What matters for choosing your neighbors? Evidence from Canadian metropolitan areas

Kristian Behrens and Oualid Moussouni

ESG-UQAM; Higher School of Economics; and CEPR
ESG-UQAM

November 8, 2018

Introduction

Introduction: Motivation

- 1^{st} Law of Geography: “Everything is related to everything else, but near things are more related than distant things”. (Tobler, 1970).

- Principal of Homophily: “Similarity breeds connection”. (McPherson et al, 2001).

$\Longrightarrow$ Does similarity breed spatial clusters?

- If the response is yes, then Corollary of the 1^{st} law would be:

  “Near things are more similar than distant things”.

Introduction: Motivation

- Theoretically, the response is: Yes, we want to be close, i.e., colocate with people like us (Preferences in the Schelling models; 1969, 1971).
- Empirically, we don't know.
- Thus, we want to open the black box of preferences:
  - What are these preferences?
  - Which ones are important?
  - Do they shape our location choices, therefore cities?

Introduction: Motivation

Why is it important ?

- Homophily will lead to stratification of cities.
- Amplify the dimension of exposure/isolation of segregation.
- Consequences: negative peer effects (Dawkins et al. 2005, Cutler and Glaeser, 1997).

Thus, understanding the roots of preferences for own type could help design better urban policies that aim at diversity.

Introduction: What we do

- Use spatially fine-grained data for 6 major Canadian metro areas (2016, 2006).
- Look at the role of:
  1. Cultural, Religious, Linguistic similarities.
  2. Politico-historical relationships.
  3. Genetics distances.
- On a measure of pairwise ethnic colocation.

Introduction: Main findings

Conditional on geographic/economic controls:

1. Religious, linguistic, cultural, and genetic proximities have positive effects on colocation patterns.

2. Results are highly robust to a large set of measures.

3. They hold for both the 2016 and 2006 Censuses.

4. There is heterogeneity across cities, but also an east-west gradient for some variables.

Data

Data: outcome variable

- We use Canadian Census data (2016, 2006) at the dissemination area level to construct our measures of colocation.
- Detailed portrait of ethnicities in Canada.
- Census question: “What were the ethnic or cultural origins of this person’s ancestors?”. Two notes:

1. “This question collects information on the ancestral origins of the population and provides information about the composition of Canada’s diverse population”.
2. “An ancestor is usually more distant than a grandparent”.

Table: Summary Statistics (2016 Census)

	Pop (Million)	# DA	# ISO3
Montreal	4.07	6,355	153
Ottawa	1.31	1,904	153
Toronto	5.87	7,293	153
Calgary	1.38	1,706	152
Edmonton	1.30	1,622	151
Vancouver	2.44	3,381	146

DAs in the Plateau, Montreal (2016)

Data: explanatory variables

- We use data from: CEPII, UNWTO, Melitz and Toubal (2014, JIE), Spolaore and Wacziarg (2009, QJE).
- For more than 13,000 unique country *pairs* we have information on:

1. Conflicts, hegemony and sharing common colonizers.
2. Cultural, religious, genetic, and linguistic proximity measures.
3. Economic controls (Trade, tourism, common currency, trade agreements).
4. Geographical controls (contiguity, distance, same region).

Empirical Strategy

K -density PDF

Consider ethnic groups i, j. There are n^i_l, n^j_m agents of each group in location l and m.

Following Duranton and Overman (2005), we estimate the density for all bilateral distances d_{ij} between individuals belonging to i and j , for each city c, and at a distance d :

$$\widehat{k}^{ij}_c(d) = \frac{1}{h \sum_{l=1}^{L_c} \sum_{m=1}^{L_c} n^i_l n^j_m} \sum_{l=1}^{L_c} \sum_{m=1}^{L_c} n^i_l n^j_m f\left(\frac{d - d_{lm}}{h}\right), \qquad (1)$$

where $f(\cdot)$ is a Gaussian kernel, h is the bandwidth.

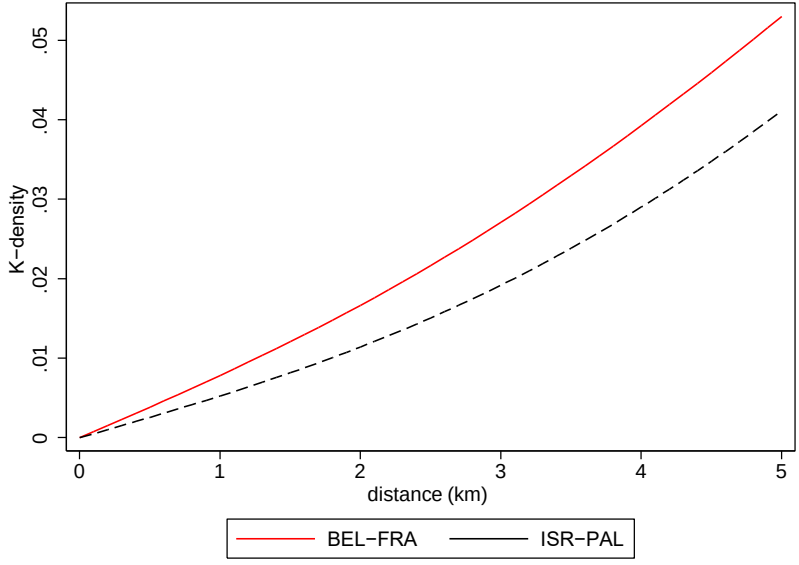
K -density CDF

Our dependent variable is obtained by computing the cumulative measures as follows:

$$\widehat{K}_c^{ij}(d) = \int_0^d \widehat{k}_c^{ij}(\zeta)\mathrm{d}\zeta. \qquad (2)$$

Interpretation: if $K_c^{ij}(1km) = 0.3$, where i = Spaniards and j = Catalans, this means that 30% of bilateral distances between pairs of Spaniards and Catalans are less than 1 kilometer in city c.

Illustration: Montreal (2016)



Baseline Model

$$\widehat{K}_c^{ij}(\overline{d}) = \alpha + X^{ij}\beta + \delta_c^i + \delta_c^j + \varepsilon_c^{ij} \qquad (3)$$

Where:

1. $K_c^{ij}(\overline{d})$: measure of geographic proximity of ethnic groups.
2. X^{ij}: set of explanatory variables.
3. δ_c^i δ_c^j: city-country fixed effects.
4. ε_c^{ij}: error term.

Results

Table: Multivariate baseline results, 2016 Census.

Dependent variable: $\hat{K}_i^j$(500m)	(1)	(2)	(3)	(4)	(5)
Contiguity	0.03[a]	0.02[a]	0.02[a]	0.01[a]	0.01[a]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Same continent	0.07[a]	0.04[a]	0.04[a]	0.04[a]	0.03[a]
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Common currency		0.02[a]	0.01[a]	0.01[a]	0.01[a]
		(0.00)	(0.00)	(0.00)	(0.00)
Free trade agreement		0.03[a]	0.03[a]	0.03[a]	0.02[a]
		(0.00)	(0.00)	(0.00)	(0.00)
Both OECD		0.03[a]	0.03[a]	0.03[a]	0.03[a]
		(0.00)	(0.00)	(0.00)	(0.00)
Bilateral trade flows		0.01[a]	0.01[a]	0.01[a]	0.01[a]
		(0.00)	(0.00)	(0.00)	(0.00)
Bilateral tourist flows		-0.01[a]	-0.01[a]	-0.01[a]	-0.01[a]
		(0.00)	(0.00)	(0.00)	(0.00)
GDP per capita gap		-0.07[a]	-0.07[a]	-0.07[a]	-0.07[a]
		(0.00)	(0.00)	(0.00)	(0.00)
Were same country			0.01[a]	0.01[a]	0.01[a]
			(0.00)	(0.00)	(0.00)
Common colonizer			0.04[a]	0.03[a]	0.03[a]
			(0.00)	(0.00)	(0.00)
Colonial relationship			0.00	0.00	0.00
			(0.00)	(0.00)	(0.00)
COL				-0.02[a]	0.01[a]
				(0.00)	(0.00)
Common religion				0.01[a]	0.01[a]
				(0.00)	(0.00)
Genetic Distance (alelle, plurality groups)					-0.03[a]
					(0.00)
Sample size	68,055	62,145	62,145	62,145	62,145
R^2	0.86	0.87	0.87	0.87	0.87

Robustness check

We run numerous robustness checks:

1. Bigger size sample only (2016).
2. Using 2006 census data.
3. Different distances (100m to 1km).
4. Alternative measures of language, genetics, religion, culture and historico-political relationships:

Table: Alternative measures of our key variables, 2016 Census.

Description	Coeff.	0.003	Sample size	R^2
Common spoken language	0.015[a]		62,145	0.872
Common native language	0.005[b]	0.002	62,145	0.872
Linguistic proximity (Tree, unadjusted)	0.008[a]	0.002	62,145	0.872
Linguistic proximity (Tree, adjusted)	0.008[a]	0.002	57,635	0.875
Linguistic proximity (ASJP, unadjusted)	0.006[b]	0.002	62,145	0.872
Linguistic proximity (ASJP, adjusted)	0.006[b]	0.002	57,635	0.875
Common Language Index (log specification)	0.015[a]	0.003	57,635	0.875
Common Language Index (level specification)	0.013[a]	0.003	62,145	0.872
Common official or primary language	0.013[a]	0.003	62,145	0.872
Language is spoken by at least 9% of the population	0.007[b]	0.003	62,145	0.872
Linguistic distance (words, plurality languages)	-0.010[b]	0.004	14,748	0.904
Linguistic distance (words, weighted)	-0.014[a]	0.005	7,760	0.931
Linguistic distance (trees, plurality languages)	0.003	0.002	52,073	0.865
Linguistic distance (trees, weighted)	0.002	0.002	52,073	0.865
Religious distance (plurality Fearon et al.)	-0.007[b]	0.003	51,594	0.866
Religious distance (weighted, Fearon et al.)	-0.011[a]	0.003	51,594	0.866
Religious distance (plurality, WCD)	-0.013[a]	0.003	59,532	0.872
Religious distance (weighted, WCD)	-0.017[a]	0.004	59,532	0.872

Table: Alternative measures of our key variables, 2016 Census.

Euclidian cultural distance, all categories	-0.032[a]	0.006	13,674	0.922
Euclidian cultural distance, category A only	-0.020[a]	0.005	13,674	0.922
Euclidian cultural distance, category C only	-0.014[a]	0.005	13,674	0.921
Euclidian cultural distance, category D only	-0.014[a]	0.005	13,674	0.921
Euclidian cultural distance, category E only	-0.019[a]	0.006	13,674	0.922
Euclidian cultural distance, category F only	-0.007[c]	0.004	13,674	0.921
Euclidian cultural distance, binary choice questions only	-0.019[a]	0.005	13,674	0.922
Euclidian cultural distance, non-binary choice questions only	-0.027[a]	0.006	13,674	0.922
Genetic distance (alelle, weighted)	-0.043[a]	0.003	59,462	0.871
Genetic distance (microsatellite variation, weighted)	-0.058[a]	0.004	57,805	0.871
Genetic distance (microsatellite variation, plurality groups)	-0.053[a]	0.004	57,805	0.871
Country was post-45 colonizer of the other	0.000	0.001	62,145	0.871
Countries in the same ‘empire’ or had common colonizer	0.017[a]	0.003	62,145	0.871
Hegemony relationship	0.003[c]	0.002	62,145	0.871
Number of years since no longer siblings (cond. on sibling = 1)	-0.035[a]	0.011	10,871	0.896
Common legal origins pre-independence	0.023[a]	0.002	62,145	0.872
Common legal origins post-independence	0.014[a]	0.002	62,145	0.871
Common legal origins across countries changed	-0.004[c]	0.003	62,145	0.871

Heterogenous effects by city

$$\widehat{K}_c^{ij}(\overline{d}) = \alpha + \beta X^{ij} + \beta_c \, Lang_{ij} * city + \delta_c^i + \delta_c^j + \varepsilon_c^{ij} \qquad (4)$$

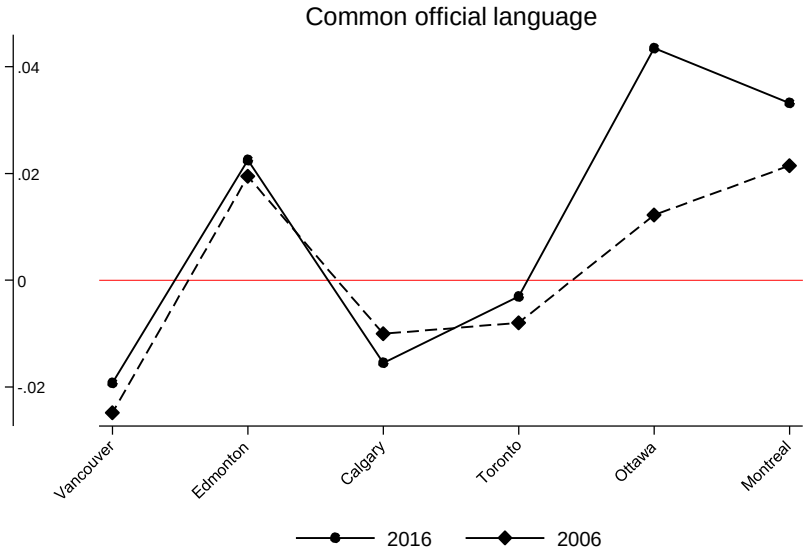
We also estimate different coefficients by metropolitan area for:

1. Religion.
2. Colonial relationships.
3. Genetics.

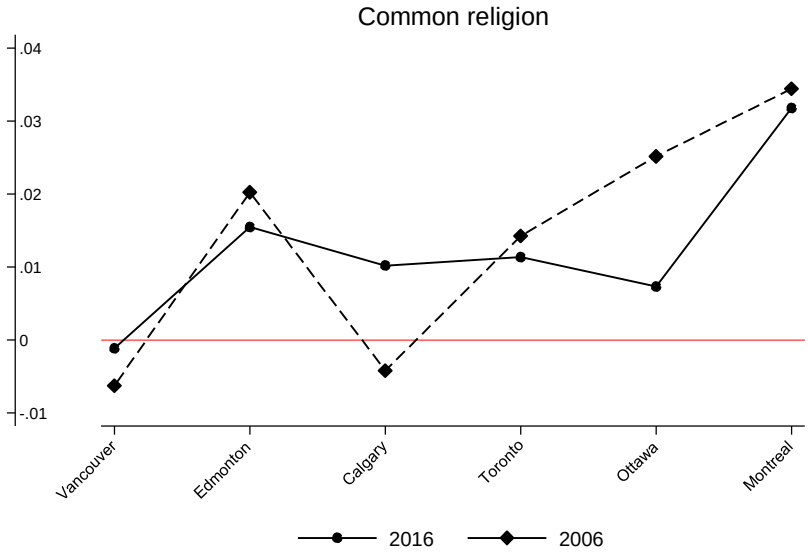
Table: Heterogeneous effects by city.

	Montreal	Ottawa	Toronto	Calgary	Edmonton	Vancouver
			2016 Census			
Common official language	0.033[a]	0.044[a]	-0.003	-0.015[a]	0.023[a]	-0.019[a]
	(0.004)	(0.005)	(0.003)	(0.005)	(0.006)	(0.005)
Common religion	0.032[a]	0.007	0.011[a]	0.010[a]	0.015[a]	-0.001
	(0.004)	(0.005)	(0.003)	(0.005)	(0.005)	(0.004)
Colonial relationship	0.012[a]	0.008[a]	0.009[a]	0.001	0.001	-0.004[c]
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
Genetic distance (alelle, plurality groups)	-0.027[a]	-0.033[a]	-0.017[a]	-0.025[a]	-0.073[a]	-0.036[a]
	(0.004)	(0.005)	(0.003)	(0.005)	(0.006)	(0.006)
			2006 Census			
Common official language	0.021[a]	0.012	-0.008[c]	-0.010	0.019[c]	-0.025[a]
	(0.006)	(0.011)	(0.005)	(0.007)	(0.011)	(0.005)
Common religion	0.034[a]	0.025[c]	0.014[a]	-0.004	0.020[c]	-0.006
	(0.005)	(0.016)	(0.004)	(0.007)	(0.012)	(0.005)
Colonial relationship	0.009[a]	0.008[a]	0.006[a]	-0.003	0.000	-0.001
	(0.004)	(0.002)	(0.002)	(0.003)	(0.004)	(0.002)
Genetic distance (alelle, plurality groups)	-0.031[a]	-0.012	-0.007[a]	0.003	-0.022[a]	-0.026[a]
	(0.005)	(0.012)	(0.004)	(0.007)	(0.011)	(0.006)

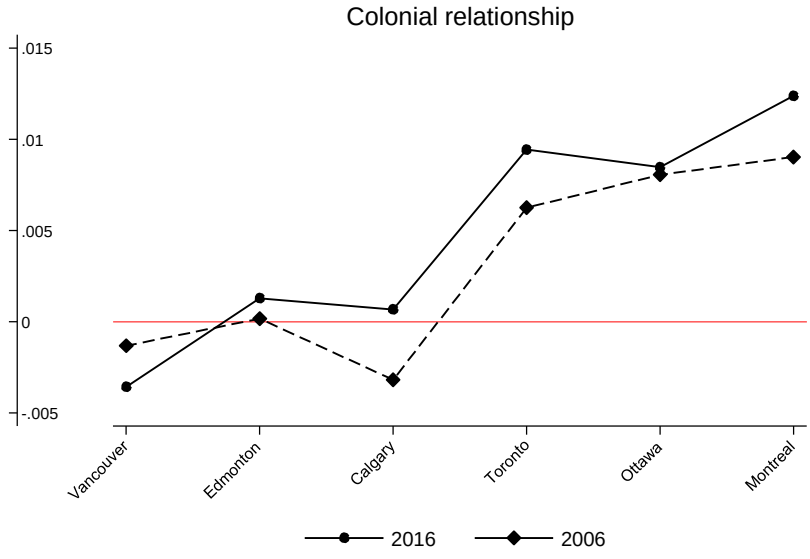
East-west gradient



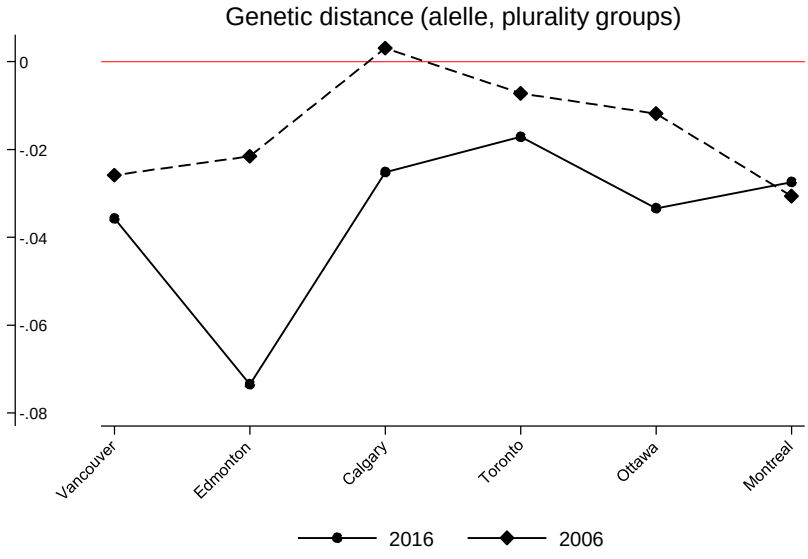
East-west gradient



East-west gradient



East-west gradient



Conclusion

Conclusion

1. Homophily has an impact on the choice of neighbors.
2. Religion, language, culture, and genetics explain in part the colocation of ethnic groups.
3. There is an east-west gradient, with preferences playing a larger role in eastern Canada.
4. Efficient public housing policy that aims to increase diversity should consider these factors.