A Gravity Model for Mobile Phone Calls

Gabriel Kreindler, LIRNE*asia*

Responsible use of mobile meta-data to support public purposes

Jetwing Lagoon, Negombo

8 August 2014

Who Calls Whom in Sri Lanka?

- We seek to understand the pattern of aggregate communication between areas.
  - depends on myriad factors: business and personal relations, ethnic, religious similarity, etc.

- We focus on three simple factors:
  - Population
  - Distance
  - IT literacy (broadly defined)

First, Visualizing the Volume of Calls

- The next slide plots the (log, normalized) volume of calls between provinces

-1	-0.75	-0.5	-0.25	0	0.25	0.5	0.75	1

Few Calls ──────────▶ Many Calls
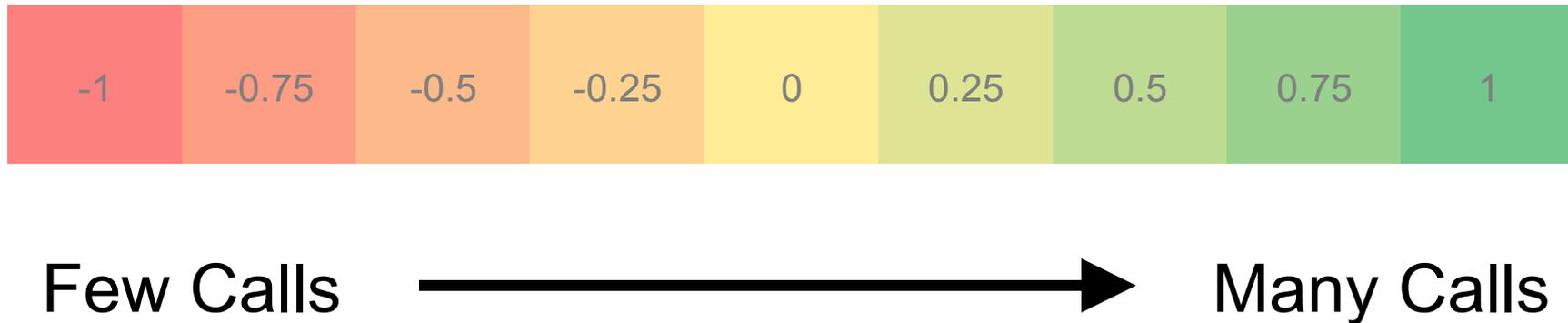
Total Volume of Calls between Provinces

- Within-province calls are generally high
- Western calls the most (both intra- and inter-province)

	N	NW	NC	E	Uva	Central	Saba	W	S
N	1.26	-2.13	-1.91	-2.01	-2.29	-1.83	-2.60	-0.77	-2.29
NW	-2.12	-0.53	-2.28	-3.42	-3.43	-2.36	-3.07	-0.76	-3.12
NC	-1.91	-2.28	-0.90	-2.89	-3.34	-2.11	-3.38	-1.19	-3.15
E	-2.01	-3.42	-2.89	-0.80	-2.85	-2.79	-3.68	-1.61	-3.47
Uva	-2.29	-3.43	-3.34	-2.85	-0.18	-2.32	-2.50	-0.80	-2.29
Central	-1.83	-2.36	-2.11	-2.79	-2.32	0.0	-2.48	-0.52	-2.83
Saba	-2.60	-3.07	-3.38	-3.68	-2.50	-2.48	-0.65	-0.32	-2.46
W	-0.77	-0.77	-1.19	-1.61	-0.80	-0.52	-0.32	2.33	-0.17
S	-2.30	-3.12	-3.15	-3.47	-2.29	-2.83	-2.46	-0.17	0.12

A Simple "Gravity" Model of Communication between pairs of DSDs

- Natural to assume:
  - Volume of calls <u>increasing in population</u> (at both ends)
  - Volume of calls <u>decreasing in distance</u> between DSDs

- Suggests a simple formula:

$$\textbf{Volume}\ (DSD_1, DSD_2) \propto \frac{\text{Pop}\ (DSD_1) \times \text{Pop}(DSD_2)}{\text{Distance}(DSD_1, DSD_2)}$$

- Similar formulas used extensively to understand transportation and trade flows.

LIRNE*asia*
Pro-poor. Pro-market.

Formula Validated by Regression Analysis

- The formula explains 42% of the variance in the data.
- Coefficients have the expected sign and *magnitude.*

	(1)	(2)
	Log(Total Volume of Calls)	
Log(DSD population)	1.029***	0.962***
	(0.00760)	(0.00646)
Log(Distance between DSDs)	-0.945***	-1.169***
	(0.00852)	(0.00703)
District fraction with internet access		0.0538***
		(0.000436)
Observations	106,276	106,276
R-squared	0.423	0.535

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Visualizing the Volume of Calls, Again

- Plot the volume of calls that remains <u>unexplained</u> after we account for population and distance.

-1	-0.75	-0.5	-0.25	0	0.25	0.5	0.75	1
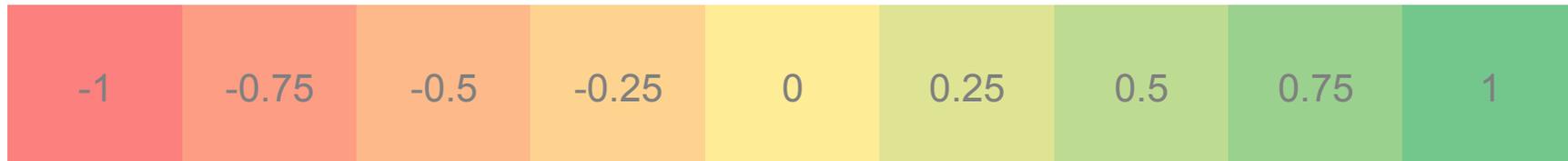
Few Calls   →   Many Calls

- Once population and distance are accounted for, the North (and East) emerge as top callers.
- Sabaragamuwa and Uva are the main laggards.

	N	NW	NC	E	Uva	Central	Saba	W	S
N	4.39	1.92	1.84	2.89	2.42	2.19	1.57	2.48	2.54
NW	1.92	0.45	0.04	0.18	-0.61	-0.78	-1.26	0.03	-0.17
NC	1.84	0.04	0.60	0.40	-0.20	0.05	-0.72	0.66	0.37
E	2.88	0.18	0.40	1.95	0.22	0.07	-0.51	0.93	0.35
Uva	2.42	-0.61	-0.20	0.22	0.73	-0.69	-0.83	0.56	0.09
Central	2.19	-0.78	0.058	0.07	-0.69	0.0	-1.50	0.05	-0.54
Saba	1.58	-1.26	-0.72	-0.51	-0.83	-1.50	-0.64	-0.49	-1.03
W	2.48	0.03	0.66	0.93	0.56	0.05	-0.49	0.17	0.59
S	2.54	-0.17	0.37	0.35	0.09	-0.54	-1.03	0.59	0.77

Data Sources

- Mobile phone data:
  - Voice calls over a period of 4 months
- Population and (district-level) internet access:
  - 2011 Census
- Distance :
  - Direct (geodesic) distance between DSD centroids

Methods

Main measure:

- Log(number of calls between two regions)
  - A call between two users is counted towards calls between the "home" DSDs of the two users.
  - Results using total duration of calls are similar.

- Scale of analysis:
  - DSDs: 326^2 = 106,276 pairs → regression analysis
  - Provinces: 9^2 = 81 pairs → graphical analysis