

Subnational Geography and Bias in Surveys of Americans Living Abroad

Federal Committee of Statistical Methodology

March 9, 2018





Presentation Overview

I. Background

II. Data and Methodology

III. Results

IV. The Mechanism

V. Summary and Conclusion



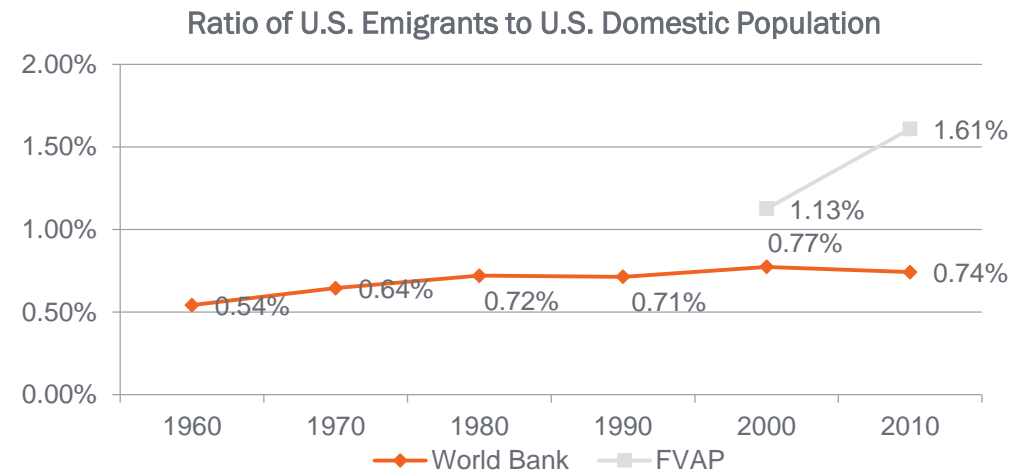
Background





Background

- **American citizens living abroad comprise increasing share of total U.S. citizen population.**
 - Data on this population with respect to engagement with United States may be of increasing interest to policymakers.
 - Example: Voting in U.S. elections





Data and Methodology





Background

Overseas Citizens Population Survey (OCPS):

- Sponsored by the Federal Voting Assistance Program (FVAP)
- Population/Frame: Registered overseas citizens who requested an absentee ballot in 2014
- Geographically stratified sample of 36,000
- Mailed an invitation to most respondents
 - Framed as a voting survey
- Each individual linked to voter history file
 - Possible to calculate population and sample vote rates for the 2015 General Election by country
- Subset of population geocoded (latitude and longitude coordinates of city of residence)
 - Possible to know whether they are in “core” or “peripheral” part of country of residence



Hypotheses

—

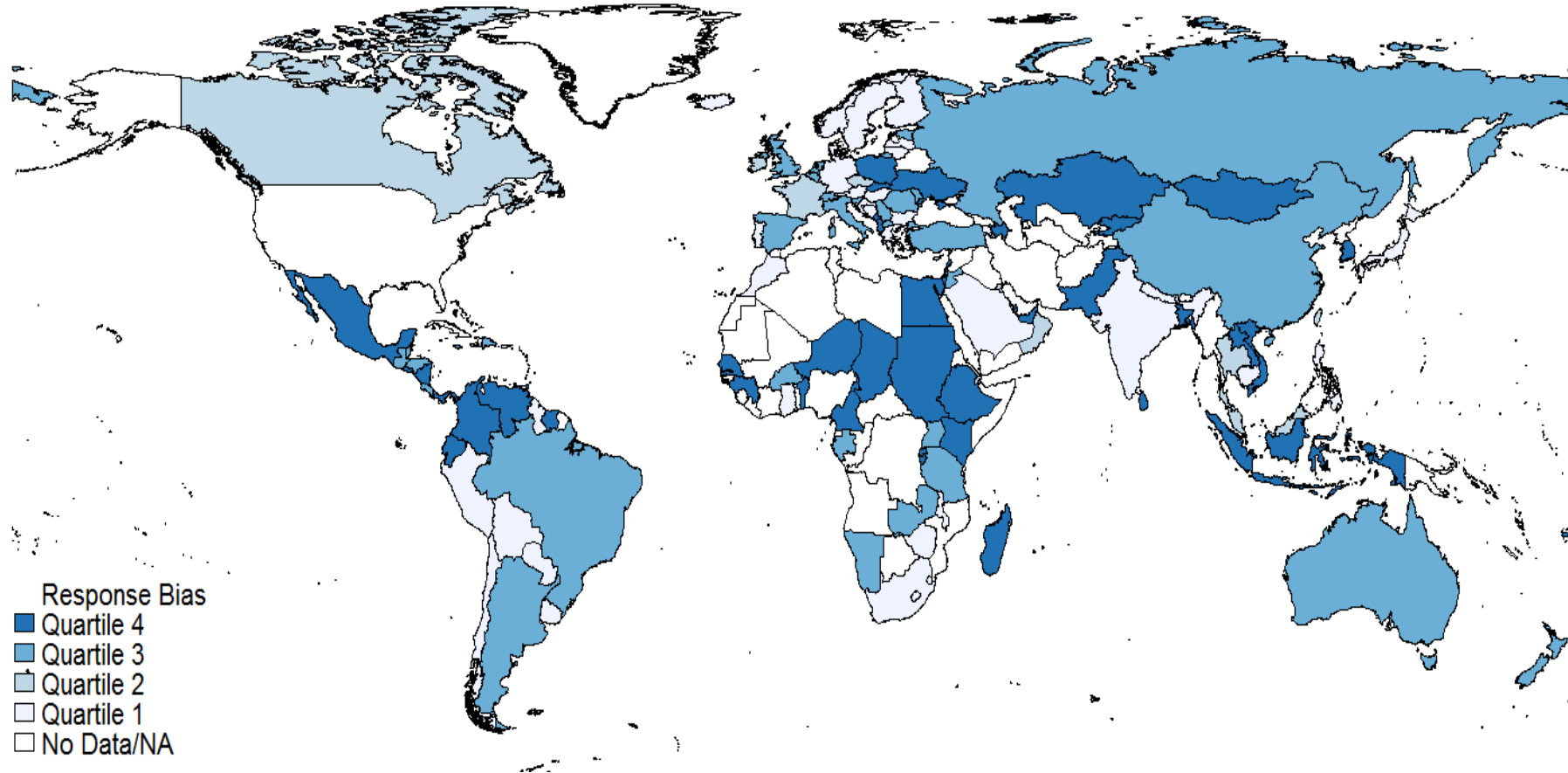
Bias in Vote Rates by Country:

$$Bias_c = Ln \left(\frac{P(Voted_{ic} = 1 | Responded_{ic} = 1)}{\overline{Voted}_c} \right)$$

- Expected to be positive due to topic bias in election turnout surveys
 - Sciarini (2016)
- Increasing in differences in response rates between voters and non-voters
- Decreasing in population vote rate
 - More scope for upward bias in vote rates in less developed countries
- Decreasing in baseline (non-voter) response rates
 - Response rate gap is more consequential when baseline response rates low.



Bias in Voting Rates by Country





Hypotheses

Hypothesis: Sample vote rates is more upwardly biased in less developed countries.

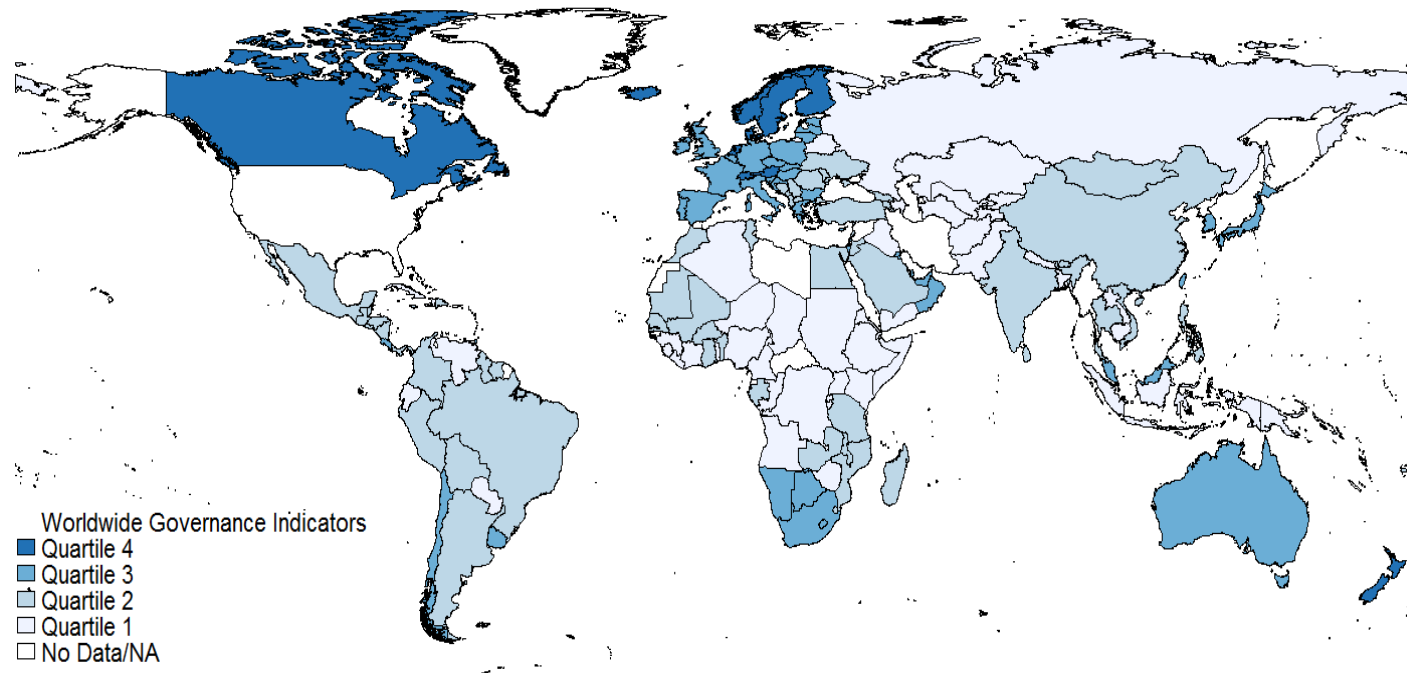
- H1:** Differences in response rates between voters and non-voters are negatively related to the country's level of development
- Poor infrastructure leads to low vote and response rates in peripheral regions of less developed countries.
 - There is a stronger correlation between voting and response propensity in less developed countries.
- H2:** Population vote rates are lower in less developed countries.
- Low vote rates in peripheral regions of less developed countries
- H3:** Baseline (Non-Voter) Response Rates Lower in Less Developed Countries
- Low response rates in peripheral regions of less developed countries



Level of Development by Country

—
Mean of World Bank's Worldwide Governance Indicators (WGI)

- Higher values or quartile = higher level of country development





Analysis Plan

Analysis of Country-Level Bias:

- **Models:**
 - Ordinary Least Squares (OLS)
 - Weighted Least Squares (WLS)
 - Weight = Population of absentee ballot requesters
- **Dependent Variable:**
 - Bias in voting rates
 - Population vote rates
 - Response rates of non-voters
 - Difference in response rates between voters and non-voters
- **Independent Variables:**
 - WGI



Results

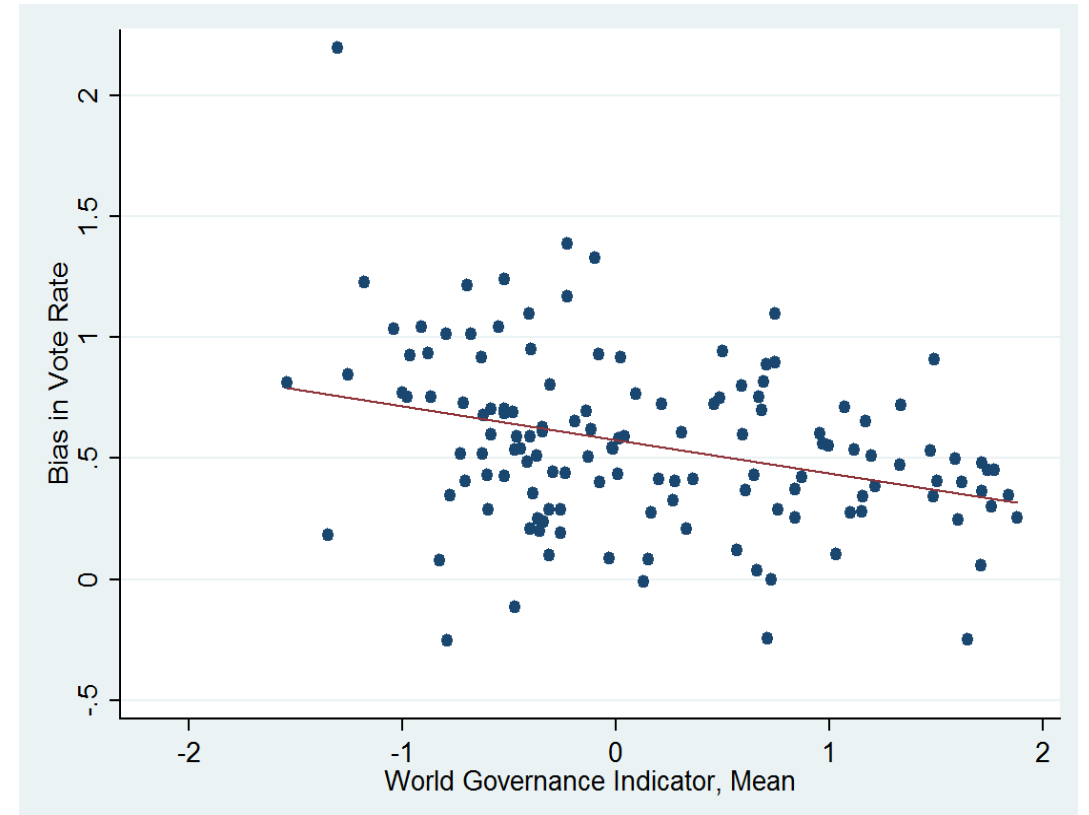




Results

Hypothesis: Sample vote rates are more upwardly biased in less developed countries.

Result: Not Rejected

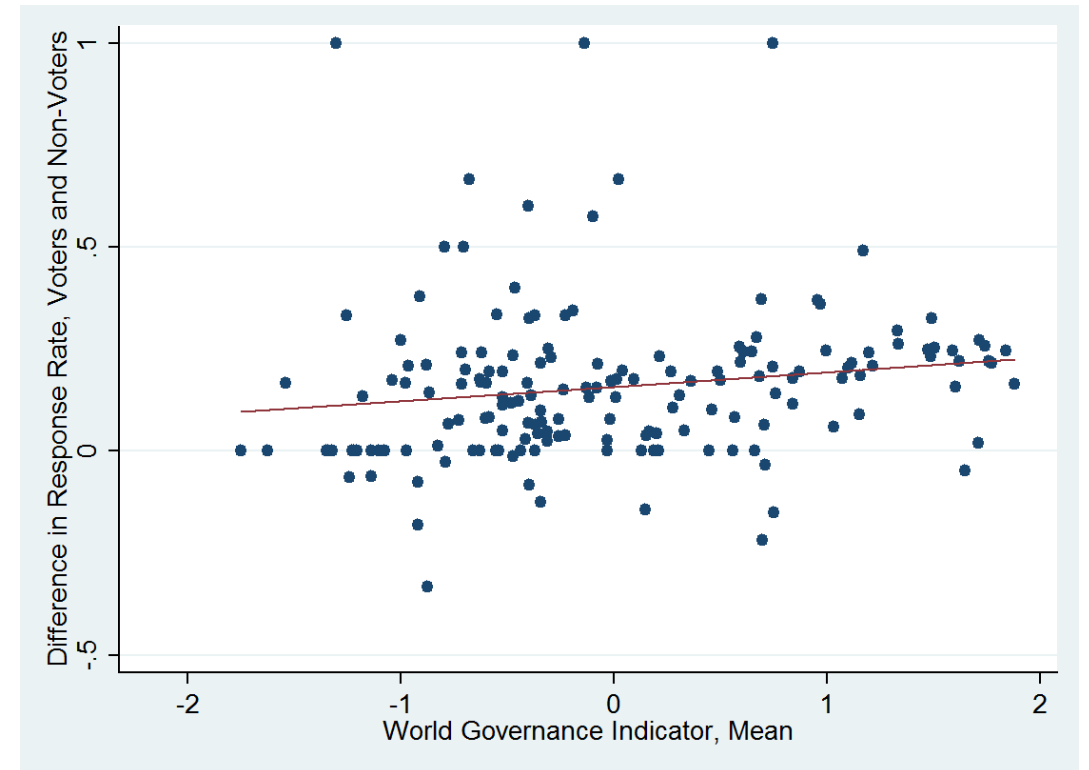




Results

H1: Differences in response rates between voters and non-voters are negatively related to a country level of development.

Result: Rejected

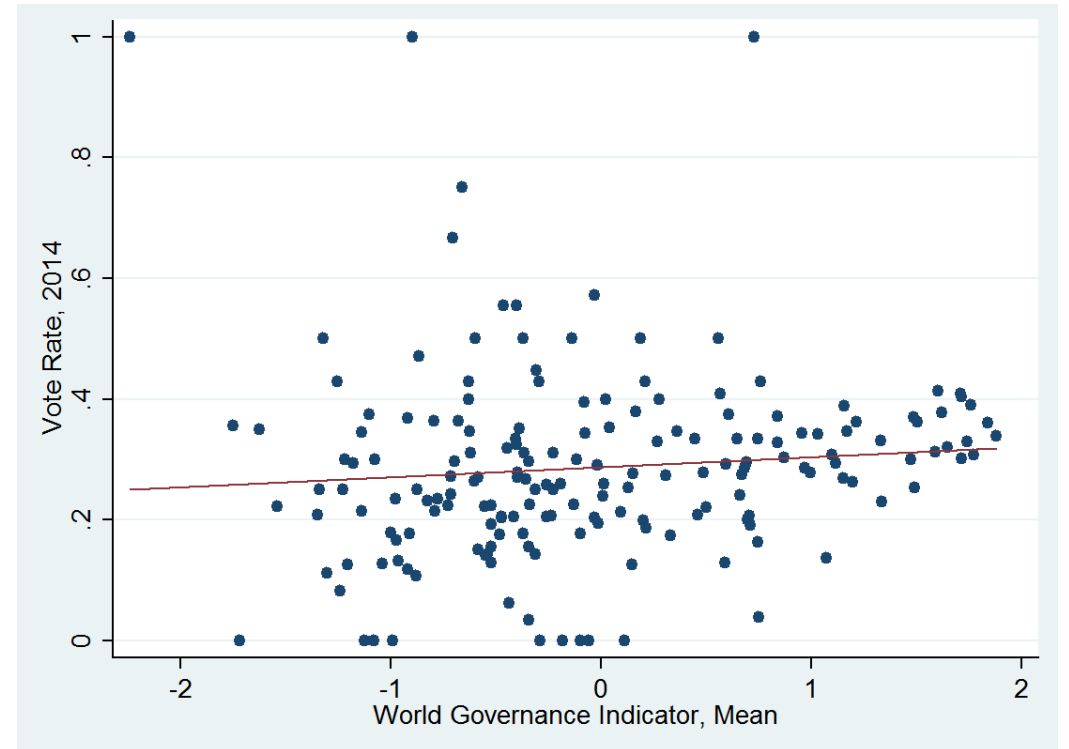




Results

H2: Population vote rates are lower in less developed countries.

Result: Not Rejected

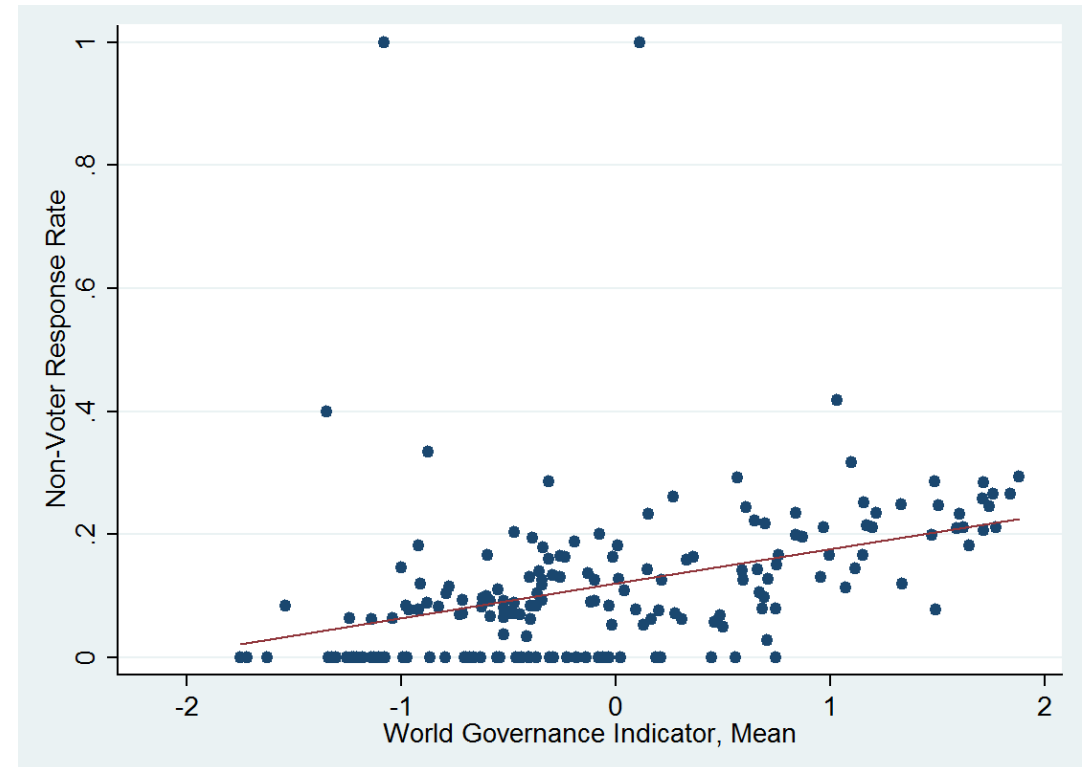




Results

H3: Baseline (non-voter) response rates are lower in less developed countries.

Result: Not Rejected





Mechanism





Within-Country Variation in Voting and Non-Response

- Why is the response gap between voters and non-voters high in more developed countries?
- Model:
 - $G_{ic} = \beta_1 y_{ic} + \beta_2 (y_{ic} * WGI_c) + \beta_3 X_{ic} + \beta_4 (X_{ic} * WGI_c) + C_c + e_{ic}$
 - G_{ic} is proxy for how accessible/peripheral individual's region of the country is.
 - y_{ic} is an indicator for whether an individual voted or responded to the OCPS.
 - WGI_c is WGI for an individual's country.
 - X is a vector of control variables (e.g. age, gender, state of legal residence).
 - C_c is a set of country fixed effects.



Within-Country Variation in Voting and Non-Response

- **Proxies for Accessibility of an Individual's Location (G_{ic}):**
 - Ln(Time to Respond to the Survey)
 - Proxy for internal mailing time
 - Ruggedness of terrain
 - Ln(Distance to Coast)
 - Ln(Distance to Nearest Large City)
 - Ln(Population of Nearest Large City)
 - Ln(Distance to Capital City)
 - Ln(Nighttime Luminosity)
 - Predicted Ln(Time to Respond to the Survey)
 - Function of other geographic covariates



Locations of Respondents and Non-Respondents

- - **Respondents are statistically significantly more likely to reside in more peripheral locations than non-respondents.**
 - Further from large city
 - Nearest large city is smaller
 - Further from capital city
 - Less luminous region
 - Difference in luminosity is smaller in less developed countries.



Locations of Voters and Non-Voters

- **Voters are statistically significantly more likely to reside in more peripheral locations than non-voters.**
 - Further from large city
 - Nearest large city is smaller
 - Further from capital city
 - Less luminous region
 - Differences between voters and non-voters' locations do not vary with level of development.



Interpretation of Results

- **Individuals in more peripheral regions are more motivated to vote.**
 - Results in higher response rates in peripheral regions due to topic interest
 - However, the response rate difference between voters and non-voters is attenuated in developing countries.
 - Voters live in regions of developing countries with less infrastructure/higher mailing times.
- **Key takeaways:**
 - Within-country variation in postal reliability may lead to lower bias by making it relatively harder for voters to respond to the survey.
 - However, bias in vote rates is still higher in developing countries due to lower population vote rates and lower baseline response rates.
 - This may reflect higher between-country (port to port) mailing times for developing countries.



Summary and Conclusion





Summary

Examined potential biases in estimates of engagement with the United States by Americans living abroad from international mail surveys:

- Case Study: Voting in the 2014 General Election by absentee ballot requesters
- Hypothesis: Larger upward bias in voting rates for Americans living in less developed countries
 - Weaker infrastructure makes Americans in more peripheral (e.g.: rural) regions of these countries both less likely to vote and less likely to respond.
- Results: Larger upward bias in vote rates for less developed countries
 - But, the gap in response rates between voters and non-voters is larger in more developed countries.
 - May be due to fact that Americans living in more rural regions are more likely to vote
 - Developing countries have generally lower response and voting rates, leading to more scope for bias.



Conclusions

—

Variation in mail reliability may not just lead to biased inferences about the engagement of Americans as a whole, but may lead one to underestimate geographic variation in levels of engagement.

- May lead to ineffective targeting of outreach efforts by policymakers and stakeholders

Attempts to boost response rates may actually increase bias in estimates of engagement.

- Longer fielding times may disproportionately increase response rates of engaged Americans who were already more likely to respond.
- The ability to mitigate this bias through weighting/post-stratification is likely to hinge on the degree to which objective proxies of outcomes of interest are available for the survey frame.
 - Example: availability of vote history in OCPS



Questions?

www.ForsmarshGroup.com

