SEGREGATION OF SYRIAN REFUGEES IN TURKEY:

EVIDENCE FROM MOBILE PHONE DATA

** PRELIMINARY – DO NOT CITE, COPY WITHOUT PERMISSION **

2nd LSE Workshop on Political Economy of Turkey

LSE, September 11, 2020

Simone Bertoli, Çağlar Özden & Michael Packard

*The usual World Bank disclaimer applies. All comments, results are the opinions of the presenter, not of the Board, senior management or the member countries.

QUESTION

- How can we use the spatial and temporal distribution of Syrian Refugees in Turkey?
- Relevant for the analysis of:
  - Labor markets
  - Product markets
  - Real estate markets
  - Political markets

QUESTION

CHALLENGE:

- Absence of high quality/frequency data on the distribution of refugees
- Existing data – too aggregated both temporally and spatially

PHONE DATA

- As part of Data for Refugees (D4R) challenge – managed by TUBITAK & Bosphorus University
- Based on Call Detail Records (CDR) of incoming and outgoing calls of Turk Telekom (25% market share)
- CDR: caller ID, receiver ID, time/date, tower location
- ~ 1 million customers, 185k tagged as refugees.
- Refugees receive lower rates and need to show ID cards
- Multiple data files. We use mainly datafile #1

PHONE DATA

- Datafile #1 includes all refugees and sample of Turkish customers, chosen from the same spatial distribution at the province level .
- Number of native and refugee calls for each tower (30k+) for each hour (8760 hours) in 2017.
- We use the call volume as a proxy for refugee and native populations' temporal and spatial distributions

DEFINITION OF IDENTITY


Consulat Impérial Ottoman à Calafat

N° G[l] 239

N° S[l] 149

39

Au Nom de Sa Majesté Impériale le Sultan.

Signalement

Age de 27 ans

Taille moyenne

Cheveux châtains

Yeux idem

Nez moyen

Bouche idem

Moustaches blondes

Barbe rasée

Menton ovale

Visage idem

Teint brun

Religion musulmane

Signes particuliers

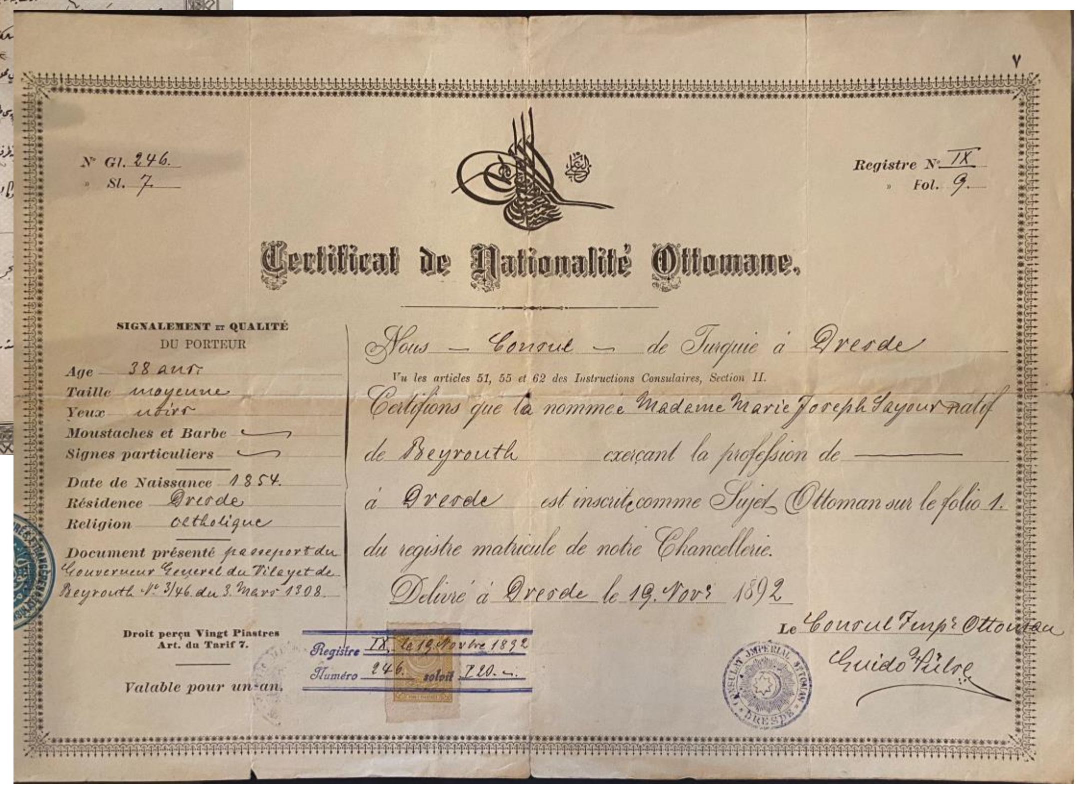
Le présent passeport est valable pour un an.

Droit perçu piastres cinquante

Nous Vice Consul de Turquie à Calafat (Roumanie) prions Messieurs les officiers civils et militaires chargés de maintenir l'ordre public dans tous les pays amis et alliés de l'Empire Ottoman de laisser passer librement le sujet Ottoman Moukhtar Ramaz[an] garde-champêtre né à Krountcha domicilié à idem se rendant à son pays et de lui donner aide et protection en cas de besoin.

Délivré le 16/28 Août 1899

Le Vice-Consul


N° Gl. 246.

" Sl. 7.

Registre N° IX

" Fol. 9.

Certificat de Nationalité Ottomane.

SIGNALEMENT ET QUALITÉ DU PORTEUR

Age 38 ans

Taille moyenne

Yeux noirs

Moustaches et Barbe —

Signes particuliers —

Date de Naissance 1854.

Résidence Dresde

Religion catholique

Document présenté passeport du Gouverneur Général du Vilayet de Beyrouth N° 3/46 du 3 mars 1308.

Droit perçu Vingt Piastres Art. du Tarif 7.

Valable pour un an.

Nous Consul de Turquie à Dresde

Vu les articles 51, 55 et 62 des Instructions Consulaires, Section II.

Certifions que la nommée Madame Marie Joseph Sayour natif de Beyrouth exerçant la profession de — à Dresde est inscrit comme Sujet Ottoman sur le folio 1. du registre matricule de notre Chancellerie.

Délivré à Dresde le 19. Nov[bre] 1892

Le Consul Imp[al] Ottoman

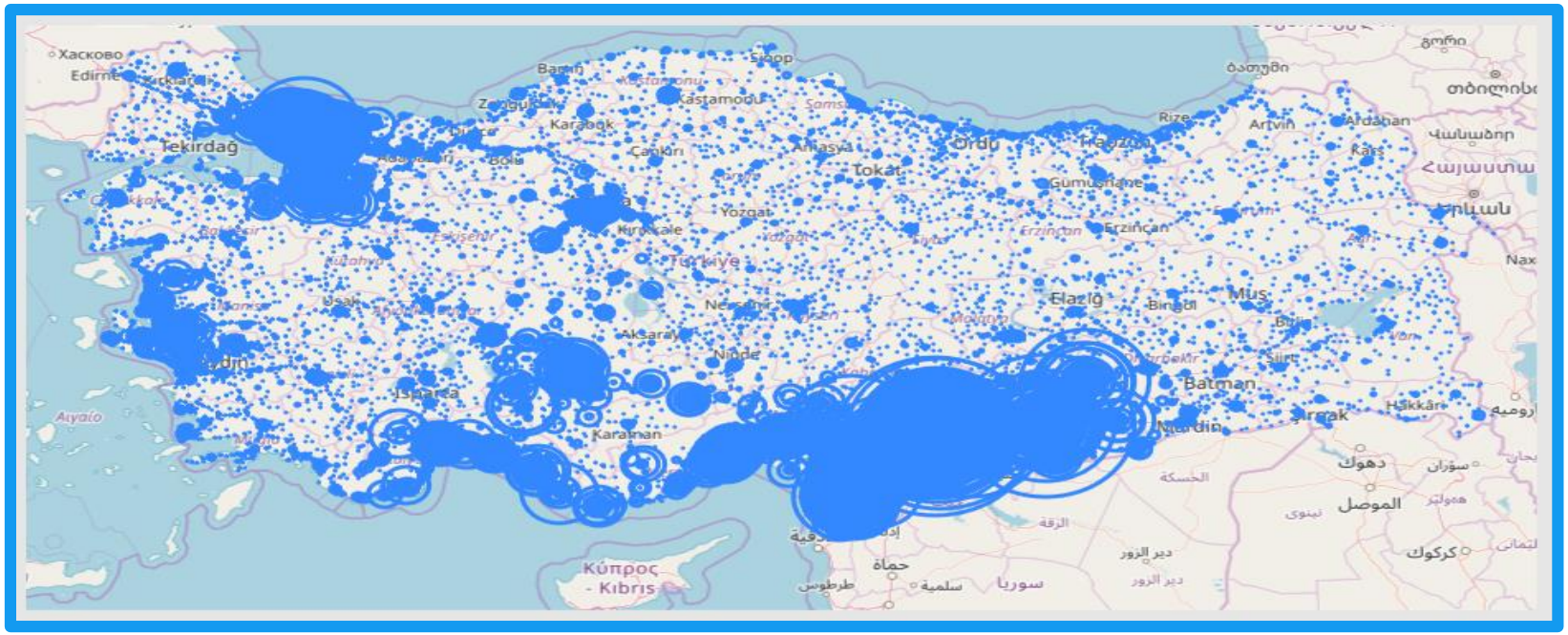
Guido Pilz

Registre IX le 19 Novbre 1892

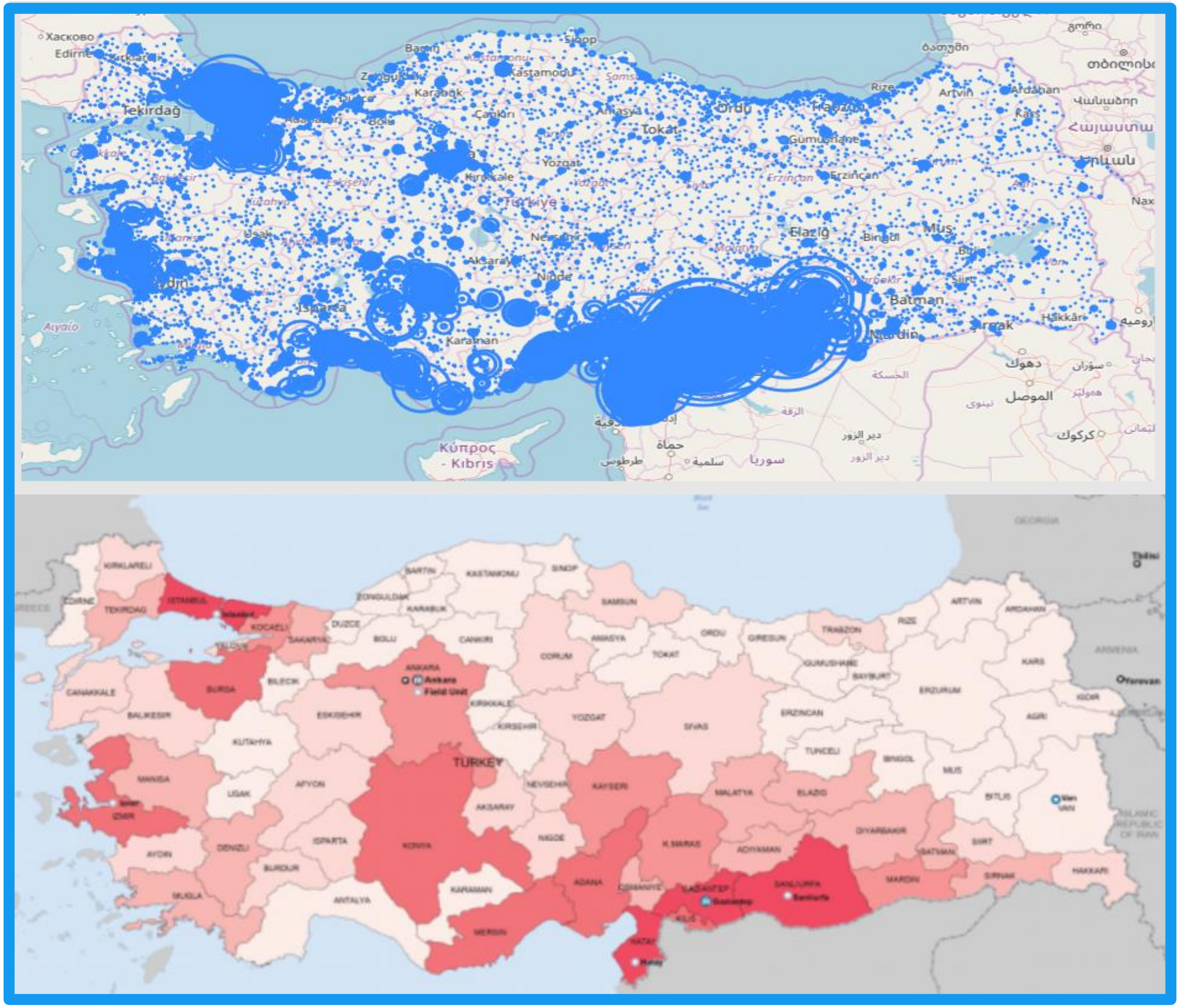
Numéro 246. soloit 120.

Syrian Refugees Distribution Across Provinces - UNHCR

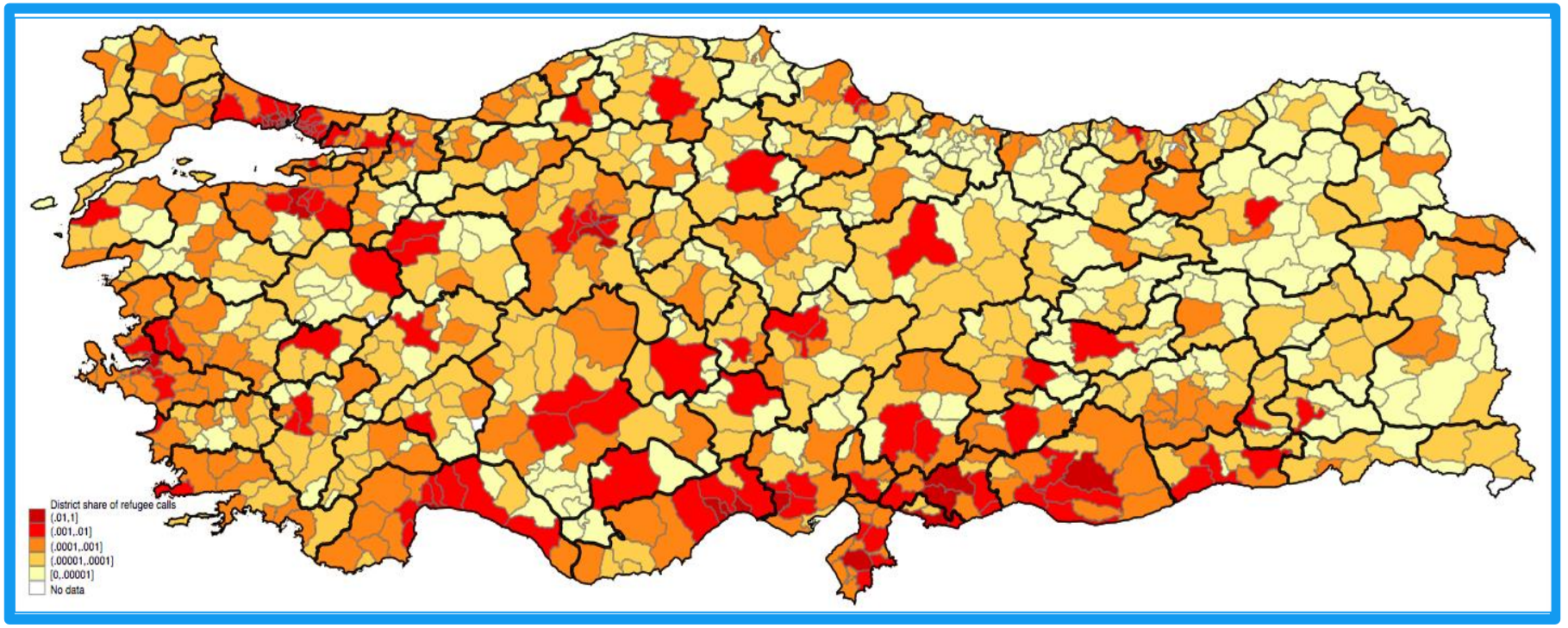
Syrian Refugees Cellular Phone Call Volumes



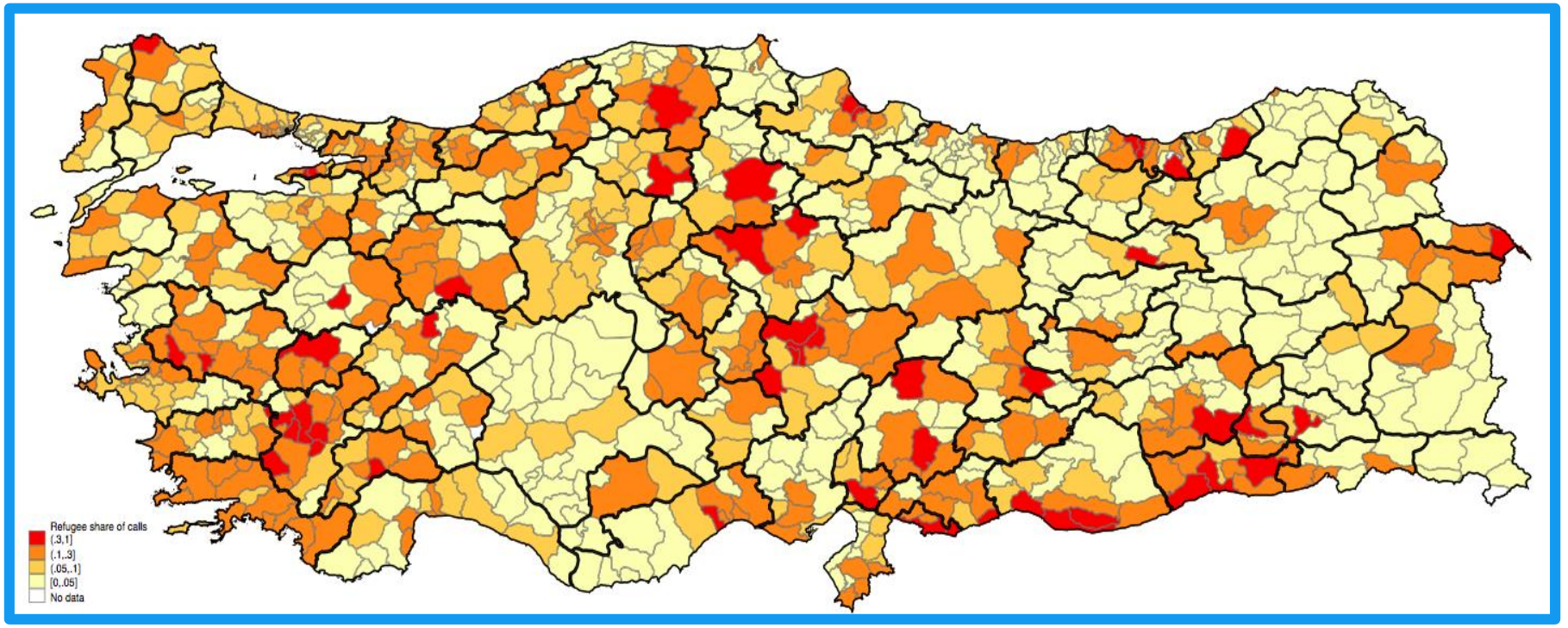
Phone Data versus Administrative Data



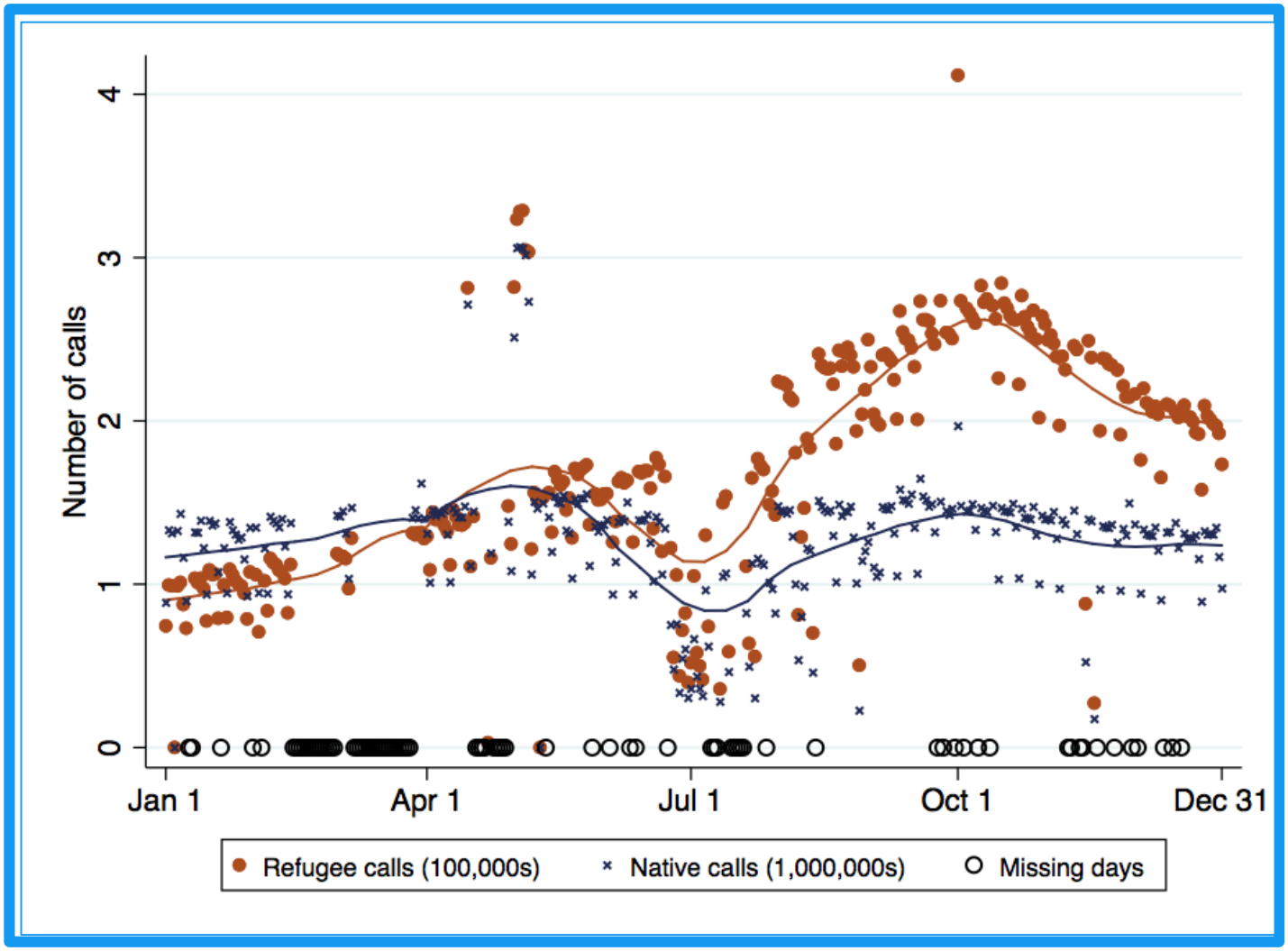
DISTRIBUTION OF REFUGEE CALLS ACROSS THE DISTRICTS



REFUGEE SHARE OF CALLS IN EACH DISTRICT



CALL VOLUME - OVER TIME



OBJECTIVE

- Identify the patterns of distribution – so can use it for impact analysis
- Main Tools from the Vast Sociology Literature
- Several Segregation indices:
  - Dissimilarity Index
  - (adjusted) Assimilation Index

SEGREGATION INDICES

Douglas S. Massey

FOLLOW

Professor of Sociology and Public Affairs, Princeton University
Verified email at princeton.edu - Homepage

Sociology Demography

TITLE	CITED BY	YEAR
American apartheid: Segregation and the making of the underclass DS Massey, NA Denton Harvard University Press	12202	1993
Theories of international migration: A review and appraisal DS Massey, J Arango, G Hugo, A Kouaouci, A Pellegrino, JE Taylor Population and development review 19 (3), 431-466	6812	1993
The dimensions of residential segregation DS Massey, NA Denton Social forces 67 (2), 281-315	3412	1988

SEGREGATION INDICES

George Borjas

Harvard Kennedy School
Verified email at harvard.edu - Homepage

Labor Economics Economics

FOLLOW

TITLE	CITED BY	YEAR
The Economics of Immigration GJ Borjas Journal of Economic Literature 32 (4), 1667-1717	3938	1994
Self-selection and the earnings of immigrants GJ Borjas American Economic Review 77 (4), 531-55	3671	1987
The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market GJ Borjas Quarterly Journal of Economics 118 (4), 1335-1374	2646	2003

DISSIMILARITY INDEX

$$D = \frac{1}{2(1-P)} \sum_{i=1}^{K} \left| \frac{m_i}{M} - \frac{t_i}{T} \right|$$

$$= \frac{1}{2} \sum_{i=1}^{K} \left| \frac{m_i}{M} - \frac{n_i}{N} \right|$$

K – number of areas

m_i – minority population in areas i

n_i – non-minority population in areas i

t_i – total population in areas i

M – total minority population

N – total non-minority (native) population

T – total population

DISSIMILARITY INDEX

- $D \in [0,1]$,

  D=0 (1) is complete integration (segregation)

- D : share of the minority (refugee) population that needs to be relocated from high to low concentration areas to match their average distribution across the country

- D : insensitive to an identical proportional increase in the size of the minority group across all areas – since "m_i/M" stays constant.

ADJUSTED ISOLATION INDEX

$$I \equiv \sum_{i=1}^{K} \left(\frac{m_i}{t_i} \frac{m_i}{M} \right) = \sum_{i=1}^{K} \left(p_i \frac{m_i}{M} \right)$$

$$I^{\text{adj}} \equiv \frac{I - P}{\min\left\{1, M/(\min_i t_i)\right\} - P}$$

m_i – minority population in areas i

t_i – total population in areas i

p_i – share of minority population in areas i

M – total minority population

T – total population

P – share of minority population in the total

ADJUSTED ISOLATION INDEX

- $I \in [0,1]$
- Regular isolation index: sensitive to proportional increase in the size of the minority group across the regions.
- So construct the adjusted isolation index to reduce, not eliminate, its dependency on P.
- A uniform increase in the share of the refugees across all districts will increase the adjusted isolation index, I.

EMPIRICAL CHALLENGES

- Objective: using the phone call distribution as a proxy for population distribution
- Problem #1: Different call propensities for Turkish and refugee customers
- Problem #2: Differences in market shares of Turkish Telekom within refugee population across provinces.
- Problem #3: Different degrees of geographic partitioning creates difference in indices

PROBLEM #1 : DIFFERENT CALL PROPENSITIES BETWEEN GROUPS

$$D = \frac{1}{2(1-P)} \sum_{i=1}^{K} \left| \frac{m_i}{M} - \frac{t_i}{T} \right|$$

$$= \frac{1}{2} \sum_{i=1}^{K} \left| \frac{m_i}{M} - \frac{n_i}{N} \right|$$

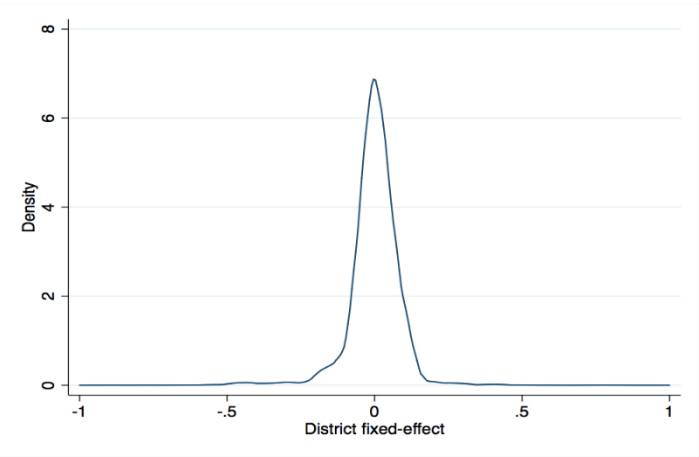
- No impact on Dissimilarity Index D if identical across geographic areas
- … but, isolation index, I, becomes bigger if the refugees have a higher propensity (since p_i will be bigger)

$$I \equiv \sum_{i=1}^{K} \left(\frac{m_i}{t_i} \frac{m_i}{M} \right) = \sum_{i=1}^{K} \left(p_i \frac{m_i}{M} \right)$$

$$I^{\text{adj}} \equiv \frac{I - P}{\min\left\{1, M/(\min_i t_i)\right\} - P}$$

PROBLEM #1 : DIFFERENT CALL PROPENSITIES BETWEEN GROUPS

- Use D4R #2 to see if propensity is different across provinces
- Smaller sample of individuals, identify time and location of ALL of their call throughout the year
- Calculate daily call volume of each individual, assign to a district for each day, regress the call volume on date and district fixed effects.



» These have mean zero with small standard deviation. 90% of observation are within 10 % of the average district. The rest are small sample districts.

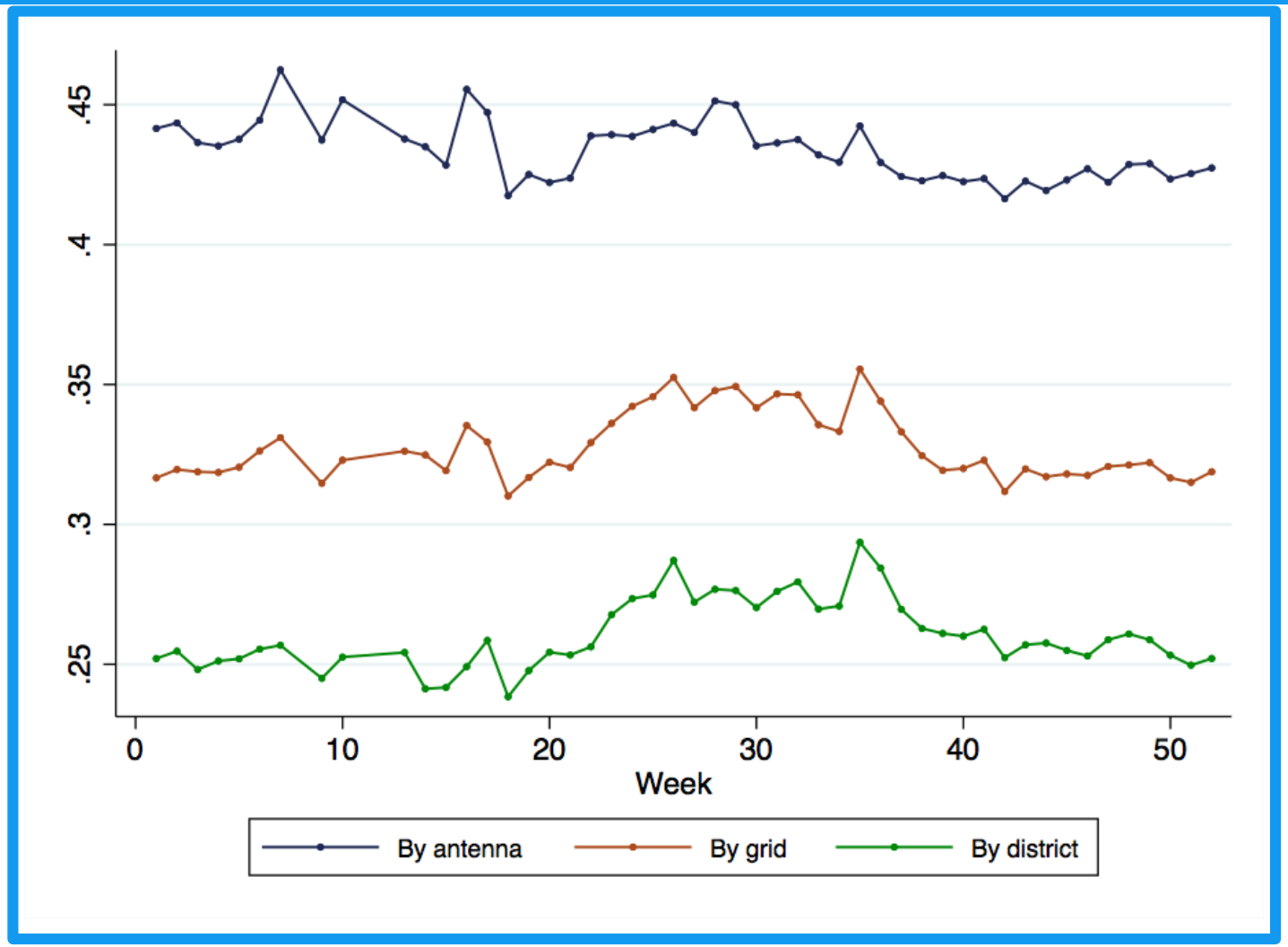
PROBLEM #2 : DIFFERENT MARKET SHARES

- Different market shares across provinces!
- Take the official statistics from Refugee Administration at the province level for January 1 and July 1, 2017.
- Use these numbers as the weights to calculate the national indices, D and I.

PROBLEM #3: GEOGRAPHIC PARTITIONING

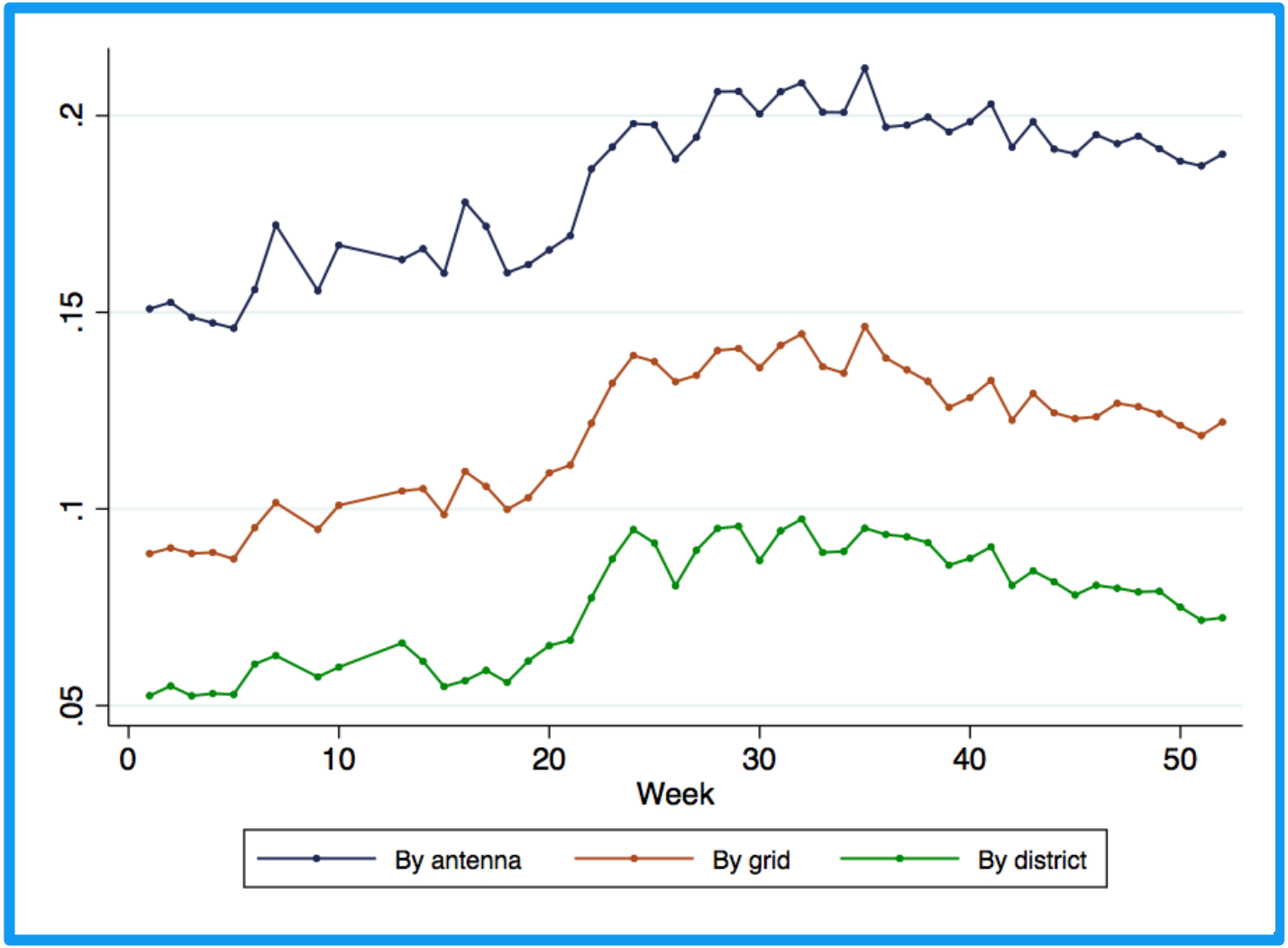
- 82 Provinces
- 957 districts
- Urban areas – at least 15 cells within 5 km radius
- Catchment areas of the 30,000+ towers

DISSIMILARITY INDEX
EVOLUTION OVER TIME FOR THE WHOLE COUNTRY

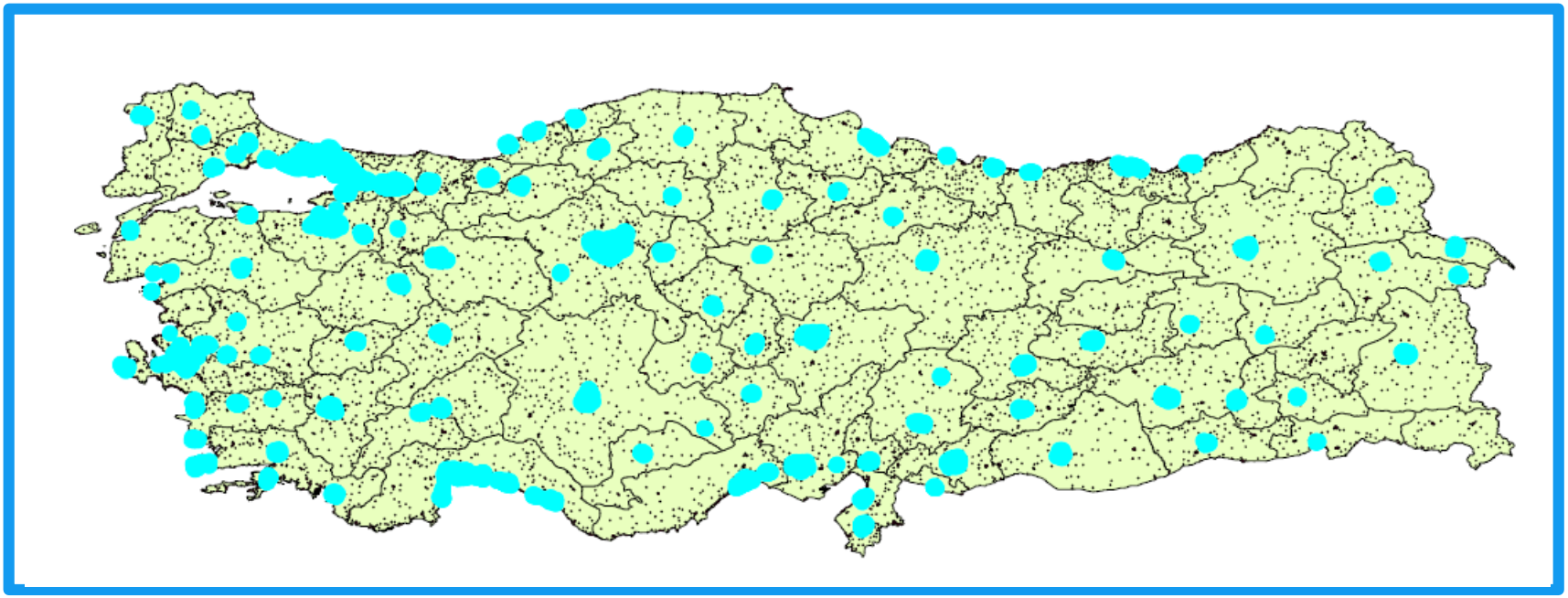


Note Weeks #26 and 35!

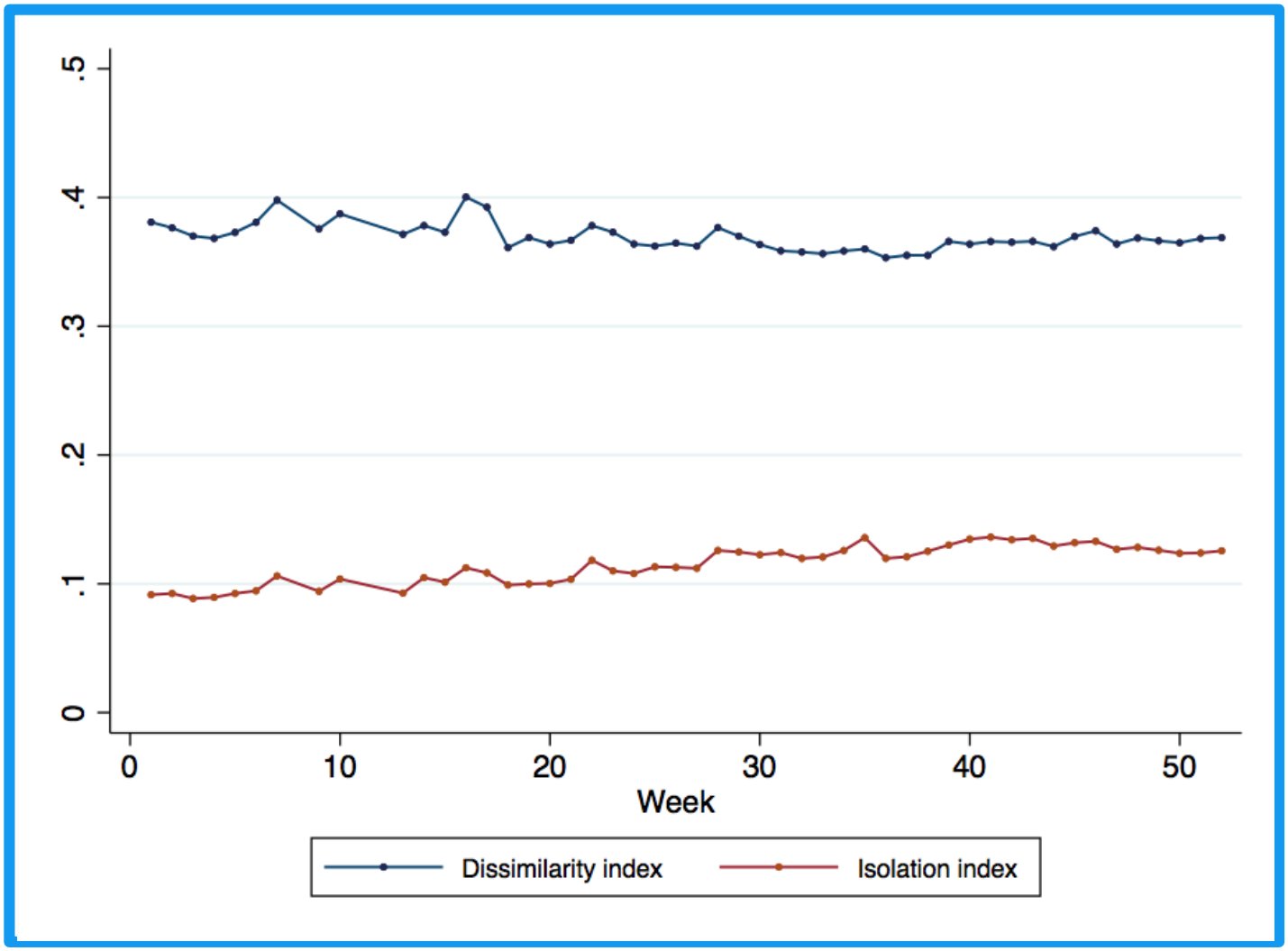
ISOLATION INDEX
EVOLUTION OVER TIME FOR THE WHOLE COUNTRY



URBAN AREAS



DISSIMILARITY AND ISOLATION INDICES
URBAN AREAS



DETERMINANTS OF MOBILITY - NATIVES

Table 1: Gravity Estimates for Natives (Dataset 3)

	(1)	(2)	(3)	(4)	(5)	(6)
$Contiguity_{jk}$	0.672	0.635	0.623	0.624	0.615	0.611
	(0.144)***	(0.138)***	(0.138)***	(0.139)***	(0.141)***	(0.141)***
$Ln(Distance_{jk})$	-0.368	-0.307	-0.292	-0.289	-0.290	-0.294
	(0.073)***	(0.077)***	(0.076)***	(0.076)***	(0.078)***	(0.078)***
$Ln(Population_k)$		0.956	0.920	0.921	0.916	0.914
		(0.040)***	(0.040)***	(0.040)***	(0.041)***	(0.041)***
$Ln(GDP/capita_k)$		0.418	0.544	0.527	0.513	0.533
		(0.095)***	(0.099)***	(0.106)***	(0.104)***	(0.107)***
$Refugeeshare_k$			2.083	1.963	2.032	2.220
			(0.731)***	(0.828)**	(0.787)***	(0.855)***
$Dissimilarityindex_k$				-0.238		0.369
				(0.647)		(0.824)
$Isolationindex_k$					-0.839	-1.033
					(0.626)	(0.800)
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	No	No	No	No	No
Pseudo-R^2	0.59	0.58	0.58	0.58	0.58	0.58
Origin-destination pairs	2,400	2,400	2,400	2,400	2,400	2,400
Aggregate sample	13,142	13,142	13,142	13,142	13,142	13,142

Table shows results from Poisson pseudo-maximum likelihood (PPML) regressions of cross-province Non-refugee migration flows calculated from dataset 3. Population and GDP variables are calculated for 2013. Province level refugee shares are taken from Turkish administrative data and correspond to stocks as

DETERMINANTS OF MOBILITY - REFUGEES

Table 2: Gravity Estimates for Refugees (Dataset 3)

	(1)	(2)	(3)	(4)	(5)	(6)
$Contiguity_{jk}$	0.428	0.447	0.422	0.381	0.417	0.408
	(0.154)***	(0.157)***	(0.152)***	(0.151)**	(0.156)***	(0.160)**
$Ln(Distance_{jk})$	-0.295	-0.401	-0.275	-0.266	-0.275	-0.260
	(0.064)***	(0.092)***	(0.069)***	(0.067)***	(0.069)***	(0.069)***
$Ln(Population_k)$		1.164	1.065	1.042	1.062	1.061
		(0.054)***	(0.044)***	(0.045)***	(0.043)***	(0.040)***
$Ln(GDP/capita_k)$		0.066	0.644	0.582	0.639	0.591
		(0.115)	(0.112)***	(0.110)***	(0.115)***	(0.114)***
$Refugeeshare_k$			7.905	7.719	7.949	7.033
			(0.365)***	(0.373)***	(0.409)***	(0.499)***
$Dissimilarityindex_k$				-2.187		-4.650
				(0.536)***		(1.380)***
$Isolationindex_k$					-0.214	2.966
					(0.701)	(1.440)**
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	No	No	No	No	No
Pseudo-R^2	0.66	0.58	0.63	0.64	0.63	0.64
Origin-destination pairs	2,400	2,400	2,400	2,400	2,400	2,400
Aggregate sample	15,497	15,497	15,497	15,497	15,497	15,497

Table shows results from Poisson pseudo-maximum likelihood (PPML) regressions of cross-province Refugee migration flows calculated from dataset 3. Population and GDP variables are calculated for 2013. Province level refugee shares are taken from Turkish administrative data and correspond to stocks as of

DETERMINANTS OF MOBILITY - NATIVES

Table 1: Gravity Estimates for Natives (Dataset 3)

	(1)	(2)	(3)	(4)	(5)	(6)
$Contiguity_{jk}$	0.672	0.635	0.623	0.624	0.615	0.611
	(0.144)***	(0.138)***	(0.138)***	(0.139)***	(0.141)***	(0.141)***
$Ln(Distance_{jk})$	-0.368	-0.307	-0.292	-0.289	-0.290	-0.294
	(0.073)***	(0.077)***	(0.076)***	(0.076)***	(0.078)***	(0.078)***
$Ln(Population_k)$		0.956	0.920	0.921	0.916	0.914
		(0.040)***	(0.040)***	(0.040)***	(0.041)***	(0.041)***
$Ln(GDP/capita_k)$		0.418	0.544	0.527	0.513	0.533
		(0.095)***	(0.099)***	(0.106)***	(0.104)***	(0.107)***
$Refugeeshare_k$			2.083	1.963	2.032	2.220
			(0.731)***	(0.828)**	(0.787)***	(0.855)***
$Dissimilarityindex_k$				-0.238		0.369
				(0.647)		(0.824)
$Isolationindex_k$					-0.839	-1.033
					(0.626)	(0.800)
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	No	No	No	No	No
Pseudo-R^2	0.59	0.58	0.58	0.58	0.58	0.58
Origin-destination pairs	2,400	2,400	2,400	2,400	2,400	2,400
Aggregate sample	13,142	13,142	13,142	13,142	13,142	13,142

Table shows results from Poisson pseudo-maximum likelihood (PPML) regressions of cross-province Non-refugee migration flows calculated from dataset 3. Population and GDP variables are calculated for 2013. Province level refugee shares are taken from Turkish administrative data and correspond to stocks as

DETERMINANTS OF MOBILITY - REFUGEES

Table 2: Gravity Estimates for Refugees (Dataset 3)

	(1)	(2)	(3)	(4)	(5)	(6)
$Contiguity_{jk}$	0.428	0.447	0.422	0.381	0.417	0.408
	(0.154)***	(0.157)***	(0.152)***	(0.151)**	(0.156)***	(0.160)**
$Ln(Distance_{jk})$	-0.295	-0.401	-0.275	-0.266	-0.275	-0.260
	(0.064)***	(0.092)***	(0.069)***	(0.067)***	(0.069)***	(0.069)***
$Ln(Population_k)$		1.164	1.065	1.042	1.062	1.061
		(0.054)***	(0.044)***	(0.045)***	(0.043)***	(0.040)***
$Ln(GDP/capita_k)$		0.066	0.644	0.582	0.639	0.591
		(0.115)	(0.112)***	(0.110)***	(0.115)***	(0.114)***
$Refugeeshare_k$			7.905	7.719	7.949	7.033
			(0.365)***	(0.373)***	(0.409)***	(0.499)***
$Dissimilarityindex_k$				-2.187		-4.650
				(0.536)***		(1.380)***
$Isolationindex_k$					-0.214	2.966
					(0.701)	(1.440)**
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	No	No	No	No	No
Pseudo-R^2	0.66	0.58	0.63	0.64	0.63	0.64
Origin-destination pairs	2,400	2,400	2,400	2,400	2,400	2,400
Aggregate sample	15,497	15,497	15,497	15,497	15,497	15,497

Table shows results from Poisson pseudo-maximum likelihood (PPML) regressions of cross-province Refugee migration flows calculated from dataset 3. Population and GDP variables are calculated for 2013. Province level refugee shares are taken from Turkish administrative data and correspond to stocks as of

DETERMINANTS OF MOBILITY - NATIVES

Table 1: Gravity Estimates for Natives (Dataset 3)

	(1)	(2)	(3)	(4)	(5)	(6)
$Contiguity_{jk}$	0.672	0.635	0.623	0.624	0.615	0.611
	(0.144)***	(0.138)***	(0.138)***	(0.139)***	(0.141)***	(0.141)***
$Ln(Distance_{jk})$	-0.368	-0.307	-0.292	-0.289	-0.290	-0.294
	(0.073)***	(0.077)***	(0.076)***	(0.076)***	(0.078)***	(0.078)***
$Ln(Population_k)$		0.956	0.920	0.921	0.916	0.914
		(0.040)***	(0.040)***	(0.040)***	(0.041)***	(0.041)***
$Ln(GDP/capita_k)$		0.418	0.544	0.527	0.513	0.533
		(0.095)***	(0.099)***	(0.106)***	(0.104)***	(0.107)***
$Refugeeshare_k$			2.083	1.963	2.032	2.220
			(0.731)***	(0.828)**	(0.787)***	(0.855)***
$Dissimilarityindex_k$				-0.238		0.369
				(0.647)		(0.824)
$Isolationindex_k$					-0.839	-1.033
					(0.626)	(0.800)
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	No	No	No	No	No
Pseudo-R^2	0.59	0.58	0.58	0.58	0.58	0.58
Origin-destination pairs	2,400	2,400	2,400	2,400	2,400	2,400
Aggregate sample	13,142	13,142	13,142	13,142	13,142	13,142

Table shows results from Poisson pseudo-maximum likelihood (PPML) regressions of cross-province Non-refugee migration flows calculated from dataset 3. Population and GDP variables are calculated for 2013. Province level refugee shares are taken from Turkish administrative data and correspond to stocks as

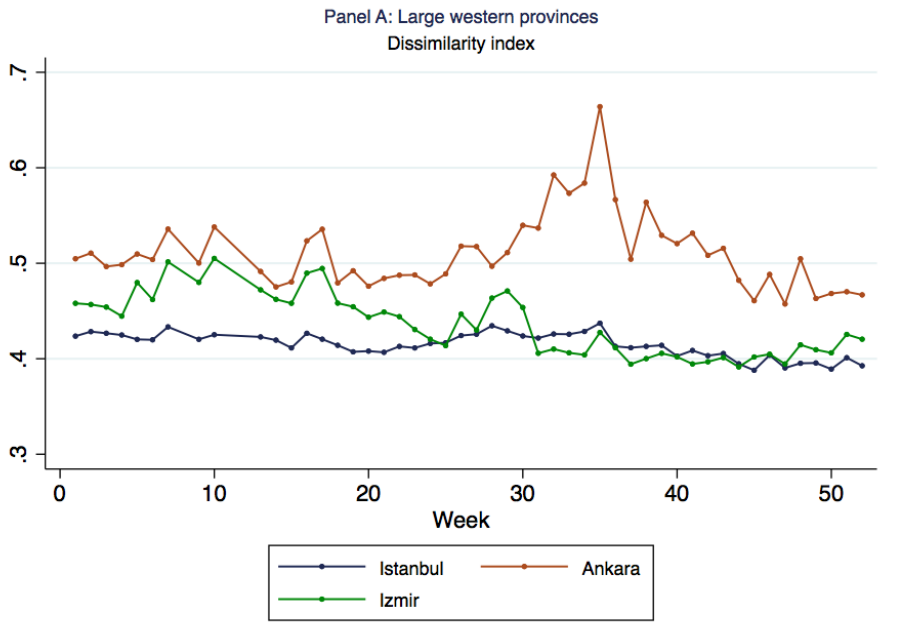
DETERMINANTS OF MOBILITY - REFUGEES

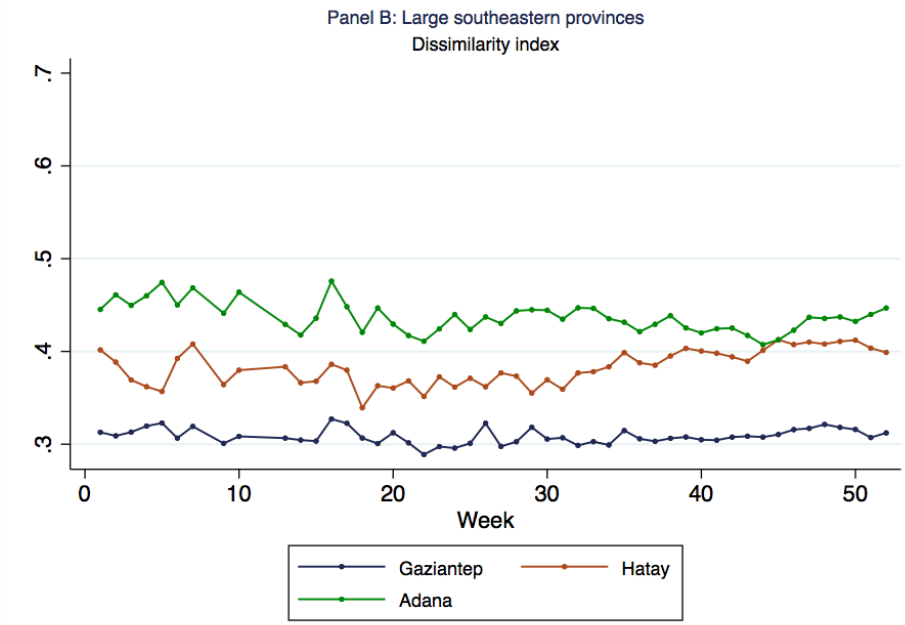
Table 2: Gravity Estimates for Refugees (Dataset 3)

	(1)	(2)	(3)	(4)	(5)	(6)
$Contiguity_{jk}$	0.428	0.447	0.422	0.381	0.417	0.408
	(0.154)***	(0.157)***	(0.152)***	(0.151)**	(0.156)***	(0.160)**
$Ln(Distance_{jk})$	-0.295	-0.401	-0.275	-0.266	-0.275	-0.260
	(0.064)***	(0.092)***	(0.069)***	(0.067)***	(0.069)***	(0.069)***
$Ln(Population_k)$		1.164	1.065	1.042	1.062	1.061
		(0.054)***	(0.044)***	(0.045)***	(0.043)***	(0.040)***
$Ln(GDP/capita_k)$		0.066	0.644	0.582	0.639	0.591
		(0.115)	(0.112)***	(0.110)***	(0.115)***	(0.114)***
$Refugeeshare_k$			7.905	7.719	7.949	7.033
			(0.365)***	(0.373)***	(0.409)***	(0.499)***
$Dissimilarityindex_k$				-2.187		-4.650
				(0.536)***		(1.380)***
$Isolationindex_k$					-0.214	2.966
					(0.701)	(1.440)**
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Dest FE	Yes	No	No	No	No	No
Pseudo-R^2	0.66	0.58	0.63	0.64	0.63	0.64
Origin-destination pairs	2,400	2,400	2,400	2,400	2,400	2,400
Aggregate sample	15,497	15,497	15,497	15,497	15,497	15,497

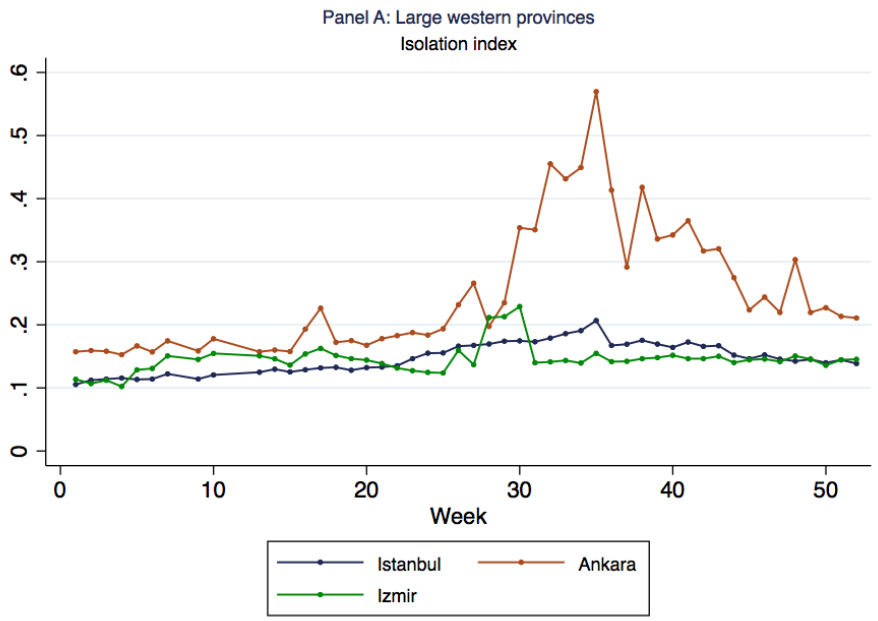
Table shows results from Poisson pseudo-maximum likelihood (PPML) regressions of cross-province Refugee migration flows calculated from dataset 3. Population and GDP variables are calculated for 2013. Province level refugee shares are taken from Turkish administrative data and correspond to stocks as of

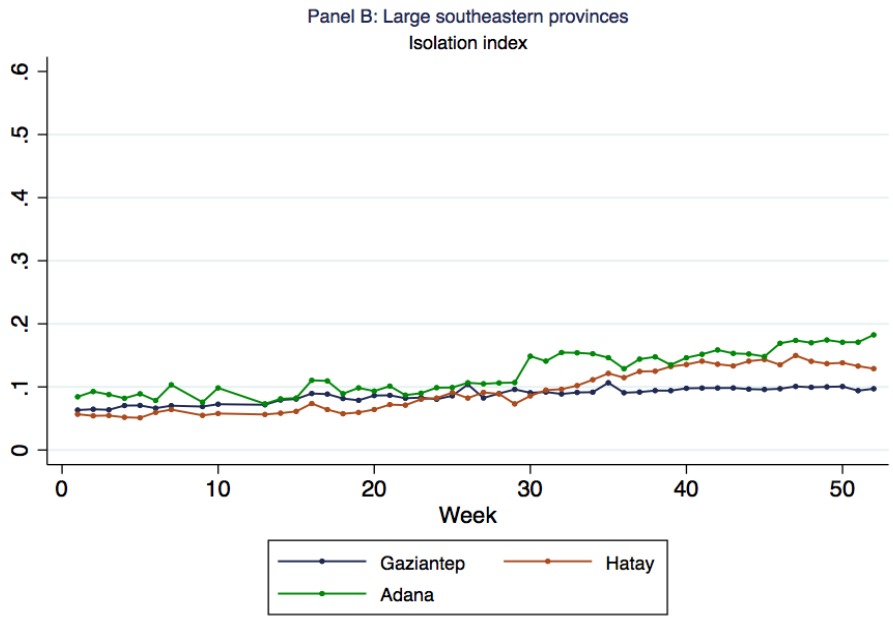
DISSIMILARITY INDEX
EVOLUTION OVER TIME FOR MAJOR PROVINCES



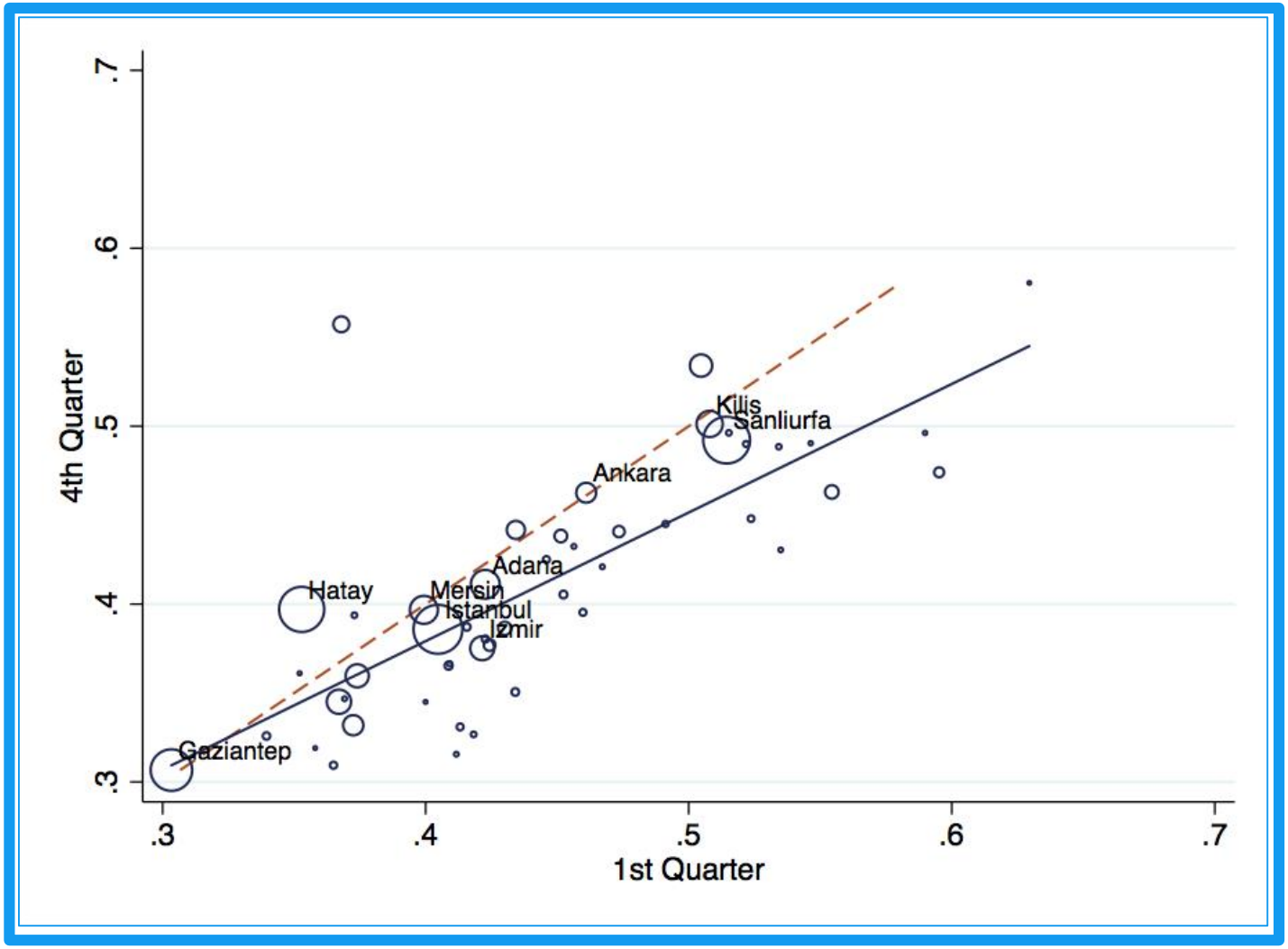


ISOLATION INDEX
EVOLUTION OVER TIME FOR MAJOR PROVINCES

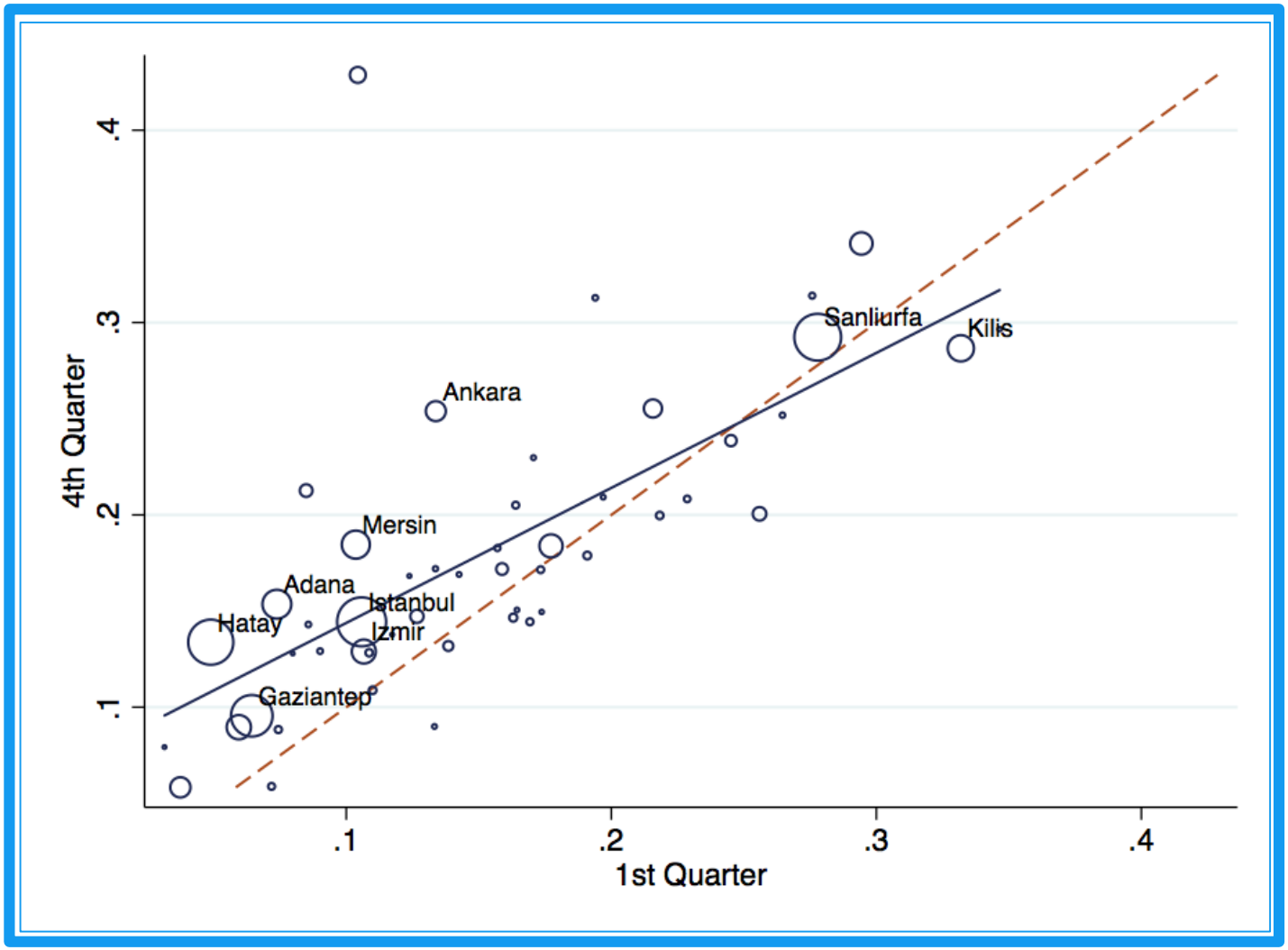




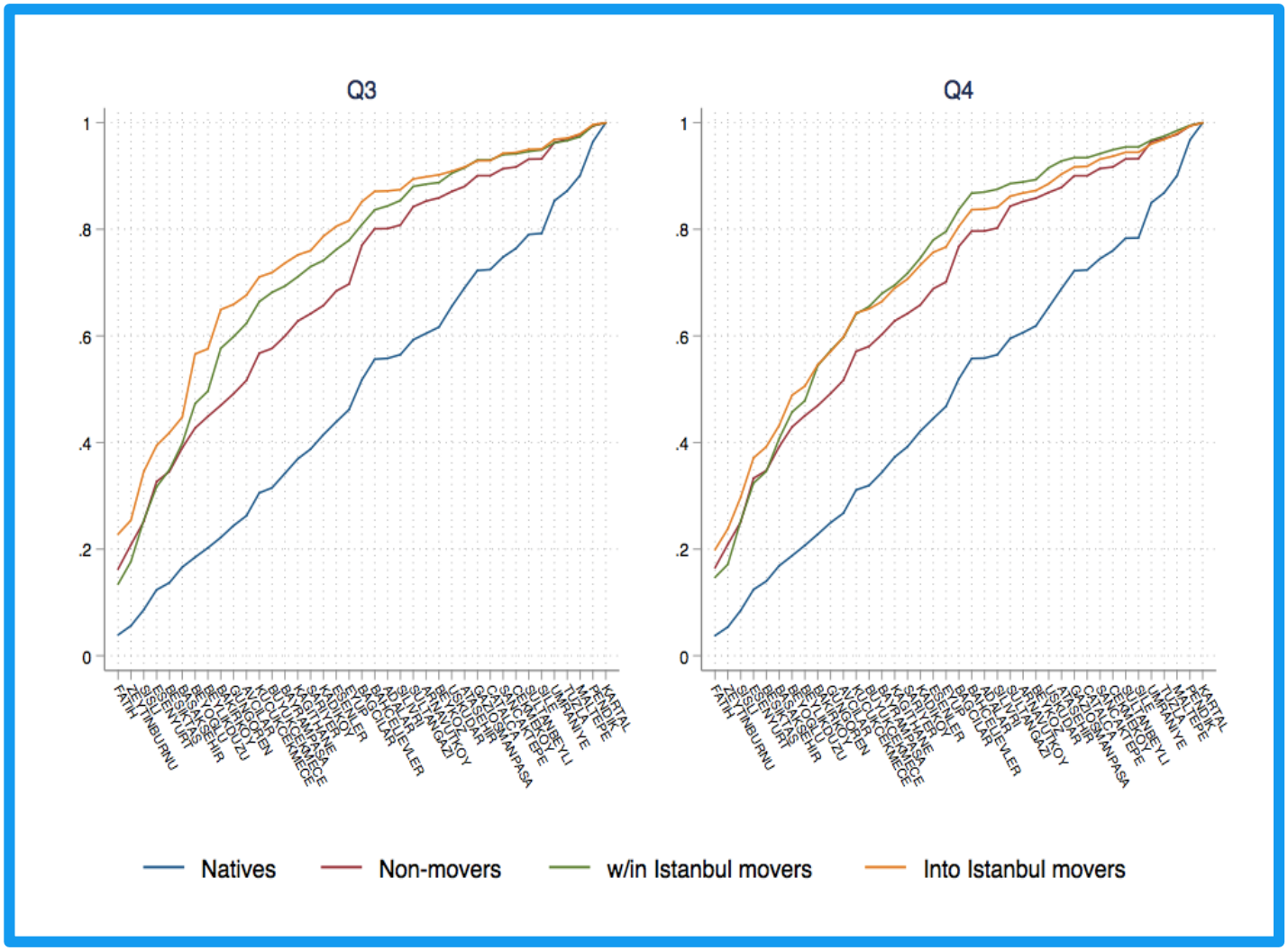
DISSIMILARITY INDEX
FIRST QUARTER VS. FOURTH QUARTER



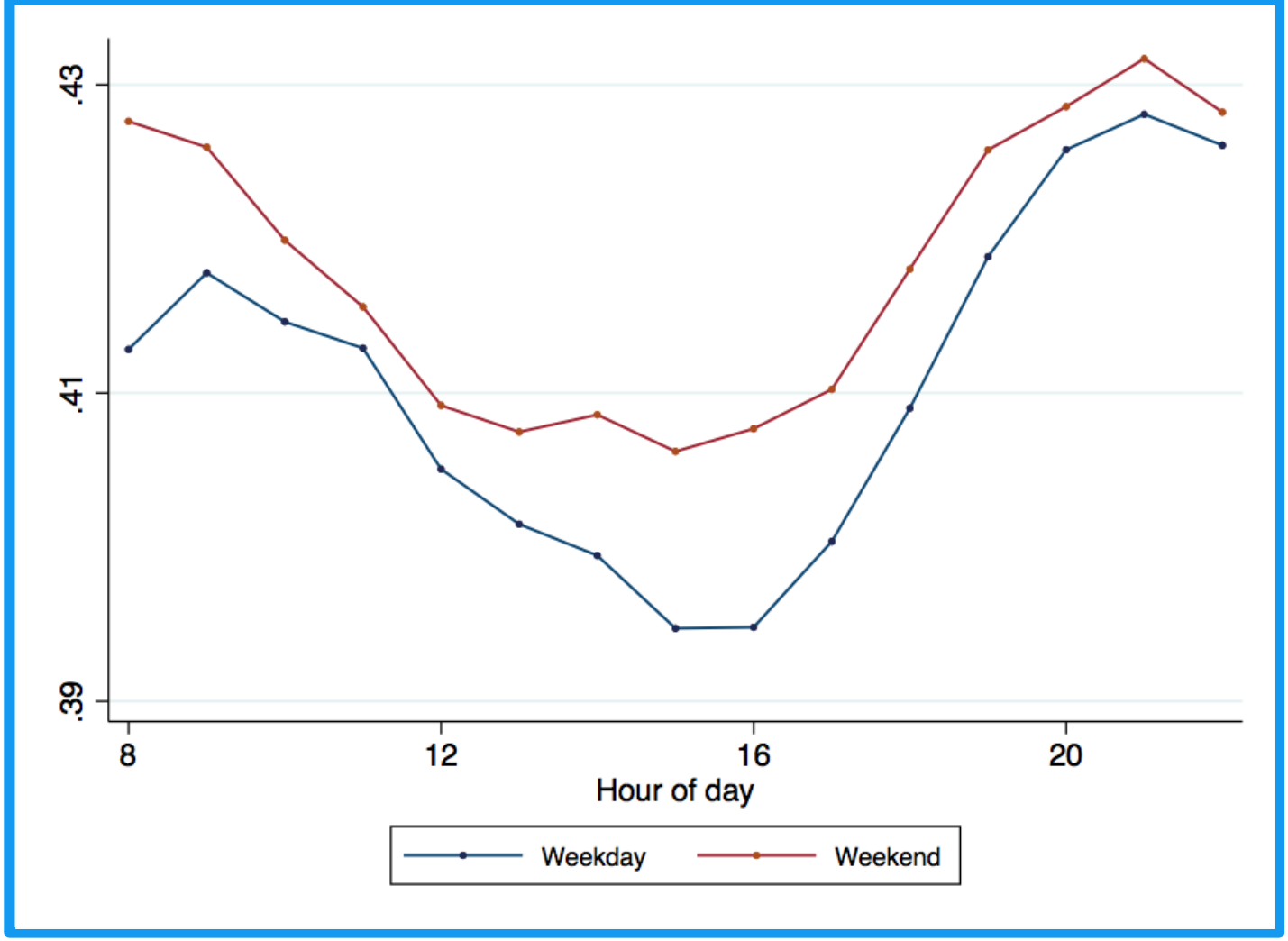
ISOLATION INDEX
FIRST QUARTER VS. FOURTH QUARTER



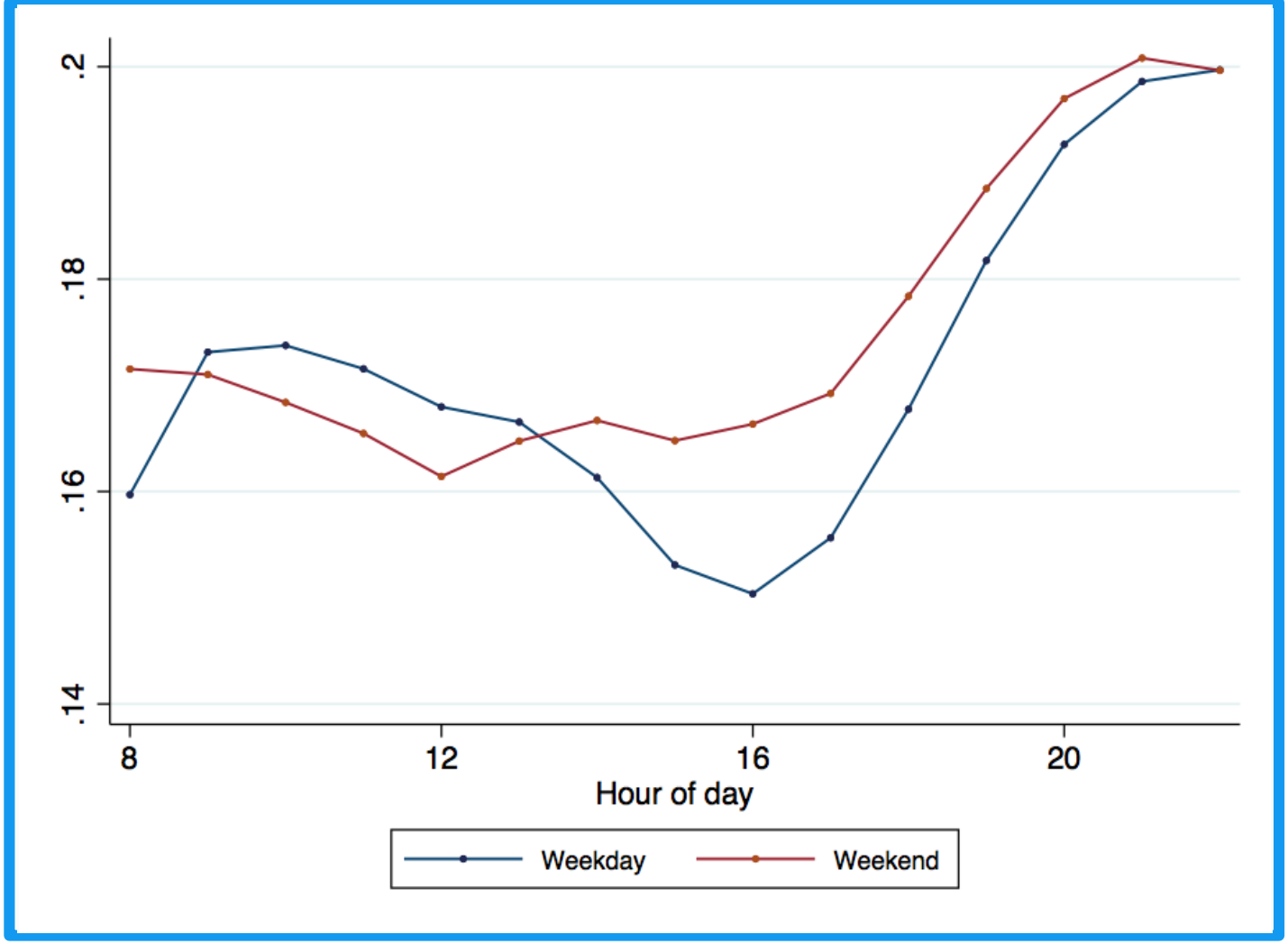
DISTRIBUTION OF REFUGEES AND NATIVES ISTANBUL



DISSIMILARITY INDEX
HOUR OF THE DAY



ISOLATION INDEX
HOUR OF THE DAY



CONCLUSION

- Big data can answer many questions
- Careful in how we use it
  - high frequency does not mean there are no measurement or sampling problems.
- Segregation measures depend on the partitioning
- Different segregation indices measure different things
- Time dimension is critical

CONCLUSION

- There are large differences in provinces
- Segregation is actually lower in larger cities in the west as opposed to border cities
- Segregation is declining over time
- There are large differences between residential and labor market segregation

THANK YOU!

cozden@worldbank.org