

Challenges of working with mobile network big data

Gabriel Kreindler, Sriganesh Lokanathan

Responsible use of mobile meta-data to support public purposes

Jetwing Lagoon, Negombo

8 August 2014



This work was carried out with the aid of a grant from the International Development Research Centre, Ottawa, Canada.

Challenges

| Challenge | Solution(s) |
|---------------------------------------|--|
| Data is biased towards frequent users | <ul style="list-style-type: none">• Understand and adjust for selection bias |
| Data sparsity | <ul style="list-style-type: none">• Interpolation techniques• Probability based models |
| Different tower densities | <ul style="list-style-type: none">• Different scale of analyses depending on region |
| Validating results | <ul style="list-style-type: none">• Using other data sources e.g. data from Dept. of Census and Statistics, transportation survey data, etc. |

Mobile phone data sample selection

Who are the Mobile Phone Users?

- Mobile phone users are not a random sample of the population.
- Important to understand the sample selection.
- This will allow us to correct the bias (e.g. using probability weights)

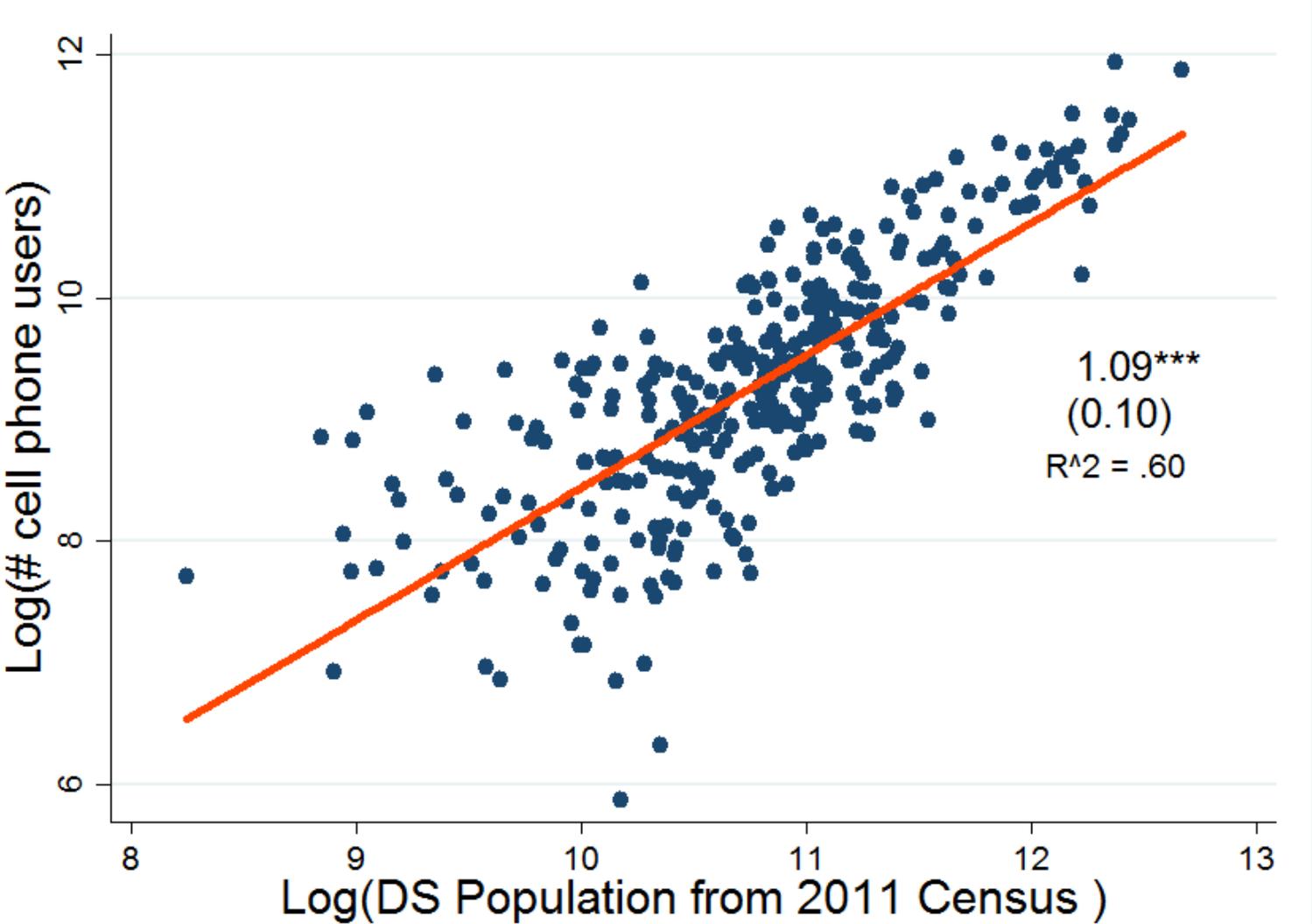
Goal: Investigate sample selection at Divisional Secretariat (DSD) level.

Limitation: ignores within-DSD selection.

How we Assign Mobile Phone Users to DSDs

- When a user makes/receives a phone call, localize the DSD where this happens.
- Analyze “home” interval: calls between 9pm and 5am
- For each user, select the “home” DSD where the user was localized the most times. (We use a 4-month period.)

Graphical Comparison with Census Data



Graphical Comparison with Census Data

- Slope close to 1 suggests that the number of cellphone users is proportional to DSD population.

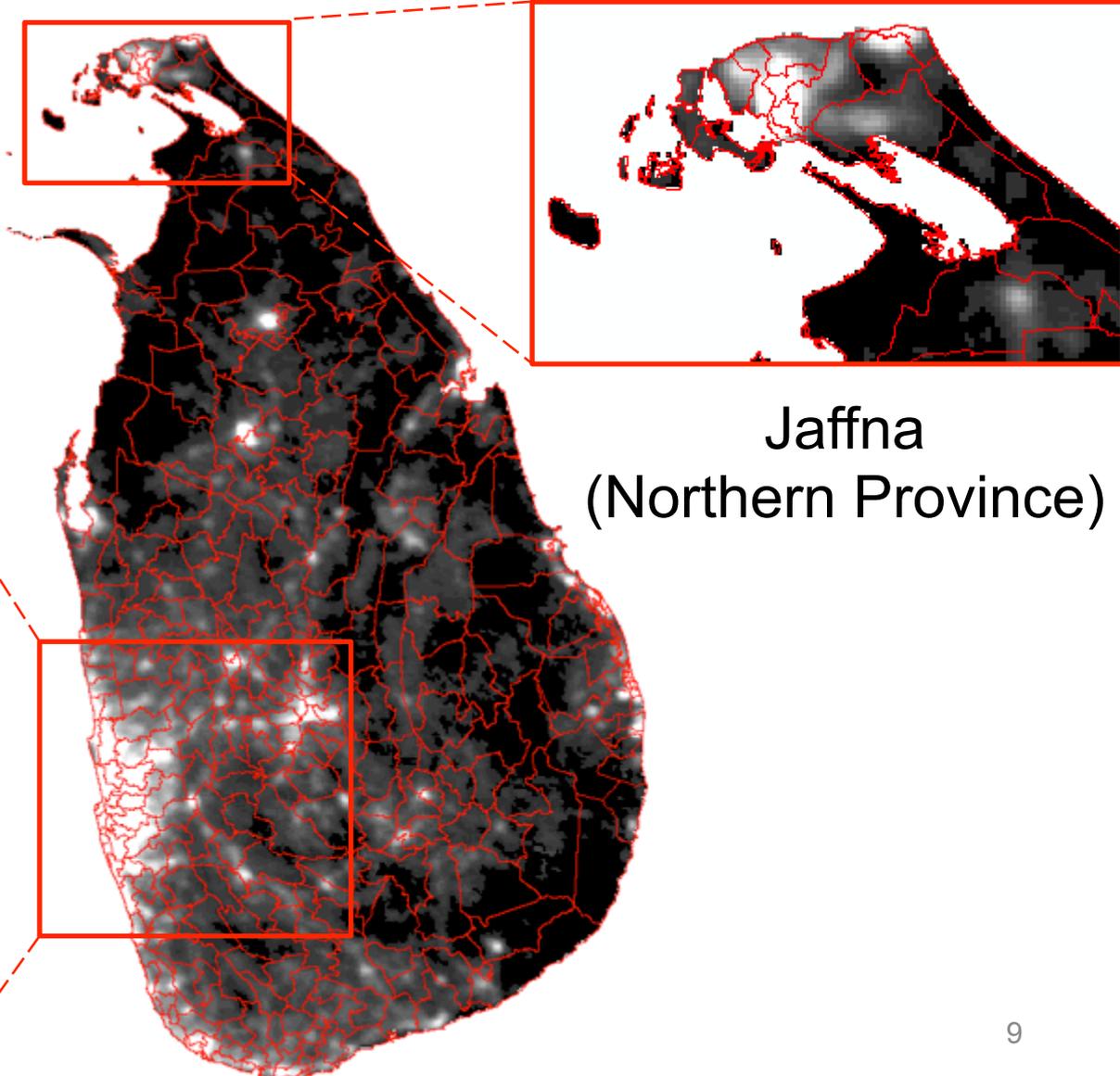
Regression Analysis

- Ordinary Least Square (OLS) regression.
- Include other explanatory variables:
 - Nighttime lights (proxy for economic development)
 - District fraction of households with mobile phones (2011 Census)
 - North and East province indicators

Satellite nighttime lights (2012)

Western
Province, Kandy
and Kurunegala

Jaffna
(Northern Province)



Adding Explanatory Power

| | (1) | (2) | (3) |
|------------------------------------|-----------------------------|-------------------|-------------------|
| | Log(number of users) | | |
| Log(DSD population) | 1.09*** (0.10) | 1.09*** (0.08) | 1.28*** (0.05) |
| Nightlight intensity per capita | | 1.35*** (0.24) | 0.71*** (0.21) |
| Province fraction has mobile phone | | | 0.03*** (0.01) |
| North Province | | | 1.34*** (0.17) |
| East Province | | | 0.60*** (0.10) |
| Observations | 326 | 326 | 326 |
| Adjusted R-squared | 0.60 | 0.64 | 0.77 |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Regression Analysis Results

- Economic development associated with more mobile phone users
- North and East provinces have more users than predicted by population, night lights, etc.
- Once North and East are accounted for, the number of mobile phone users grows faster than linear in DSD population (coefficient = 1.28)

High-frequency Mobile Phone Users

- Also important: sample selection of frequent mobile phone users.
- Results:
 - Large DSDs have higher-than-proportional share of high-frequency users.
- These results can help us compensate for sample selection bias in analyses of mobile phone data.

Data Sources

- Mobile phone data:
 - Voice calls over a period of 4 months
- Population and (district-level) fraction of households with mobile phones:
 - 2011 Census population data at DSD level.