PUBL0055 Introduction to Quantitative Research Methods - Seminar

Irvin Chen-Yu, Lee

University College of London

7th October 2022

What are We Going to Learn Today.....

- Environment of R
- Working on Dataframes
- Some basic functions
- A little bit statistical analysis

Who am I?

- 

Who am I?

- 
- Second-year Ph.D. student in Government at Essex

Who am I?

- 
- Second-year Ph.D. student in Government at Essex
- The spirit of social science research is about telling a good story, and methods are the tools that enable us to tell good stories!

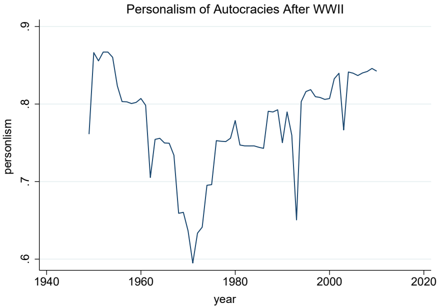
Who am I?

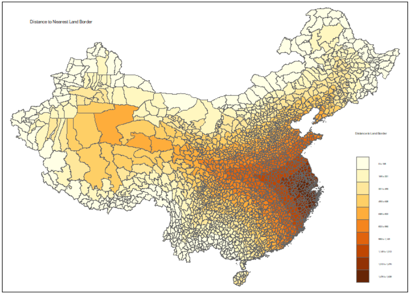
- 
- Second-year Ph.D. student in Government at Essex
- The spirit of social science research is about telling a good story, and methods are the tools that enable us to tell good stories!
- I like cycling, travelling, and watching good movies and TV series when I am not working

What am I Studying?

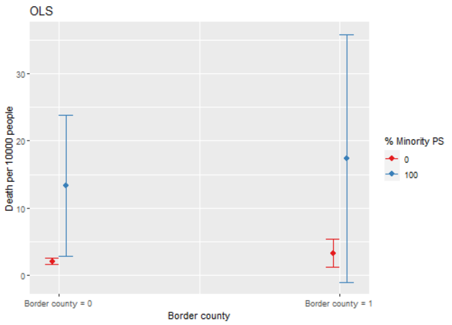
- Legacies of state repression in autocratic regimes
- Dynamic interaction between mobilization and repression in civil wars
- I apply quantitative methods, especially causal inference, geospatial analysis, and network analysis

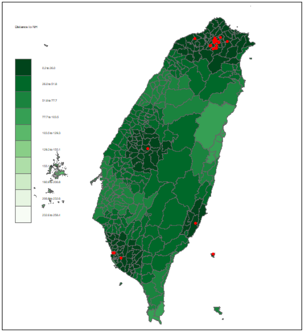
Some of the Projects I'm Working on





Some of the Projects I'm Working on





PUBL0055 Seminar: Causality and Experiments

Irvin Chen-Yu, Lee

University College of London

14th October 2022

Let's refresh what we learned this week

- Causality
- Potential outcome framework
- Randomized treatment assignment
- Experiments

Thinking about Causality

- A critical aim of social science research is to establish the causality between social phenomena
- Does democratization cause economic growth?
- What is the effect of electoral system on turnout?
- Does trade reduce conflict between states?

Potential Outcome Framework

- Instead of asking "whether X causes Y", ask "what would Y become in the absence/presence of X"
- The fundamental problem of causal inference: potential outcomes can never be observed!
- Randomization as a solution $\rightarrow$ Experiment as the "Golden Standard" of causal inference

Potential Outcome Framework

- Instead of asking ' … what would Y become in the absence/pre …
- The fundamental p … potential outcomes can never be observed!
- Randomization as … the "Golden Standard" of causal inference

The Democracy of Dating

- Easton and Holbein (2021) examines whether political preference influences people's decisions to develop romantic relationships

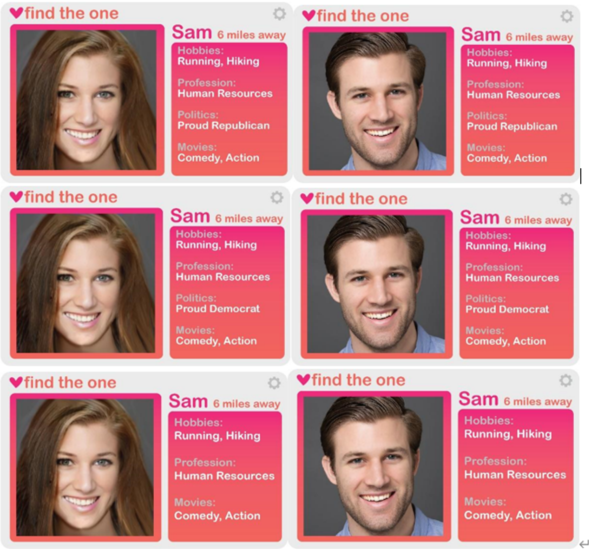
The Democracy of Dating

- Easton and Holbein (2021) examines whether political preference influences people's decisions to develop romantic relationships
- The researchers create fake profiles on dating apps and randomly assign these profiles to participants

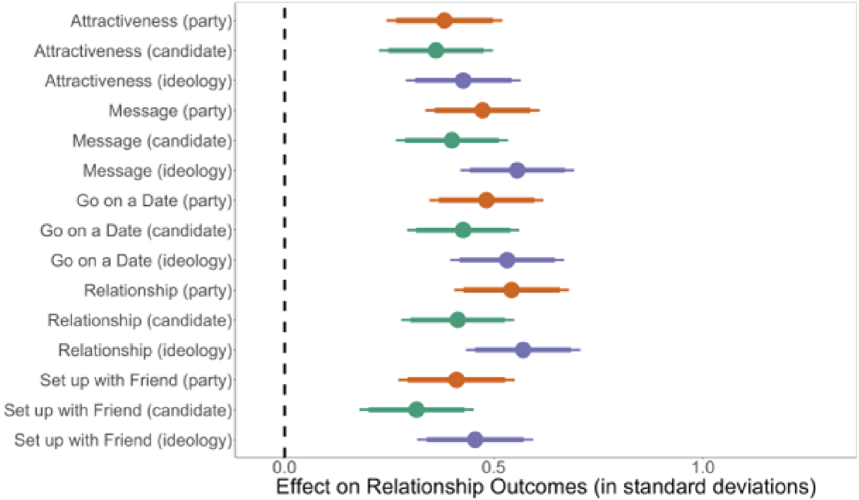
The Democracy of Dating

- Easton and Holbein (2021) examines whether political preference influences people's decisions to develop romantic relationships
- The researchers create fake profiles on dating apps and randomly assign these profiles to participants
- After viewing the assigned profiles, respondents were asked questions regarding relationship formation (e.g. attractiveness, willing to message/date, forming relationship)

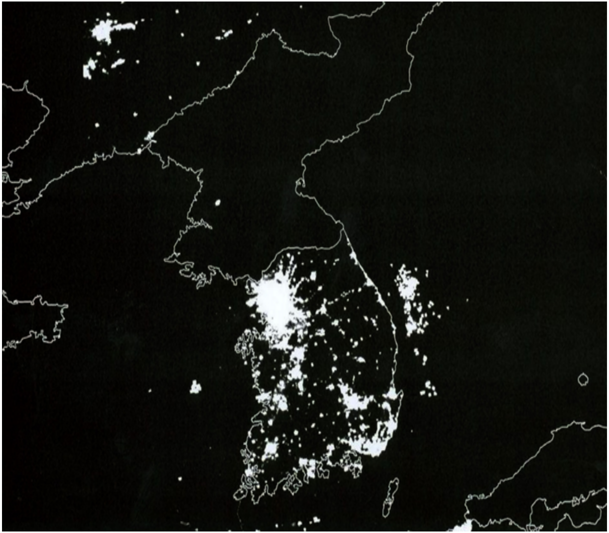
Figure A1: Survey Treatment Profiles



Political Preference Does Have an Effect on Formation of Romantic Relationship!



Quasi-experiments in Human Societies



PUBL0055 Seminar: Descriptive Data Analysis

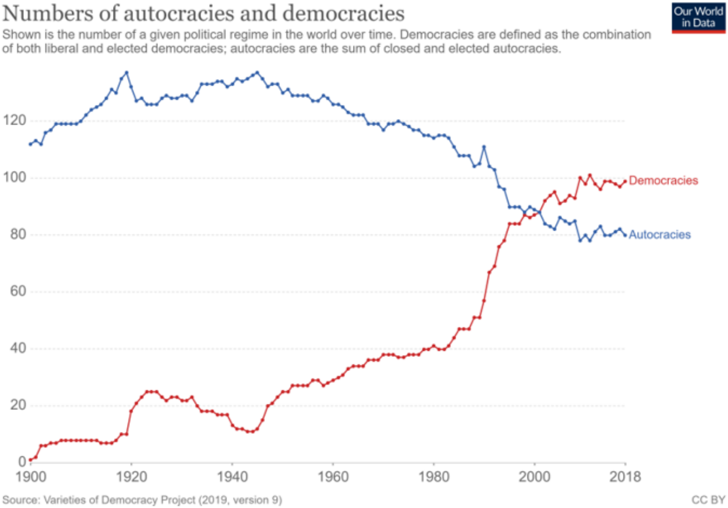
Irvin Chen-Yu, Lee

University College of London

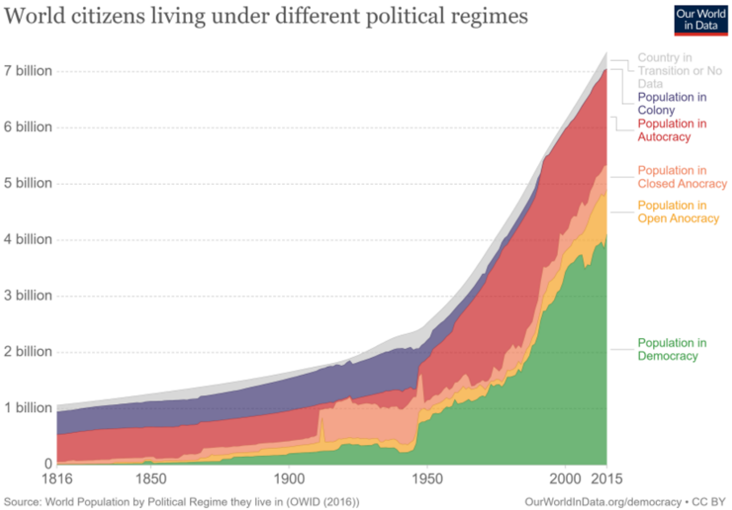
21th October 2022

- Descriptive data analysis can provide us valuable information even without statistical inference
- Data Visualization can be a powerful tool to display descriptive statistics

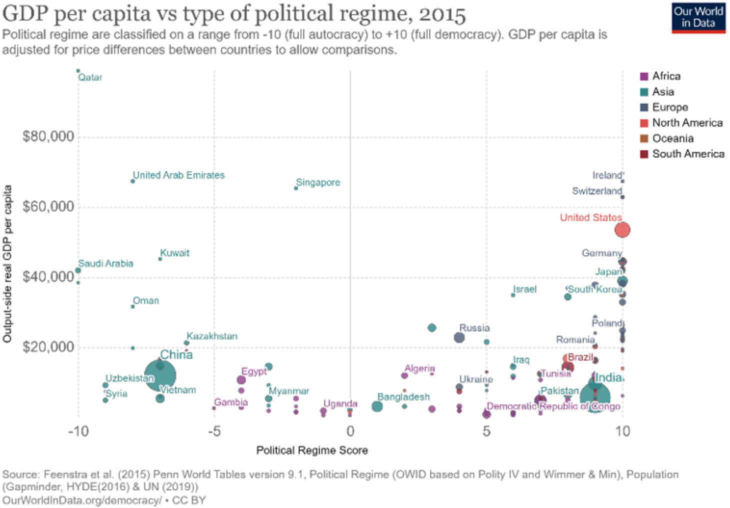
Evolution of Regime Type Over Time



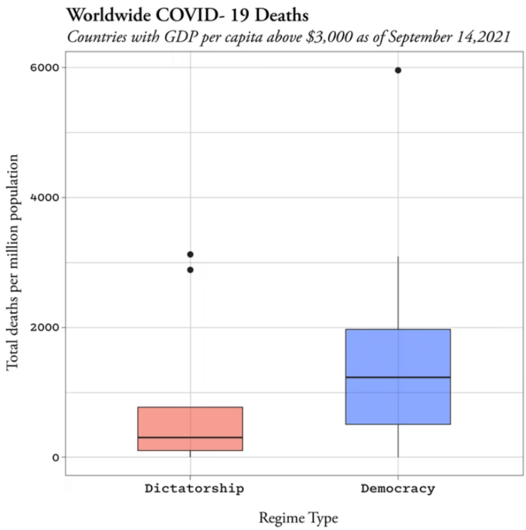
More People Live in Democracies Nowadays



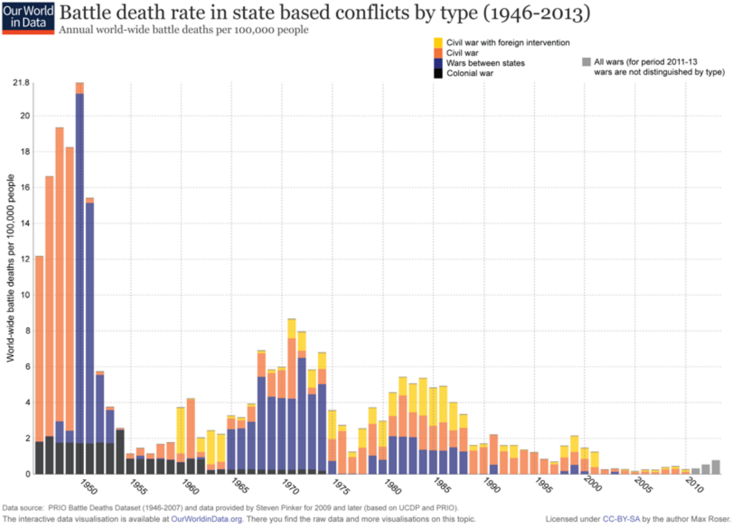
Are Democracies Richer?



Did Democracies Perform Worse in Covid 19?



We Live in a More Peaceful World



PUBL0055 Seminar: Regression and Causal Inference

Irvin Chen-Yu, Lee

University College of London

16th November 2022

Linear Regression

- Multiple linear regression:

$$Y = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \varepsilon$$

- Linear regression finds the slope (β) that fits the least summation of squared residuals $\rightarrow \sum_{n=1}^{i} (\hat{Y} - Y)^2$
- Interpretation: One unit increase in X is associated with β change in the dependent variable Y
- The coefficients of dummy and categorical variables has to be interpreted in a relative sense $\rightarrow$ It has to be compared against the referenced group!

Interaction Effect

- Consider an interaction model:

$$ImmiSupport = \beta_1 Age + \beta_2 Female + \beta_3 Age \cdot Female + \varepsilon$$

Interaction Effect

- Consider an interaction model:

$$ImmiSupport = \beta_1 Age + \beta_2 Female + \beta_3 Age \cdot Female + \varepsilon$$

- The main effect of age (β_1) is the effect of age conditional on **Female = 0**

Interaction Effect

- Consider an interaction model:

$$ImmiSupport = \beta_1 Age + \beta_2 Female + \beta_3 Age \cdot Female + \varepsilon$$

- The main effect of age (β_1) is the effect of age conditional on **Female = 0**
- ImmiSupport $= \beta_1$Age $+ \beta_2 \times (\mathit{Female} = 0) + \beta_3$ Age $\times (\mathit{Female} = 0)$

Interaction Effect

- Consider an interaction model:

$$ImmiSupport = \beta_1 Age + \beta_2 Female + \beta_3 Age \cdot Female + \varepsilon$$

- The main effect of age (β_1) is the effect of age conditional on **Female = 0**
- ImmiSupport $= \beta_1$Age $+ \beta_2 \times (Female = 0) + \beta_3$ Age $\times (Female = 0)$
- $= \beta_1$Age

Interpretation of Interaction Effect

- β_3 is the **change** in the effect of *Age* conditional on $Female = 1$

Interpretation of Interaction Effect

- β_3 is the **change** in the effect of *Age* conditional on $Female = 1$
- When $Female = 1$,

Interpretation of Interaction Effect

- β_3 is the **change** in the effect of *Age* conditional on $Female = 1$
- When $Female = 1$,
- $ImmiSupport = \beta_1 \text{Age} + \beta_2 \times (Female = 1) + \beta_3 \text{Age} \times (Female = 1)$

Interpretation of Interaction Effect

- β_3 is the **change** in the effect of *Age* conditional on $Female = 1$
- When $Female = 1$,
- $ImmiSupport = \beta_1\text{Age} + \beta_2 \times (Female = 1) + \beta_3\text{Age} \times (Female = 1)$
- $\qquad\qquad\qquad = \beta_1\text{Age} + \beta_2 + \beta_3\text{Age}$

Interpretation of Interaction Effect

- β_3 is the **change** in the effect of *Age* conditional on $Female = 1$
- When $Female = 1$,
- $ImmiSupport = \beta_1 \text{Age} + \beta_2 \times (Female = 1) + \beta_3 \text{Age} \times (Female = 1)$
- $= \beta_1 \text{Age} + \beta_2 + \beta_3 \text{Age}$
- $= (\beta_1 + \beta_3) \text{Age} + \beta_2$

Omitted Variable Bias

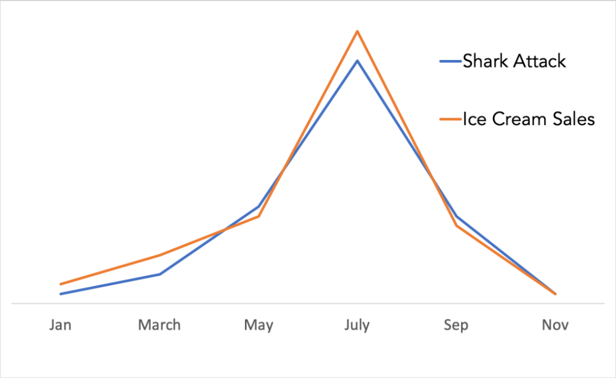
- When there are other variable **Z** both correlated to **X** and **Y**, the relationship between **X** and **Y** could be spurious and driven by variable **Z**

Omitted Variable Bias

- When there are other variable **Z** both correlated to **X** and **Y**, the relationship between **X** and **Y** could be spurious and driven by variable **Z**
- Think about the correlation between shark attack and ice cream sale......

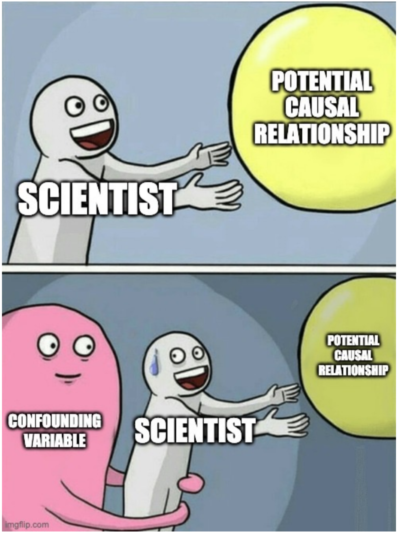
Omitted Variable Bias

- When there are other variable **Z** both correlated to **X** and **Y**, the relationship between **X** and **Y** could be spurious and driven by variable **Z**
- Think about the correlation between shark attack and ice cream sale......



-

Omitted Variable Bias could be Problematic......



Controlling Confounders

- We can "control" the confounding by including potential confounding variables in our regression models
- Control confounding variables that might correlate to your explanatory/independent **X** and dependent variable **Y**
- More controls does not always mean better! Please don't do this: ▸ Link

PUBL0055 Seminar: Panel Data, Fixed Effects, and DID

Irvin Chen-Yu, Lee

University College of London

2th December 2022

Panel Data

- In a panel data, each unit i is observed at multiple time period t

Country	WC Year	Win	Rank	Goals
England	2002	2	6	6
England	2006	3	7	6
England	2010	1	13	3
England	2014	0	26	2
England	2018	3	4	12

Table: Panel data example

- A balanced panel with N units and T time periods have $\mathrm{N}\times\mathrm{T}$ observations

Panel Data

- In a panel data, each unit i is observed at multiple time period t

Country	WC Year	Win	Rank	Goals
England	2002	2	6	6
England	2006	3	7	6
England	2010	1	13	3
England	2014	0	26	2
England	2018	3	4	12

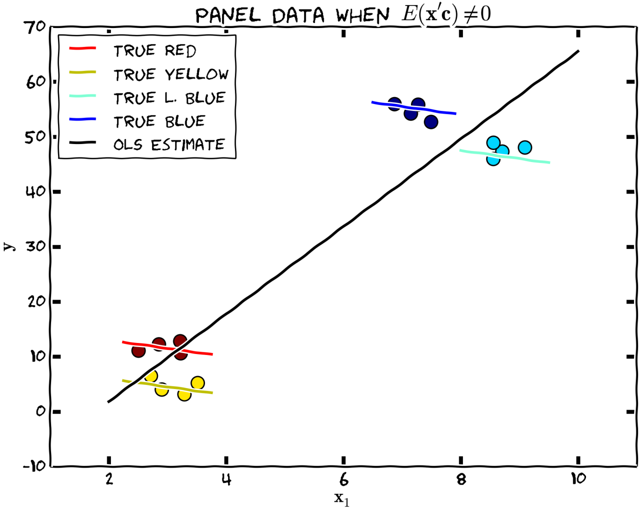
Table: Panel data example

- A balanced panel with N units and T time periods have N$\times$T observations
- Advantage of panel data: it enables us to trace the dynamic relationship between variables over time!

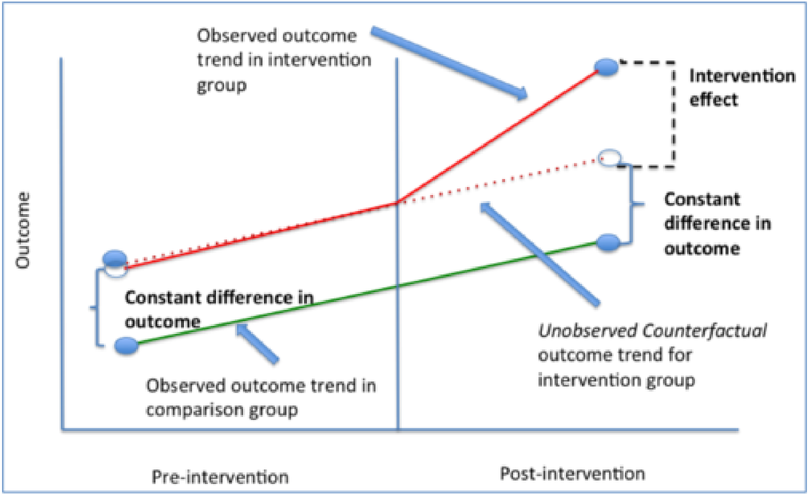
Fixed Effect (FE) Model

- Fixed effect removes variances between units and thus is also known as **within** estimator
- **Unit fixed effect** absorbs time-invariant factors within each unit
- **Time fixed effect** accounts for common exogenous shocks across all units
- A two-way fixed effect (TWFE) model includes both unit and time fixed effects

Pooled OLS vs Fixed Effect Model



Difference-in-differences (DID) Estimator



Difference-in-differences (DID) Estimator

	Treatment	Control
Pre	65	60
Post	85	75

Table: DID estimator

$$\begin{aligned}\text{ATE} = {} & [E(Post = 1 | Treat = 1) - E(Post = 0 | Treat = 1)] - [E(Post = 1 | Treat = 0) - E(Post = 0 | Treat = 0)] \\ & = (85 - 65) - (75 - 60) \\ & = 20 - 15 \\ & = 5\end{aligned}$$

- In a regression setting, a DID estimator can be specified as:

$$Y = \beta_1 Treat + \beta_2 Post + \beta_3 Treat \cdot Post + \varepsilon$$

Identifying Assumption

- Parallel trends assumption: Outcome in treatment and control groups has no differential trend in the absence of treatment intervention
- Untestable assumption! $\rightarrow$ Plot the trends in group means to bolster the confidence of parallel trends assumption

Does Victory of National Football Team Reduce Civil Conflict?

- Depetris-Chauvin et al. (2020) exploit the quasi-random variation in national football teams' qualification for African Cup of Nations to estimate the effect of qualification on civil conflcit

Team	Pld	W	D	L	GF	GA	GD	Pts
Mali	5	3	0	2	7	4	3	9
Zimbabwe	5	2	2	1	6	3	3	8
Cape Verde	5	2	1	2	5	6	-1	7
Liberia	5	1	1	3	5	10	−5	4

08/10/2011	Liberia	2 – 2	Mali
	Cape Verde	2 – 1	Zimbabwe

Team	Pld	W	D	L	GF	GA	GD	Pts
Mali	6	3	1	2	9	6	3	10
Cape Verde	6	3	1	2	7	7	0	10
Zimbabwe	6	2	2	2	7	5	2	8
Liberia	6	1	2	3	7	12	−5	5

FIGURE 3. EXAMPLE OF CLOSE QUALIFICATION: GROUP A, ACN 2012

-

Winning Football Does Reduce Civil Conflict, But Only in the Short Run!