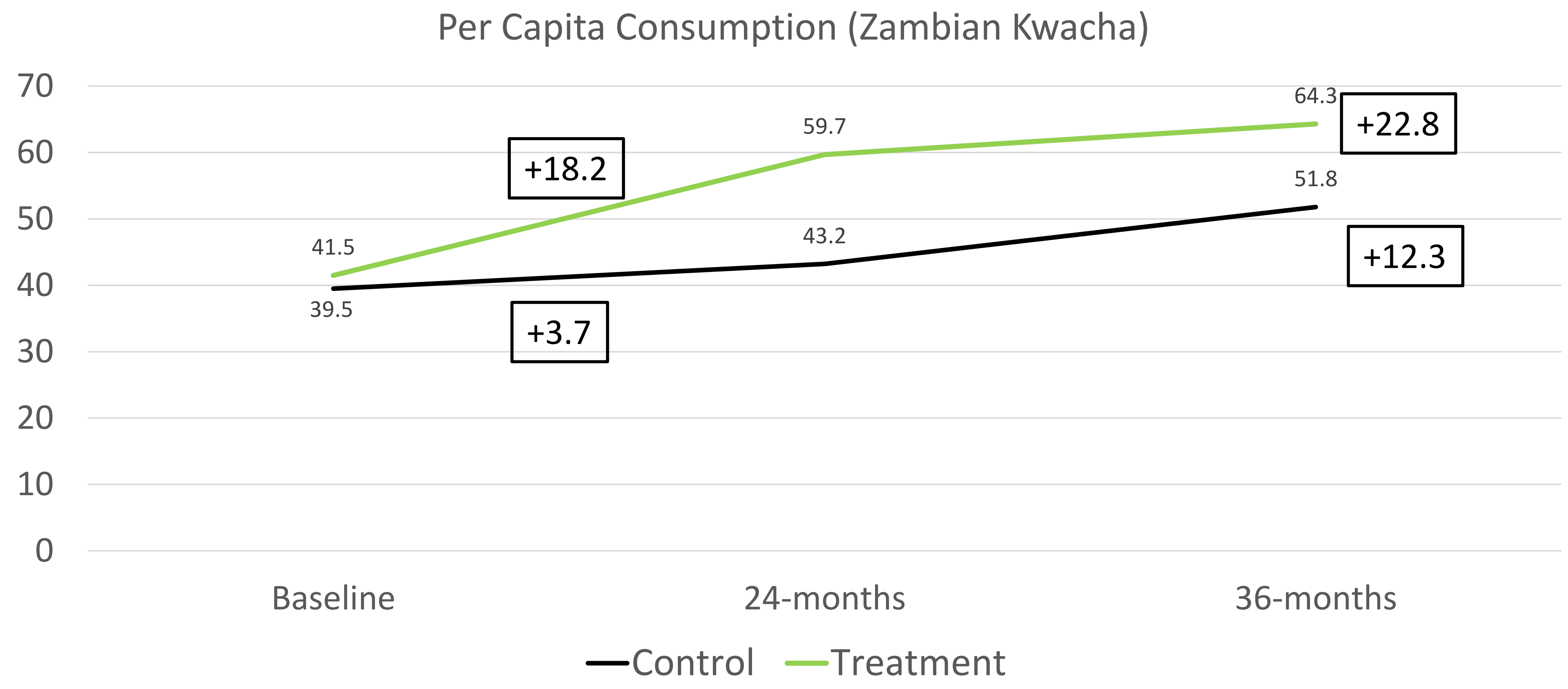
Key references if your study deviates from 2x2
[Red papers have ready-made code you can use!]

- Baker, Andrew, David F. Larcker, and Charles C. Y. Wang. 2021. “How Much Should We Trust Staggered Difference-in-Differences Estimates?” Working paper 246. Rock Center for Corporate Governance at Stanford University. https://doi.org/10.2139/ssrn.3794018.
- Callaway, Brantly, and Pedro H. C. Sant’Anna. 2020. “Difference-in-Differences with Multiple Time Periods.” *Journal of Econometrics*, December. https://doi.org/10.1016/j.jeconom.2020.12.001.
- de Chaisemartin, Clément, and Xavier d’Haultfoeuille. 2020. “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects.” *American Economic Review* 110 (9): 2964–96. https://doi.org/10.1257/aer.20181169.
- Goodman-Bacon, Andrew. 2021. “Difference-in-Differences with Variation in Treatment Timing.” *Journal of Econometrics*, June. https://doi.org/10.1016/j.jeconom.2021.03.014.
- Jakiela, Pamela. 2021. “Simple Diagnostics for Two-Way Fixed Effects,” March. https://arxiv.org/abs/2103.13229.
- Sun, Liyang, and Sarah Abraham. 2020. “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects.” *Journal of Econometrics*, December. https://doi.org/10.1016/j.jeconom.2020.09.006.
- https://www.andrewheiss.com/blog/2021/08/25/twfe-diagnostics/#a-different-way-of-thinking-about-ols-coefficients

STATA demonstrations

- Zambia Child Grant
  - 2x2 DiD
  - Three periods
- Rural-urban differences in male youth outcomes over time in Ghana
  - Repeated cross-sections using DHS
  - Data is ‘stacked’
  - ‘Treatment’ is ‘living in rural region’
  - **Hundreds of PhD dissertations available using this methodology**
- Measuring effect of climate shocks (time permitting)

Zambia Child Grant Program Impact Evaluation



<u>DiD estimates</u>
Baseline-24-months: (18.2 – 3.7) = +14.5
Baseline-36-months: (22.8 – 12.3) = +10.5 (from baseline)

$$Y_{ijt} = \beta_0 + \beta_1 Treat_j + \beta_2 T1_t + \beta_3 T2_t + \beta_4 Treat_j * T1_t + \beta_5 Treat_j * T2_t + \varepsilon_{ijt}$$

39.5 2.0 3.8 12.4 14.4 10.4

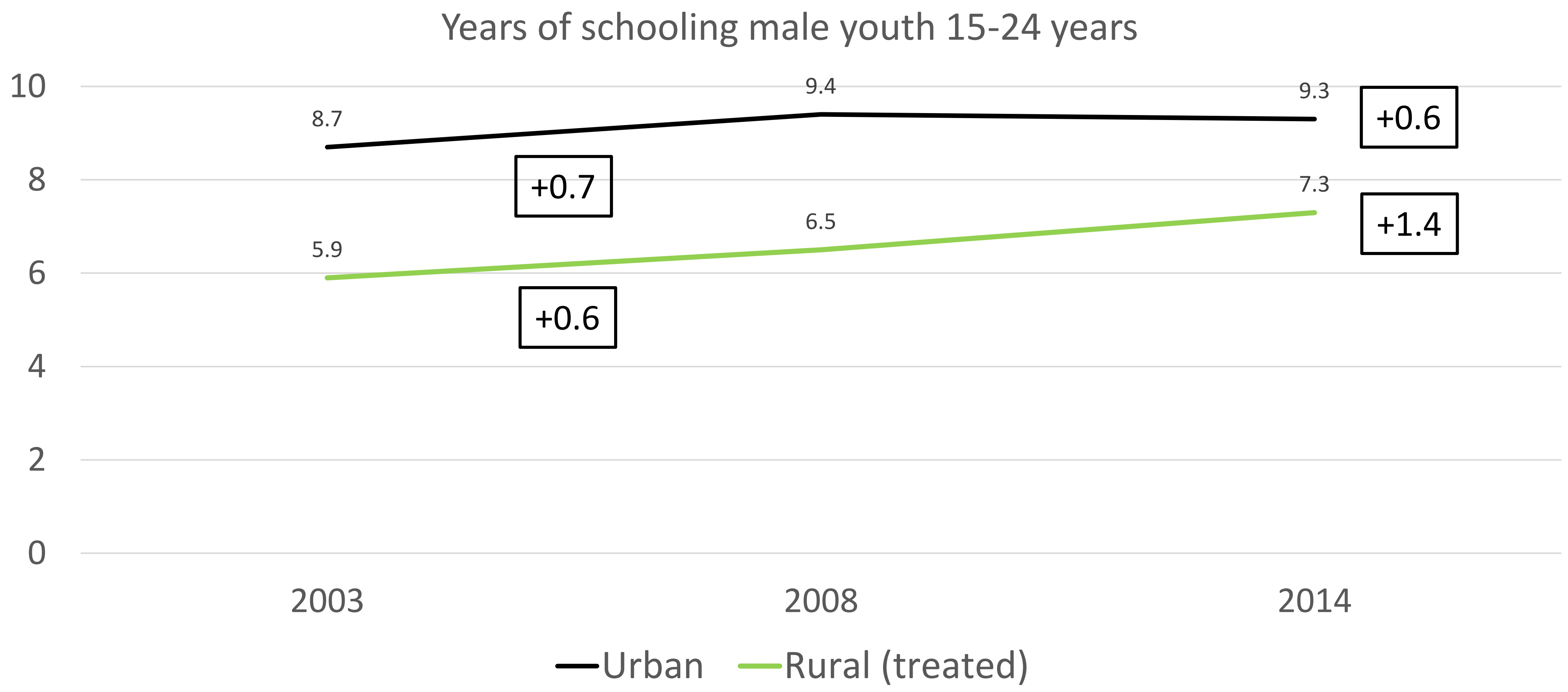
```
. reg exp_per_capita treat period24 period36 DD24 DD36
```

Source	SS	df	MS
Model	644287.314	5	128857.463
Residual	9627429.76	7,267	1324.81488
Total	10271717.1	7,272	1412.50235

Number of obs	=	7,273
F(5, 7267)	=	97.26
Prob > F	=	0.0000
R-squared	=	0.0627
Adj R-squared	=	0.0621
Root MSE	=	36.398

exp_per_ca~a	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
treat	2.019155	1.451573	1.39	0.164	-.8263499	4.86466
period24	3.789231	1.486378	2.55	0.011	.8754988	6.702963
period36	12.35714	1.45655	8.48	0.000	9.501876	15.2124
DD24	14.40795	2.100744	6.86	0.000	10.28988	18.52602
DD36	10.44709	2.064304	5.06	0.000	6.400455	14.49373
_cons	39.47611	1.025805	38.48	0.000	37.46524	41.48699

Rural-urban differences in youth over time in Ghana



<u>DiD estimates</u>
2003-2008: (0.6 - + 0.7) = -0.1
2003-2014: (1.4 - +0.6) = 0.8

$$Y_{ijt} = \beta_0 + \beta_1 Treat_j + \beta_2 T1_t + \beta_3 T2_t + \beta_4 Treat_j * T1_t + \beta_5 Treat_j * T2_t + \varepsilon_{ijt}$$

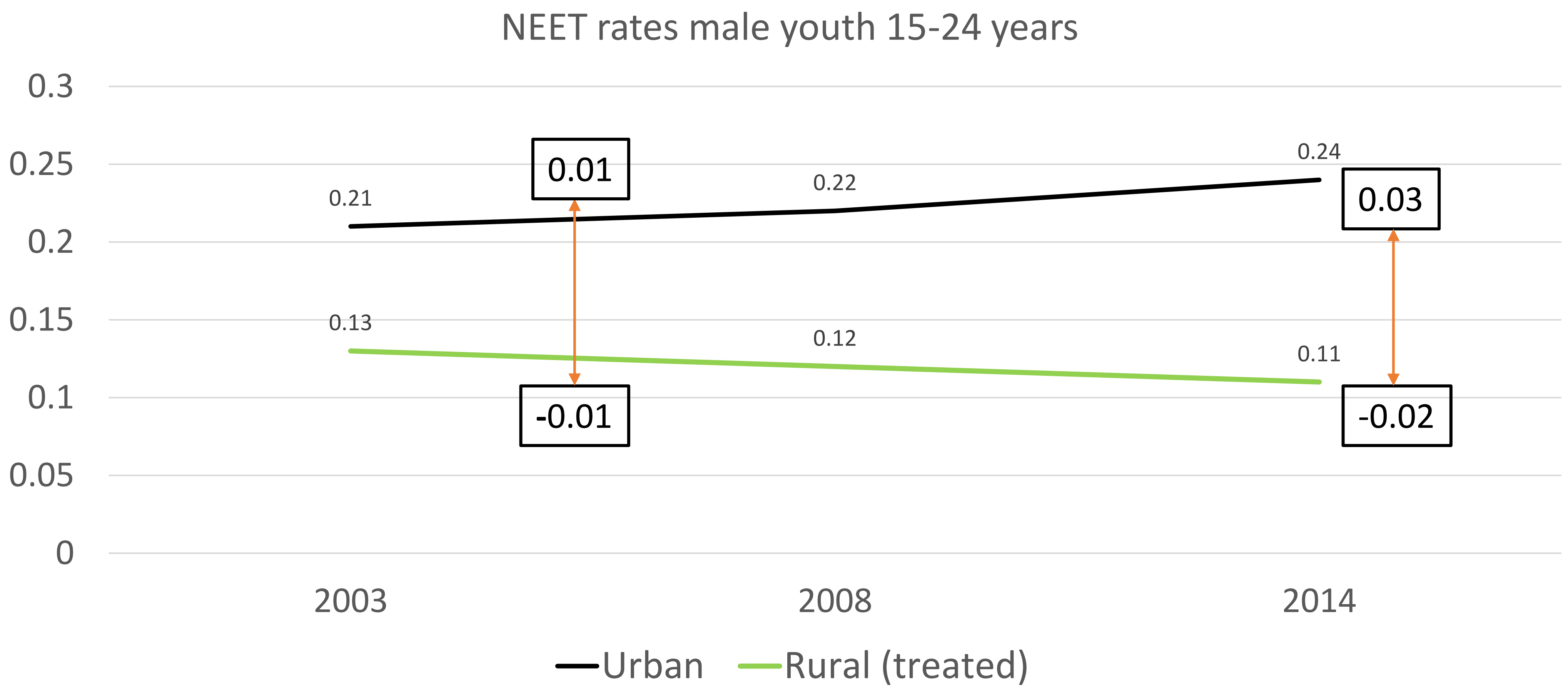
8.6 -2.8 0.69 0.58 -0.04 0.80

```
. reg yearsed year2008 year2014 rural DD2008 DD2014

      Source |       SS           df       MS      Number of obs   =     4,940
-------------+----------------------------------   F(5, 4934)      =    169.03
       Model |  9241.59087         5  1848.31817   Prob > F        =    0.0000
    Residual |  53952.6073     4,934  10.9348616   R-squared       =    0.1462
-------------+----------------------------------   Adj R-squared   =    0.1454
       Total |  63194.1982     4,939  12.7949379   Root MSE        =    3.3068

------------------------------------------------------------------------------
     yearsed | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    year2008 |   .6932637   .1752934     3.95   0.000     .3496106    1.036917
    year2014 |   .5824653   .1767417     3.30   0.001      .235973    .9289577
       rural |  -2.804055   .1592733   -17.61   0.000    -3.116301   -2.491808
      DD2008 |  -.0387703   .2293317    -0.17   0.866    -.4883624    .4108219
      DD2014 |   .8023604   .2338181     3.43   0.001      .343973    1.260748
       _cons |   8.674931   .1227265    70.69   0.000     8.434333     8.91553
------------------------------------------------------------------------------
```

Rural-urban differences in youth over time in Ghana



<u>DiD estimates</u>
2003-2008: (-0.01 - + 0.01) = -0.02
2003-2014: (-0.02 - + 0.03) = -0.05

$$Y_{ijt} = \beta_0 + \beta_1 Treat_j + \beta_2 T1_t + \beta_3 T2_t + \beta_4 Treat_j * T1_t + \beta_5 Treat_j * T2_t + \varepsilon_{ijt}$$

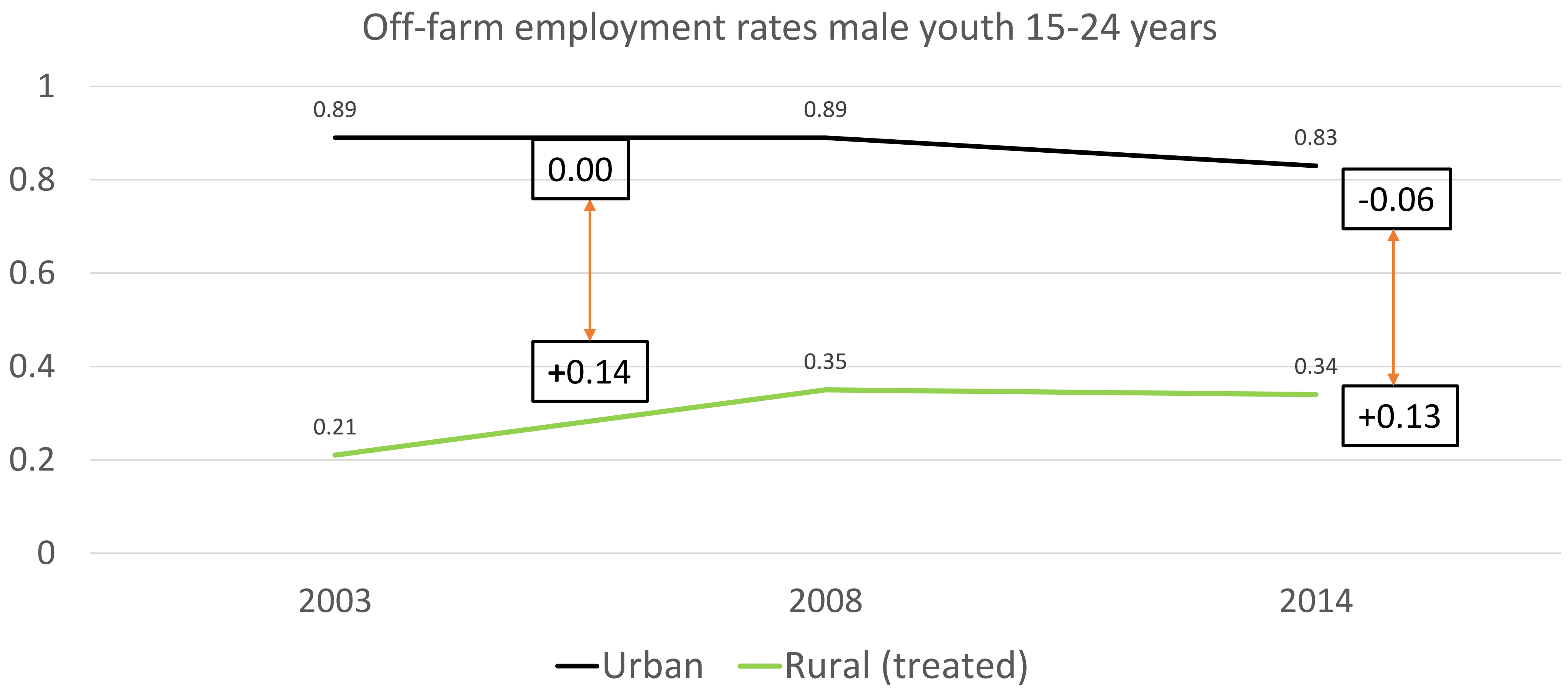
0.21 -0.08 0.02 0.03 -0.03 -0.05

```
. reg neet year2008 year2014 rural DD2008 DD2014

      Source |       SS           df       MS      Number of obs   =     4,944
-------------+----------------------------------   F(5, 4938)      =     20.46
       Model |  13.7062056         5  2.74124111   Prob > F        =    0.0000
    Residual |  661.568673     4,938  .133975025   R-squared       =    0.0203
-------------+----------------------------------   Adj R-squared   =    0.0193
       Total |  675.274879     4,943  .136612357   Root MSE        =    .36603

------------------------------------------------------------------------------
        neet | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    year2008 |   .0162456    .019389     0.84   0.402    -.0217654    .0542565
    year2014 |   .0345127   .0195634     1.76   0.078    -.0038403    .0728656
       rural |  -.0784306   .0176298    -4.45   0.000    -.1129929   -.0438683
      DD2008 |  -.0273379   .0253678    -1.08   0.281    -.0770702    .0223943
      DD2014 |    -.05345   .0258811    -2.07   0.039    -.1041885   -.0027114
       _cons |   .2066116   .0135845    15.21   0.000     .1799799    .2332432
------------------------------------------------------------------------------
```

Rural-urban differences in youth over time in Ghana



<u>DiD estimates</u>
2003-2008: (+0.14 - 0.0) = +0.14
2003-2014: (+0.13 - - 0.06) = +0.19

$$Y_{ijt} = \beta_0 + \beta_1 Treat_j + \beta_2 T1_t + \beta_3 T2_t + \beta_4 Treat_j * T1_t + \beta_5 Treat_j * T2_t + \varepsilon_{ijt}$$

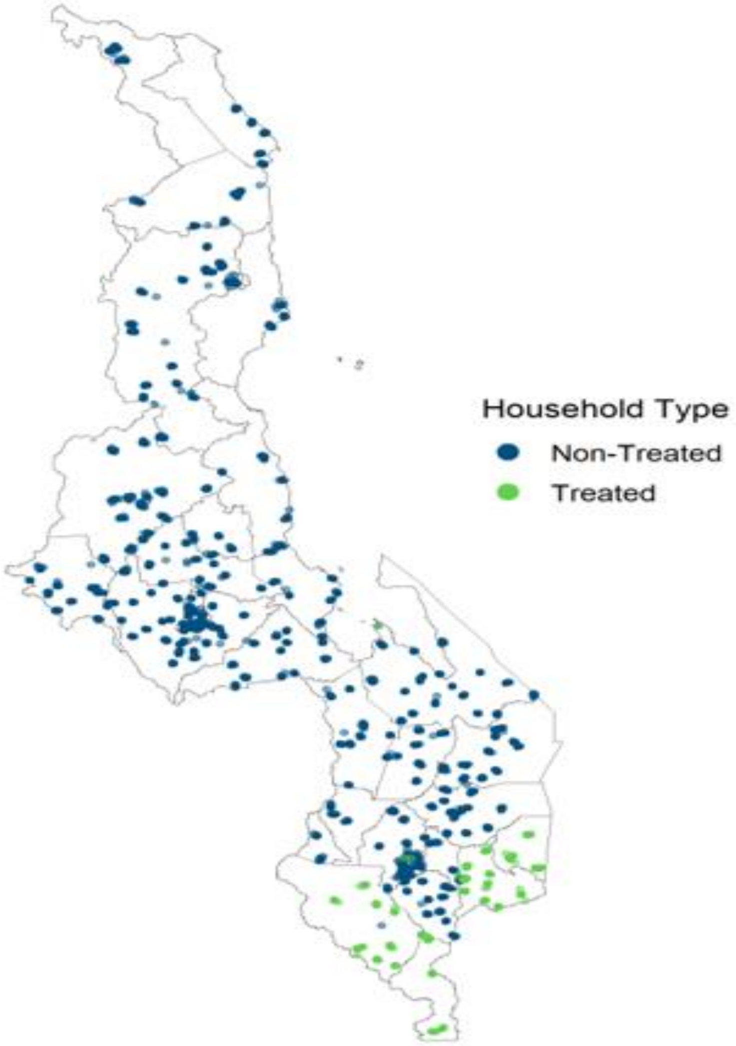
0.89 -0.68 0.01 -0.06 0.13 0.18

```
. reg offfarm year2008 year2014 rural DD2008 DD2014

      Source |       SS           df       MS      Number of obs   =     2,539
-------------+----------------------------------   F(5, 2533)      =    222.83
       Model |  193.908114         5  38.7816227   Prob > F        =    0.0000
    Residual |  440.837062     2,533  .174037529   R-squared       =    0.3055
-------------+----------------------------------   Adj R-squared   =    0.3041
       Total |  634.745175     2,538  .250096602   Root MSE        =    .41718

------------------------------------------------------------------------------
     offfarm | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    year2008 |    .005668   .0352151     0.16   0.872    -.0633854    .0747214
    year2014 |  -.0578935   .0351686    -1.65   0.100    -.1268556    .0110686
       rural |  -.6751369   .0320546   -21.06   0.000    -.7379927   -.6122811
      DD2008 |    .133893   .0432617     3.09   0.002     .0490612    .2187249
      DD2014 |   .1834484   .0434029     4.23   0.000     .0983396    .2685572
       _cons |   .8866397   .0265444    33.40   0.000     .8345887    .9386906
------------------------------------------------------------------------------
```

What was the ‘impact’ of Tropical Storms Ana and Gombe (Q1, 2022) on households in Malawi?

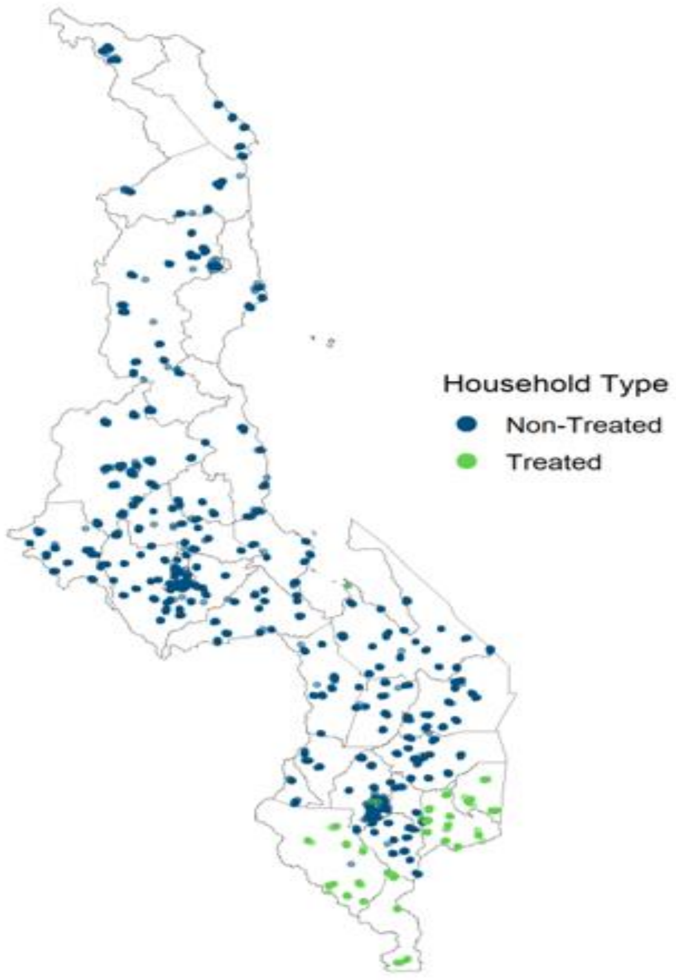


1. The ‘treatment’ is the storm locations (areas marked in green)

2. National household survey 2019 (IHS5)

3. Special national survey conducted in 2023

What was the ‘impact’ of Tropical Storms Ana and Gombe on households in Malawi?

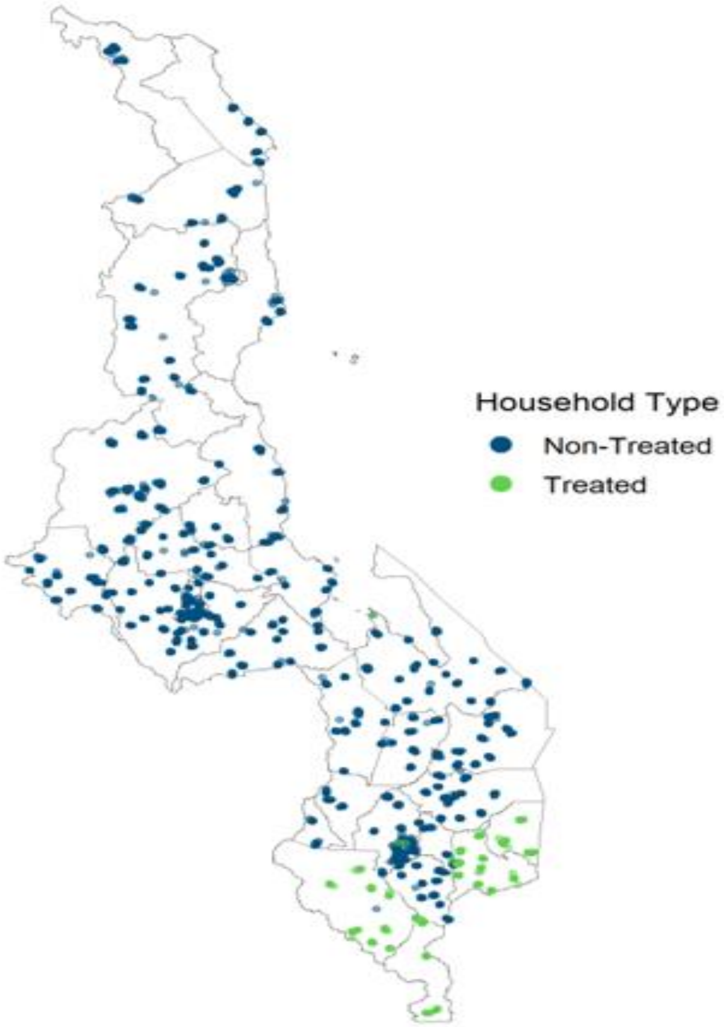


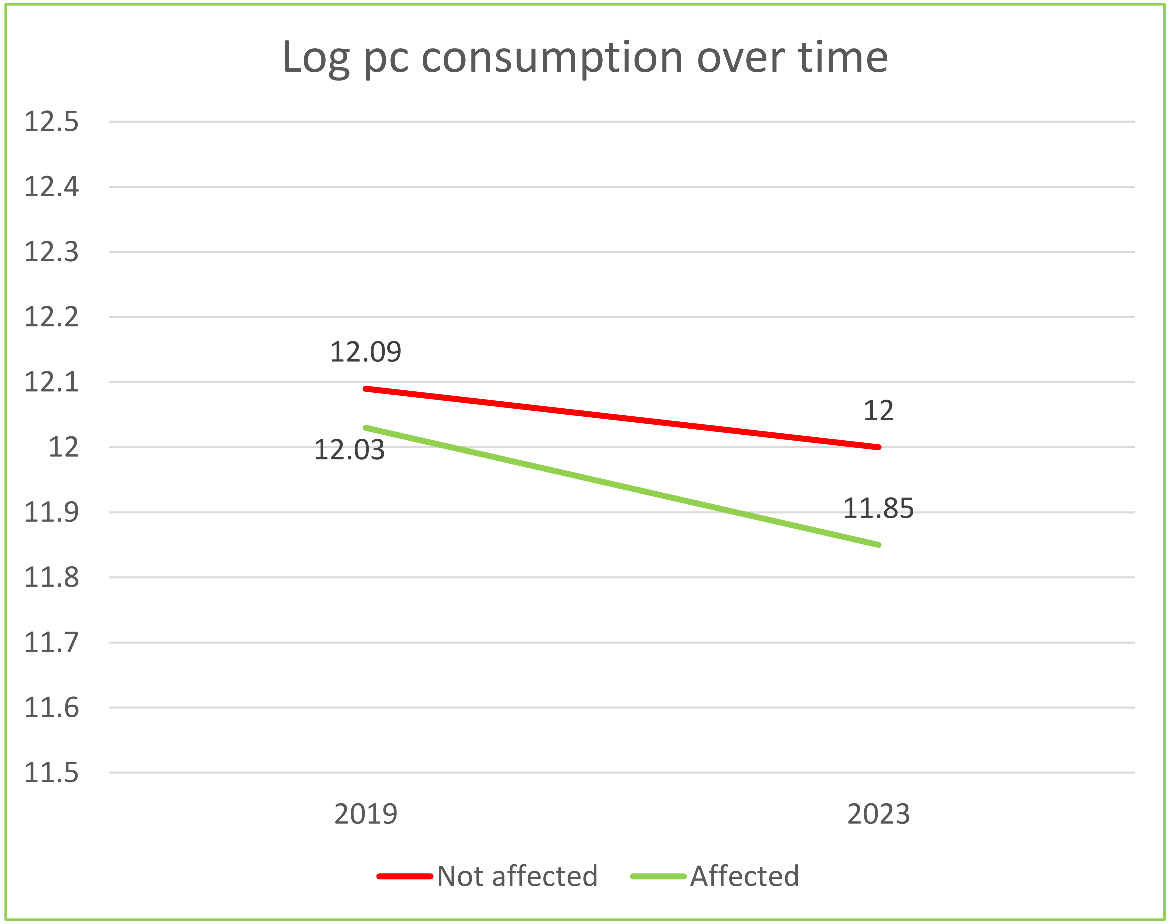
1. Identified households in most intensely hit areas

2. Found ‘similar’ households in less intensely hit areas in the South

3. Had data from 2019 (before) and after (2023)—compared trends

DiD estimate of ‘impact’ of Tropical Storms Ana and Gombe in Malawi





Affected: (12 – 12.09) = -0.09
Not affected: (11.85 – 12.03) = -0.18
DiD: (-0.18 – (-0.09)) = -0.09 (9% reduction)

DiD estimate of ‘impact’ of Tropical Storms Ana and Gombe in Malawi

```
. reg logexp time T DD if bl_rural==1

      Source |       SS           df       MS      Number of obs   =     6,170
-------------+----------------------------------   F(3, 6166)      =     17.84
       Model |  23.1011513         3  7.70038376   Prob > F        =    0.0000
    Residual |  2661.79059     6,166  .431688386   R-squared       =    0.0086
-------------+----------------------------------   Adj R-squared   =    0.0081
       Total |  2684.89174     6,169  .435223171   Root MSE        =    .65703

------------------------------------------------------------------------------
      logexp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        time |  -.0780356   .0179186    -4.35   0.000    -.1131623   -.0429088
           T |  -.0531358   .0353647    -1.50   0.133    -.1224629    .0161913
          DD |  -.1089284   .0500132    -2.18   0.029    -.2069718   -.0108851
       _cons |   12.08734   .0126704   953.98   0.000      12.0625    12.11218
------------------------------------------------------------------------------
```