Webinar Series in Applied Quantitative Analysis - Updated

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	Propensity score matching (Theory)	
May 30	Propensity score matching (Application)	
	Session Five	
June 20	Fixed-effects I (Theory)	
June 27	Fixed-effects II (Application)	
	Session Six	
July 25	Instrumental variables I (Theory)	
August 1	Instrumental Variables II (Application)	
	Session Seven	
August 22	Lagged dependent variables and the Arellano-Bond Estimator I (Theory)	
August 29	Lagged dependent variables and the Arellano-Bond Estimator II (Application)	

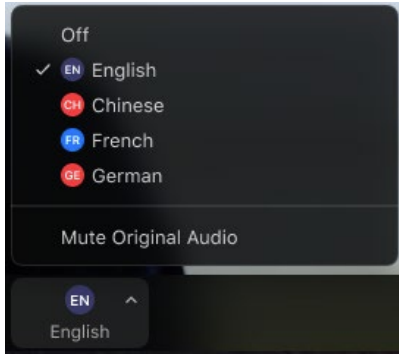
AIR®
Advancing Evidence.
Improving Lives.

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

AERC
CREA

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



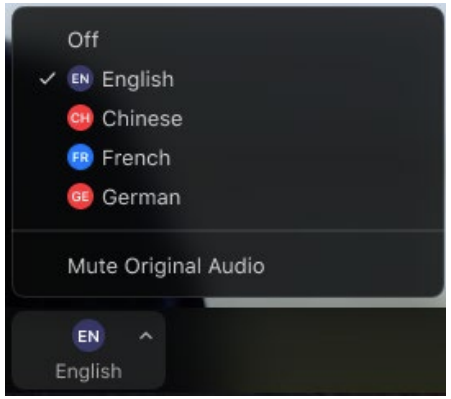
3. (Facultatif) Pour entendre uniquement la langue interprétée, cliquez sur Couper le son original.

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.

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.

Matching and Propensity Score Matching II

Ashu Handa

Institute Fellow – AIR

Kenan Eminent Professor of Public Policy – UNC-CH

May 30, 2024

AIR®

Advancing Evidence.
Improving Lives.

AERC
CREA

Finish example from last week and move to STATA: Estimating the impact of a poverty targeted program

- We have data on the program participants (P=1)
- We have a national data set (household budget survey)—can we use that to estimate the impact of this program?
- Key assumption behind matching: ‘Selection on observables’

When is the assumption behind matching valid (or more valid)?

[In each case you want to identify a comparison group from the national survey using matching, in order to conduct an impact assessment]

Case	
1	Poor, rural households with children in three districts are targeted and receive a cash transfer from the government
2	A business training program for women aged 18-28 is being offered in a few settlement areas of Nairobi for free, eligible people must apply and are selected based on their application

Poll

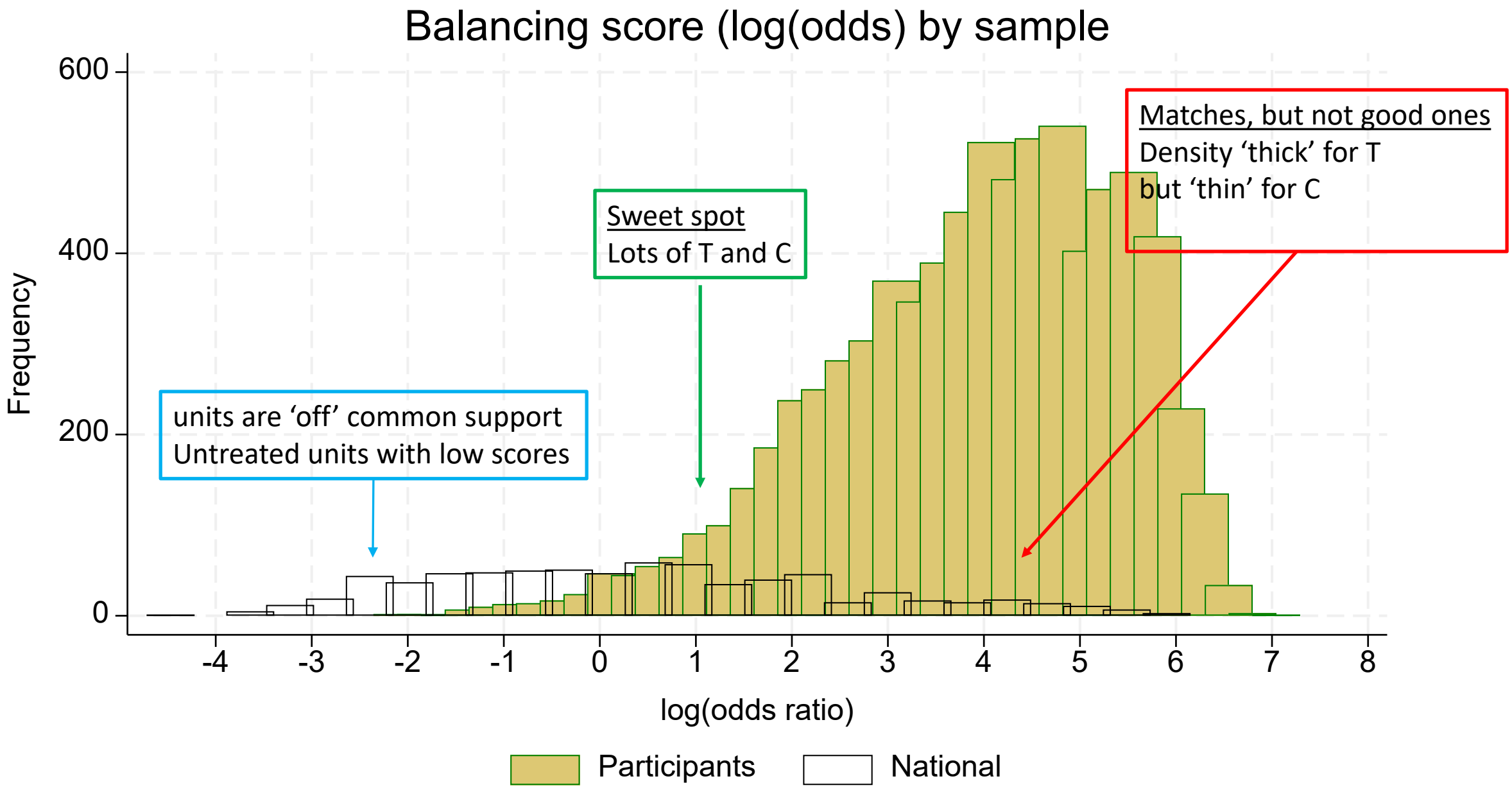
The propensity score matching recipe

1. Assemble the relevant data sets
   - Two data sets (program data + secondary data) or one

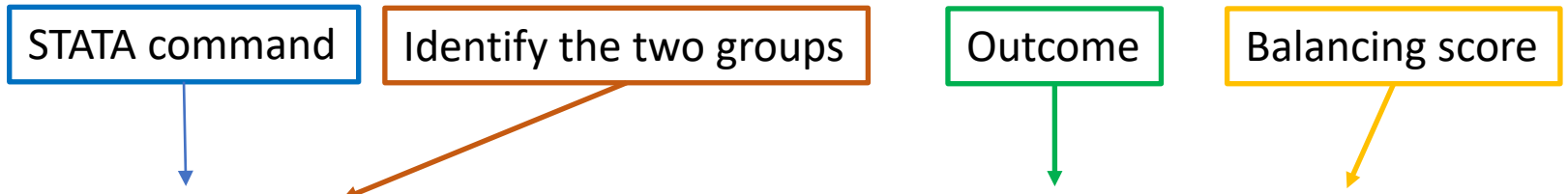
2. Estimate balancing (propensity) score
   - Logit T = f(X)

3. Assess the quality of the matches
   - Look at distribution of scores; common support
   - Conduct matching and test for balance

4. Finalize the estimates
   - Doubly robust estimator


Balancing score (log(odds) by sample
Frequency
600
400
200
0
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
log(odds ratio)
Matches, but not good ones
Density 'thick' for T
but 'thin' for C
Sweet spot
Lots of T and C
units are 'off' common support
Untreated units with low scores
Participants
National

Match on the generated log(odds ratio) using psmatch2: Nearest neighbor w/ replacement

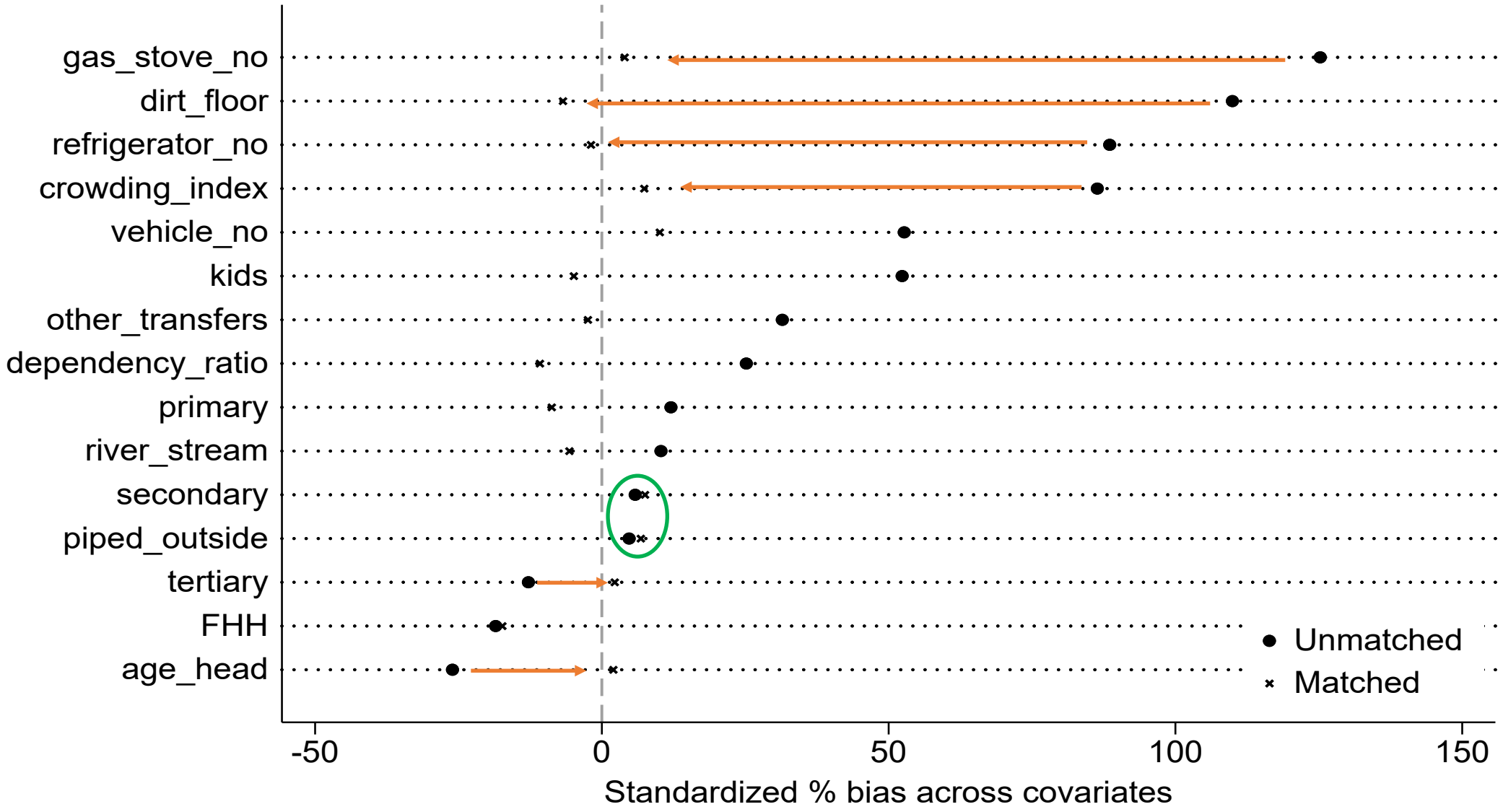



```
. psmatch2 D if comsup==1 & type~=2, out(foodexp) pscore(lo)
```

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
foodexp	Unmatched	511.96147	848.814674	-336.853204	17.3779495	-19.38
	ATT	511.96147	701.981711	-190.02024	70.4417245	-2.70

408/736 untreated units used as matches

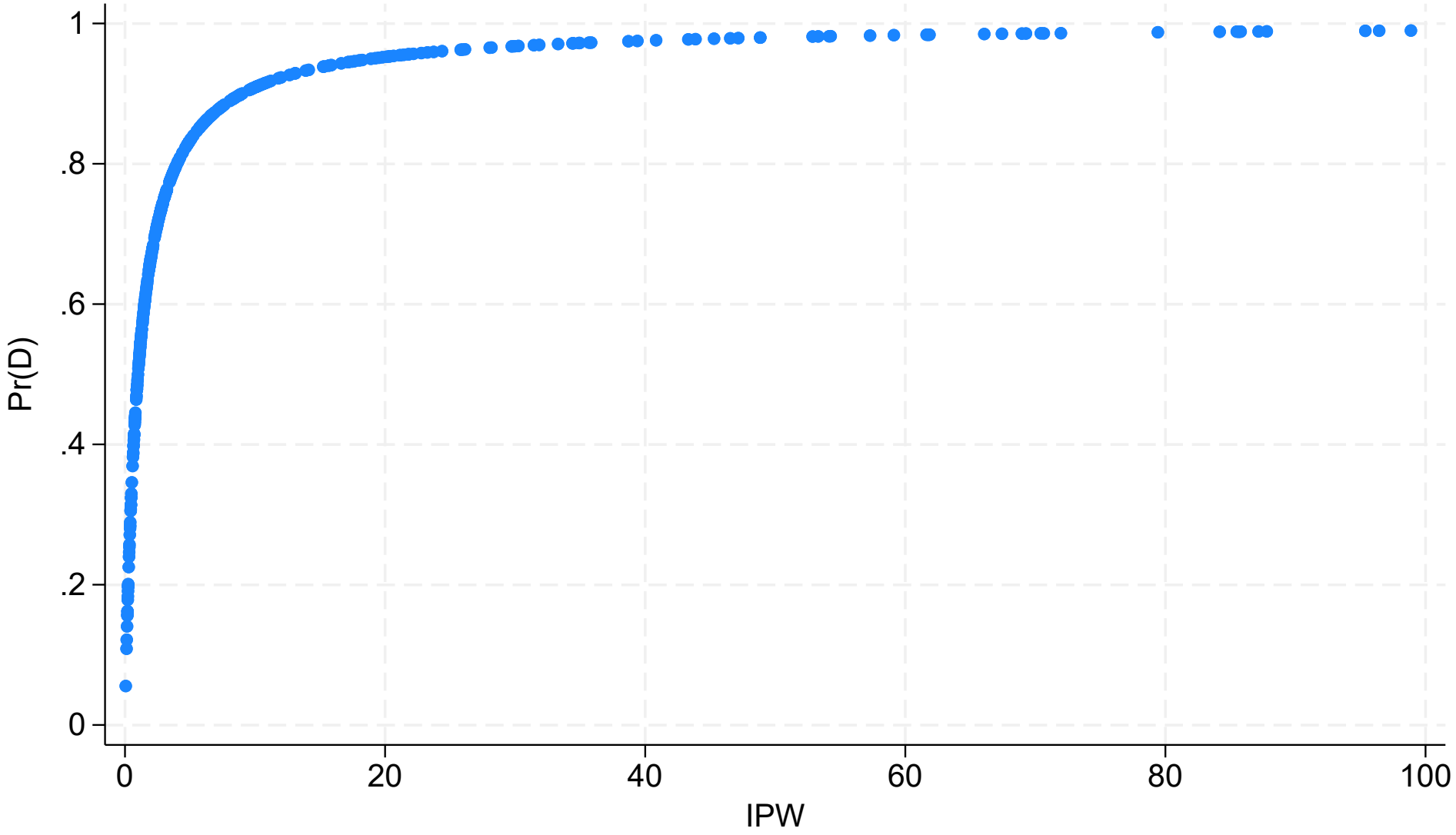
Reduction in imbalance with matched sample



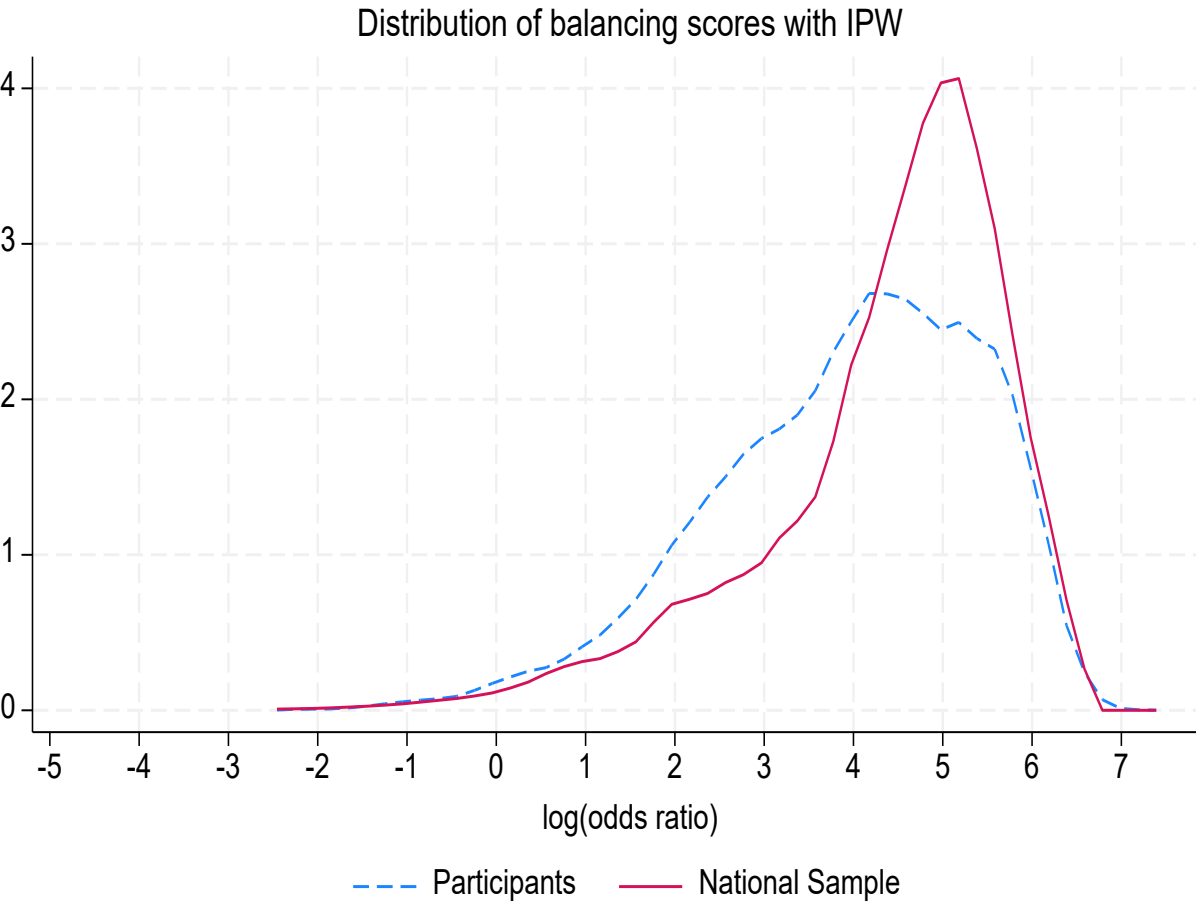
Doubly robust estimator

- Comparison of means between matched samples has been replaced with a ‘doubly robust’ estimator which performs better
- Identify matched sample using standard procedure
- Then estimate treatment effect via regression on the matched sample
  - Add covariates to absorb remaining imbalance
  - Use inverse probability weights (IPW) in the regression
  - **IPW = (ps/(1-ps)) for untreated units, 1 for treated**

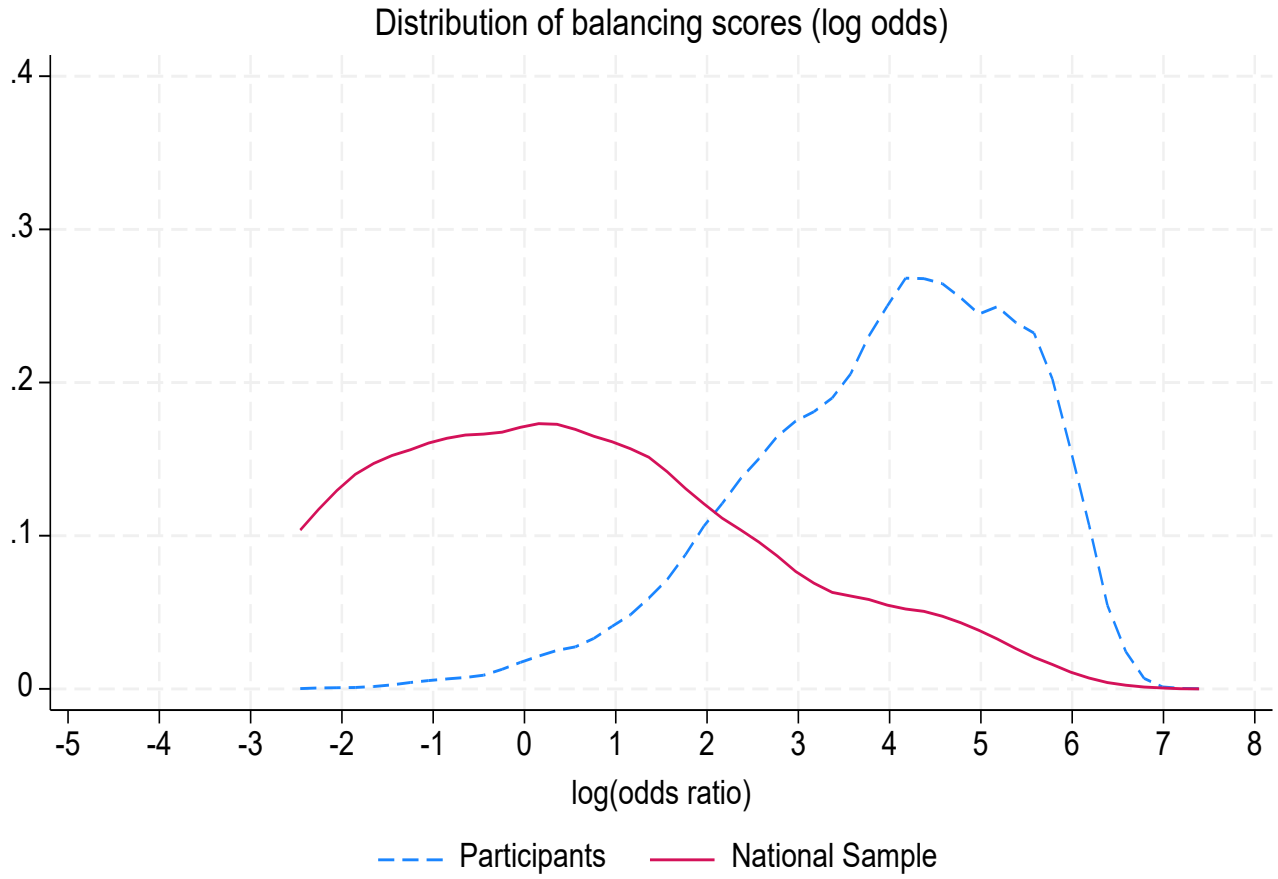
IPW values among untreated units vs propensity score



Balancing scores weighted by IPW – more overlap



Weighted by inverse probability eights



Unweighted (raw)

```
. regress foodexp D $xvar if matched==1 [aw=IPW]
(sum of wgt is 17,861.9699854255)
```

Source	SS	df	MS		
				Number of obs	= 8,113
				F(21, 8091)	= 93.96
Model	521093439	21	24813973.3	Prob > F	= 0.0000
Residual	2.1369e+09	8,091	264104.317	R-squared	= 0.1961
				Adj R-squared	= 0.1940
Total	2.6580e+09	8,112	327657.972	Root MSE	= 513.91

foodexp	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
D	-242.3174	11.93493	-20.30	0.000	-265.7129	-218.9218
dependency_ratio	-267.6263	21.27996	-12.58	0.000	-309.3405	-225.9121
FHH	-69.94924	19.95517	-3.51	0.000	-109.0665	-30.83197
primary	-159.814	15.55408	-10.27	0.000	-190.304	-129.324
secondary	29.70706	19.0703	1.56	0.119	-7.675627	67.08974
tertiary	26.27208	30.87374	0.85	0.395	-34.24838	86.79255
age_head	32.22766	9.321138	3.46	0.001	13.95584	50.49949
age_sq	-.363994	.1946579	-1.87	0.062	-.7455736	.0175855
age_cubed	.001075	.0012712	0.85	0.398	-.0014168	.0035668
kids	172.9239	11.52241	15.01	0.000	150.337	195.5108

(Regression table has been truncated…)

Doubly Robust Method

1) Regression on matched sample
2) Covariates
3) IPW

An Assessment of Propensity Score Matching as a Nonexperimental Impact Estimator

Evidence from Mexico's PROGRESA Program

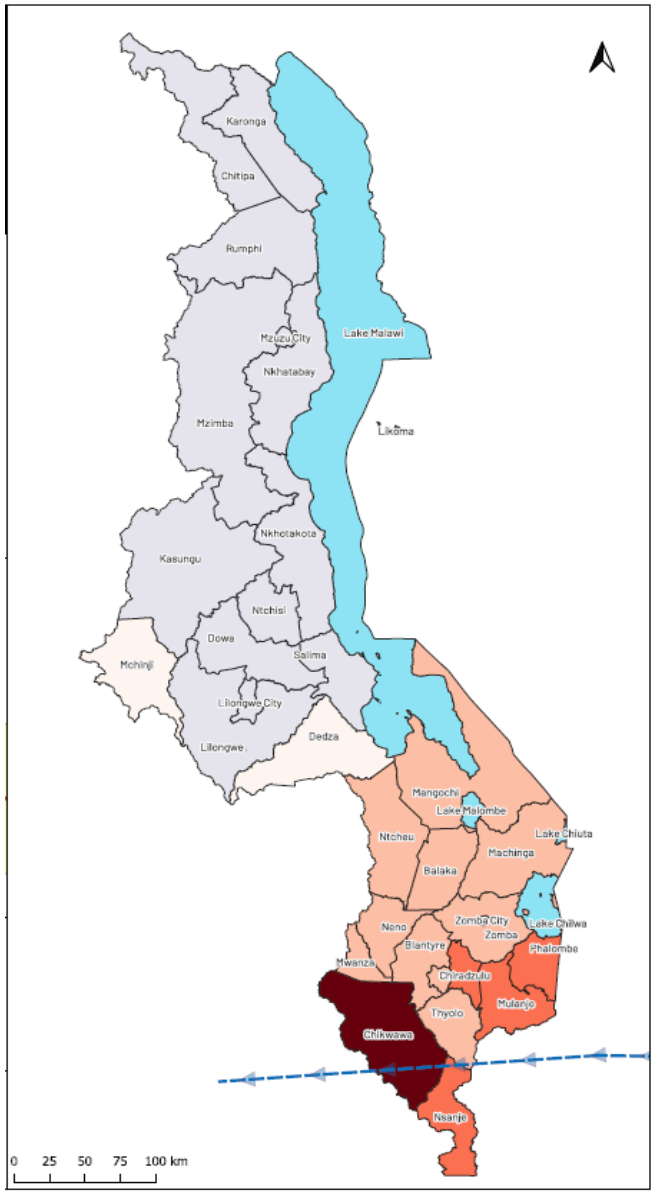
Juan Jose Diaz
Sudhanshu Handa

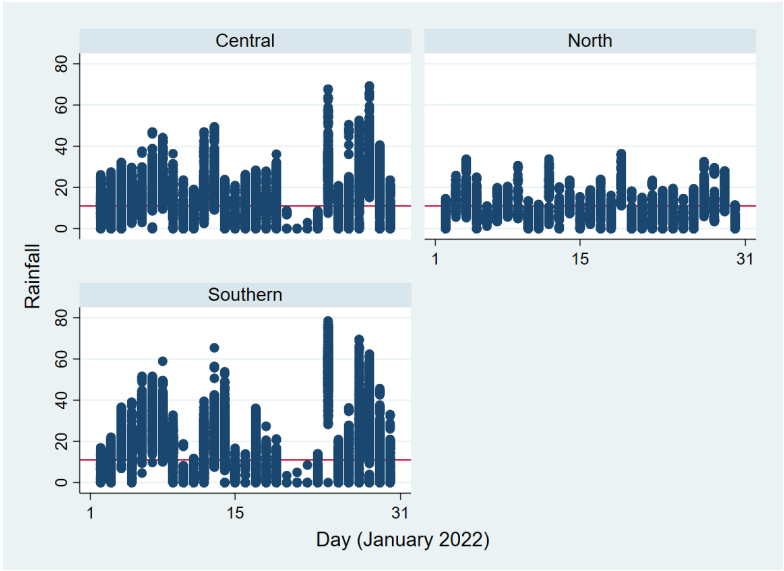
ABSTRACT

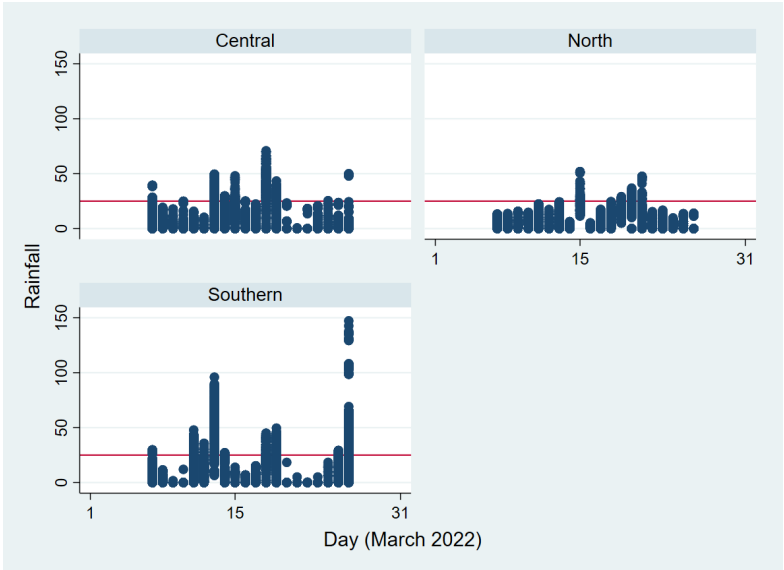
Not all policy questions can be addressed by social experiments. Nonexperimental evaluation methods provide an alternative to experimental designs but their results depend on untestable assumptions. This paper presents evidence on the reliability of propensity score matching (PSM), which estimates treatment effects under the assumption of selection on observables, using a social experiment designed to evaluate the PROGRESA program in Mexico. We find that PSM performs well for outcomes that are measured comparably across survey instruments and when a rich set of control variables is available. However, even small differences in the way outcomes are measured can lead to bias in the technique.

Data for this example comes from this Article: Journal of Human Resource Vol.41(2), 2006.

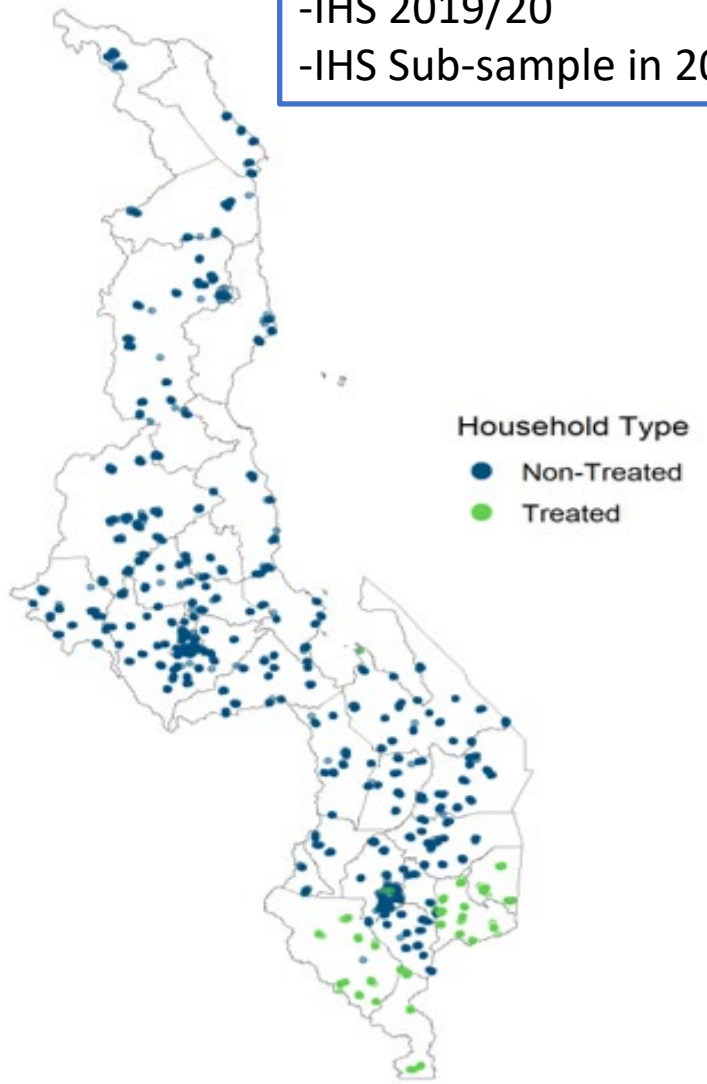
Background for STATA example: Impact of Tropical Storms Ana and Gombe in Malawi Q1 2022







-IHS 2019/20
-IHS Sub-sample in 2023



https://www.unicef.org/malawi/reports/impact-multiple-shocks-most-vulnerable-malawi-0

Government of Malawi

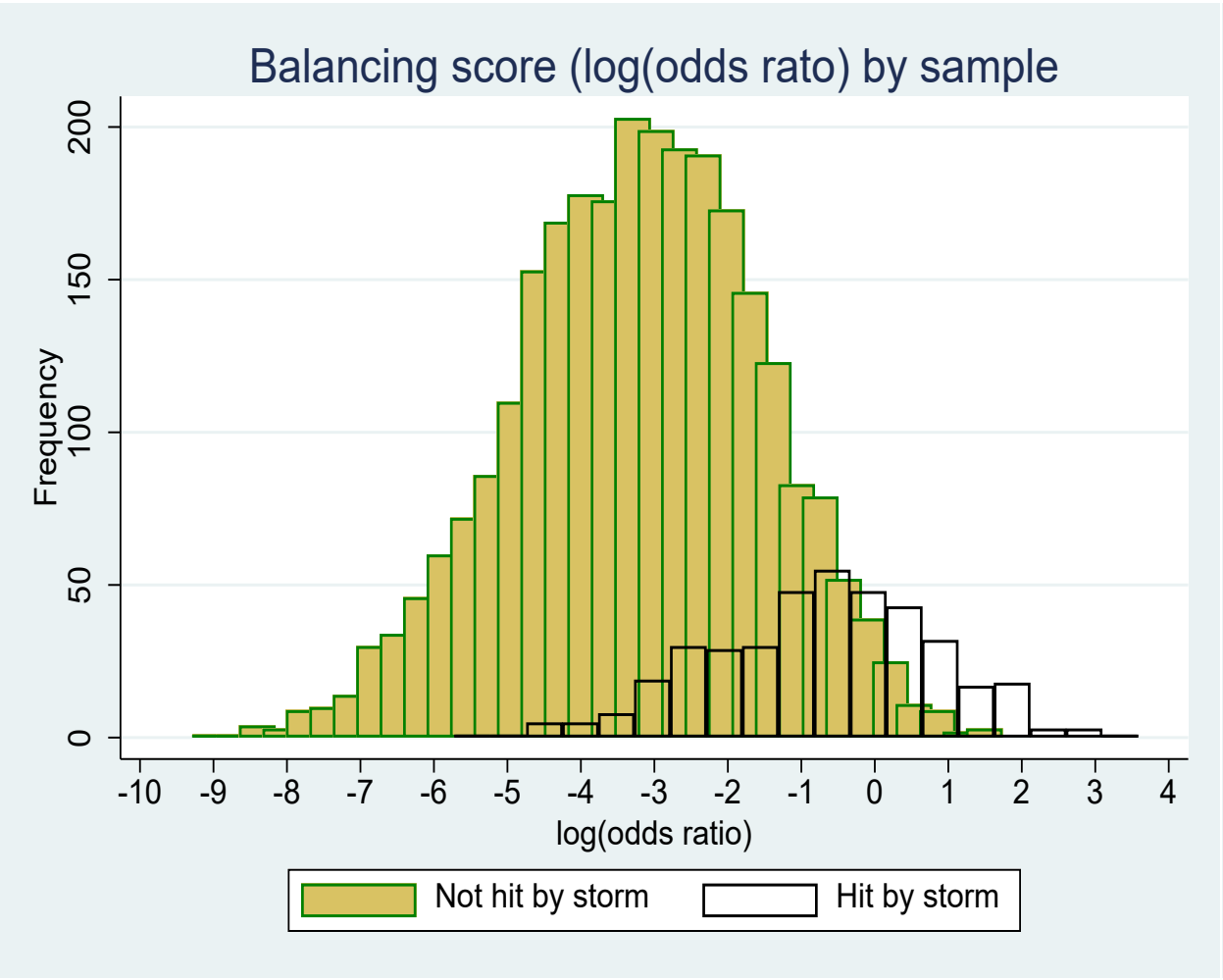
National Statistical Office

JOINT SDG FUND

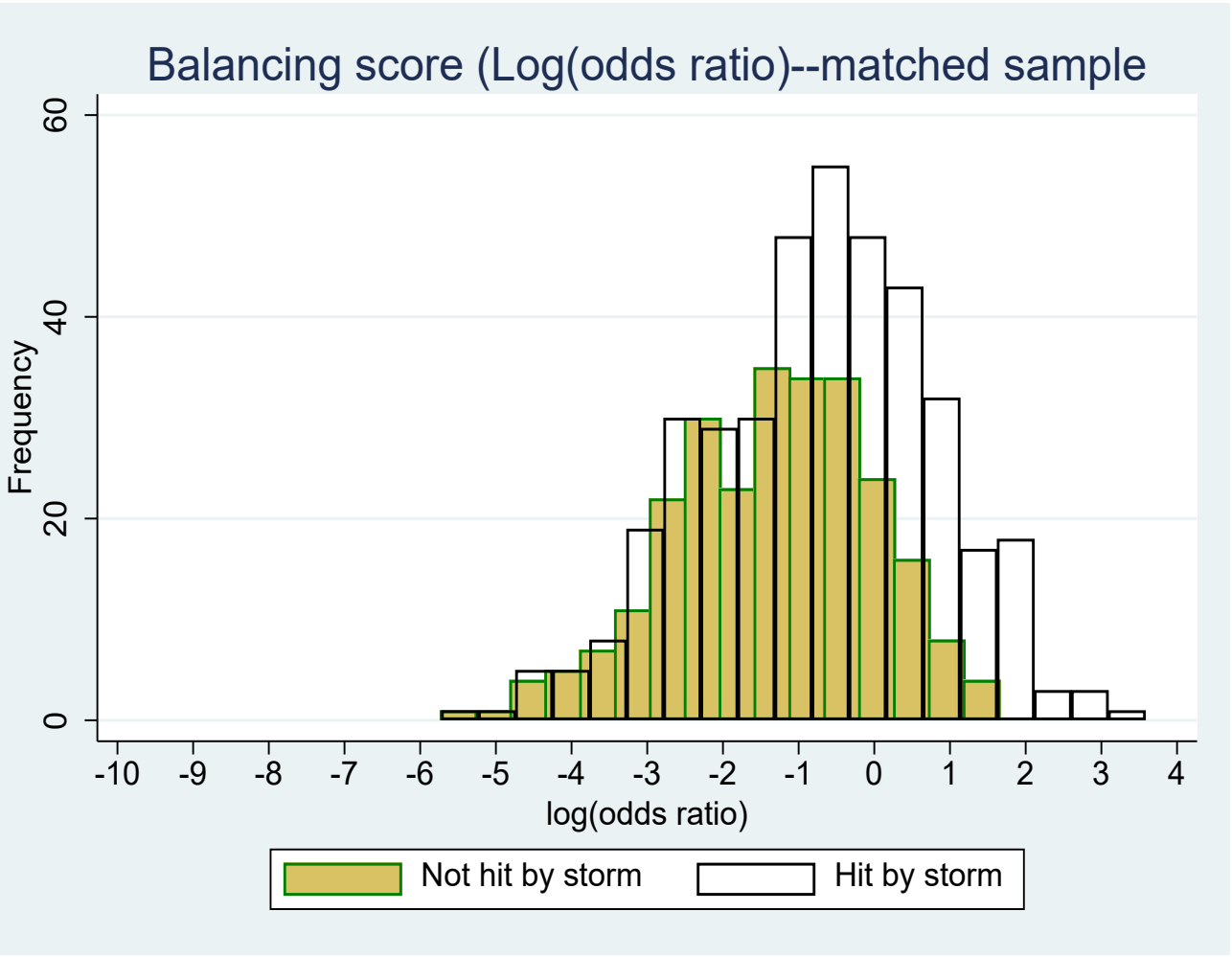
Impact of Multiple Shocks on the Most Vulnerable in Malawi

Some results from the Malawi shocks example

Balancing scores for entire sample



Balancing scores for matched sample only



Summary of estimates for effect of tropical storms on # of months of food shortage L12 months

Estimation	N	Coefficient ('impact')	T-statistics
Uncontrolled mean difference	3083	1.85	9.15
Regression adjusted difference	3083	1.09	7.28
NN matching w replacement	655	1.85	4.49
NN matching w/o replacement	792	1.51	5.34
Doubly robust w replacement	655	1.22	2.59
Doubly robust w/o replacement	792	1.09	2.95

Range of impacts is 1.85 to 1.09, all are statistically significant
In the end, matching doesn't do better than regression!

NN=nearest neighbor matching

Most common uses of PSM

- Typical application is a national data set with some treated units
  - Use matching to ‘select’ an equivalent comparison group and compare outcomes
  - Some farmers receive a treatment, use other farmers as a comparison group

- Other application is with program data on a treatment group, and use a national survey to pick matches (our example)
  - Use log(odds ratio) for matching due to over-sample of treated units

What you should take away….

- If the assumption behind matching doesn't hold, everything else is irrelevant
  - Assumption is 'selection on observables'
  - Matching does remove some bias due to differences in unobservables, but how much?
  - **Your paper MUST address the appropriateness of the assumption**

- Consensus now is that doubly robust approach is the best
  - Results not sensitive to the matching algorithm (caliper, NN, kernel)
  - Spending days exploring different matching algorithms is not a great use of time, unless you consider it leisure

  - **Get to the doubly robust method quickly**