Love thy neighbor?

An empirical test of neighborhood ethnicity change and Schelling behavior

Jørgen Baun Høst

University of Copenhagen · Department of Economics · May 2025

Overview

- **Motivation**: Demographic transition and Schelling's model
- **Research Question**: Does the ethnicity of your nearest neighbor affect propensity to move?
- **Methods**: Nearest-neighbor research design with comprehensive administrative data
- **Results**: Schelling behavior?
- **Heterogeneity Analysis**: SES
- **Conclusion**

Defintions:

1. Native households: All members are of Danish origin
2. Non-Western households: At least 1 member is of non-Western origin
3. Western households: At least 1 member is of Western origin (but no non-Western household members)

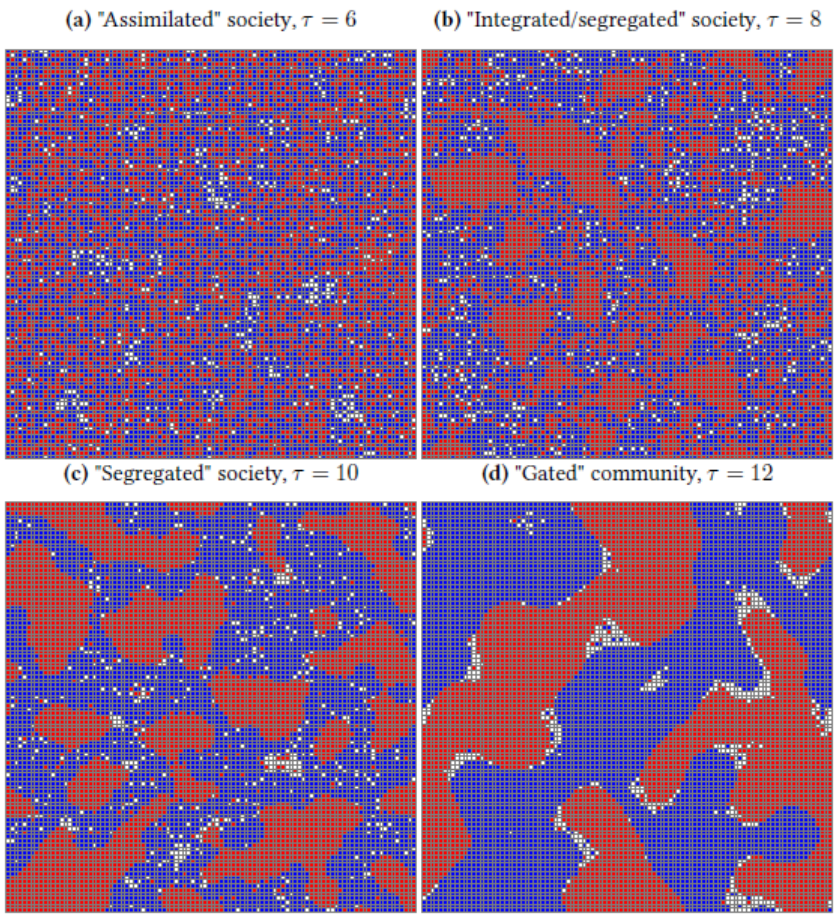
Introduction

Theoretical Background: Schelling's Model

- (Schelling, 1971) proposed that neighborhoods may "tip" when minority share reaches a threshold
- Even with relatively tolerant preferences toward diversity
- Three types of segregation:
  1. Organized segregation (e.g., historical Jim Crow laws)
  2. Economically induced segregation (clustering by income/education)
  3. Individually motivated segregation ← **Focus of this paper**

Schelling's key insight: Small individual preferences can lead to macro-level segregation

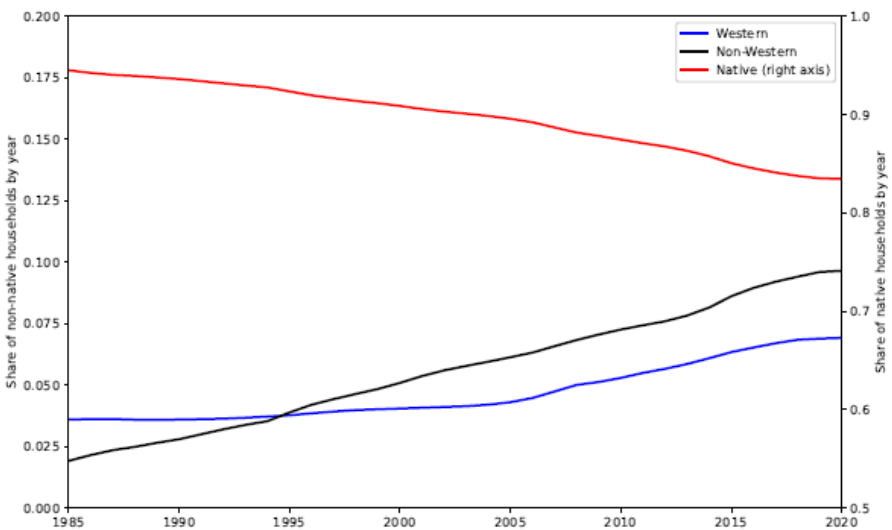
Figure C.1: Schelling model simulations by τ required same-type neighbors



Demographic context

- Denmark has transformed from a relatively homogeneous society to increasing ethnic diversity
- Non-Western households grew from ~2% in 1985 to ~10% by 2020
- Limited empirical evidence on how ethnic background directly influences residential sorting

Figure 1: Distribution of households (1985-2020)



Note: This is based on own calculations and shows the distribution of households by type as defined in section 1.1. I sample households that are present at December 31st in each year from 1985.

Methods

Identification Challenge

$$V_{i,j,t} = f(Z_{i,t}, X_{j,t}, \xi_{j,t}) + \sum_k g(Z_{i,t}, Z_{k,t}, D_{i,k}) + \delta E\left[V_{i,j,t+1}\right] + \varepsilon_{i,j,t}$$

Where:

- $f(\cdot)$: Utility from neighborhood amenities
- $g(\cdot)$: Utility from characteristics of each neighbor k at distance $D_{i,k}$
- Z_i: Observable household attributes
- X_j: Observable neighborhood attributes
- ξ_j: Unobservable neighborhood attributes
- $\varepsilon_{i,j,t}$: Idiosyncratic preferences

Key identification challenges:

- Unobserved neighborhood amenities
- Dynamic preferences (expectations of future changes)
- Selection effects (who moves where is not random)

Nearest neighbor research design

Innovative approach from (Bayer *et al.*, 2022):

Compare households within the same neighborhood who receive different-type neighbors. Why does this work? Consider two households:

- Household a: New different-type e' neighbor among their nearest (rank 1-3) neighbors
- Household b: New different-type e' neighbor slightly further away (rank 4-6)

Difference in moving propensity:

$$\begin{aligned} Y_a(e', k_{\text{nearest}}) - Y_b(e', k_{\text{near}}) &= (\mathbb{P}[e', k_{\text{nearest}}]) - \mathbb{P}[e', k_{\text{near}}])) \\ &+ (\xi_a B(e', k_{\text{nearest}}) - \xi_b B(e', k_{\text{near}})) \\ &+ (\rho_a - \rho_b) + \left(\omega_j - \omega_j\right) \leftrightarrow \\ &= \mathbb{P}[e', k_{\text{nearest}}]^* + \rho_a - \rho_b \end{aligned}$$

1. $\mathbb{P}[e', k_{\text{nearest}}]) - \mathbb{P}[e', k_{\text{near}}]) > 0$
2. $\xi_a B(e', k_{\text{nearest}}) - \xi_b B(e', k_{\text{near}}) \approx 0$: (almost) no difference in future neighborhood quality expectation
3. $Y_a(e', k_{\text{nearest}}) - Y_b(e', k_{\text{near}}) \perp \rho_a - \rho_b$: For existing households, location of new neighbors are not related to idiosyncratic factors ρ.

Nearest neighbor research design

- **Treatment group**: Households with new different-type neighbors among their 3 nearest neighbors
- **Control group**: Households with new different-type neighbors "just down the road" (ranks 4-6)

$$Y_{i,j,t} = \beta_1 I[e', k = n_{\text{nearest}}] + \beta_2 I[e', k = n_{\text{near}}] + \beta_3 I[e', k = n_{\text{close}}] + \gamma Z_{i,j,t} + \omega_{j,t} + \varepsilon_{i,j,t}$$

Parameter of interest:

$$\beta_1 - \beta_2$$

This design addresses key identification challenges by comparing households experiencing same neighborhood conditions but different micro-geography of new neighbors.

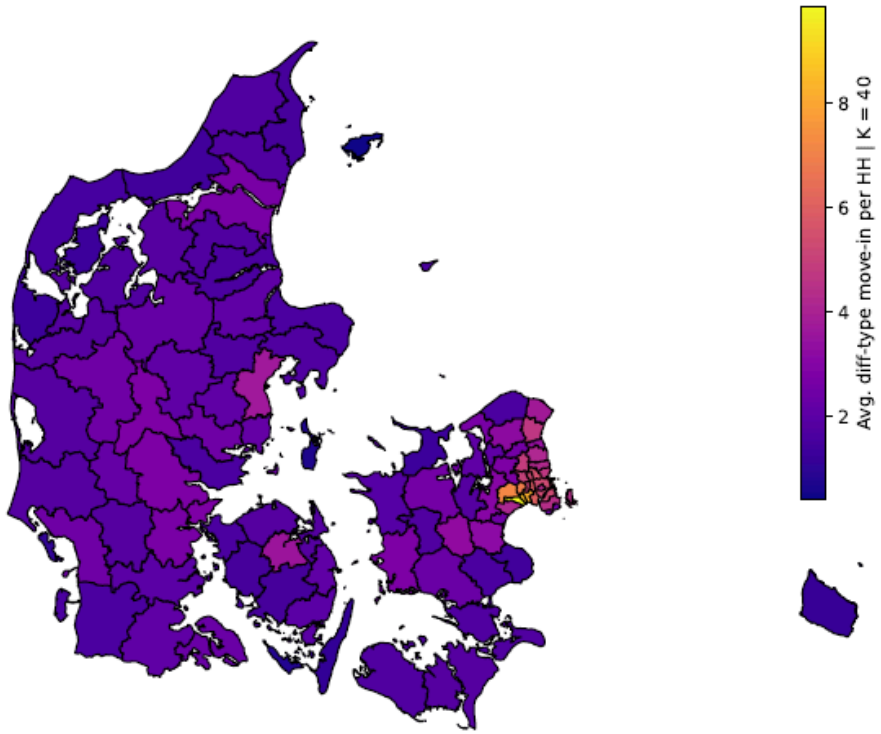
Data

Spatial Patterns of New Different-Type Neighbors

Key spatial patterns:

- Clear east-west and urban-rural divide
- Concentration in Copenhagen and surroundings
- Highest incidence in Ishøj (~9 new different-type neighbors)
- Copenhagen (~6), Aarhus and Odense (~4)

Figure 3: Incidence of new different-type neighbors (1985-2020)



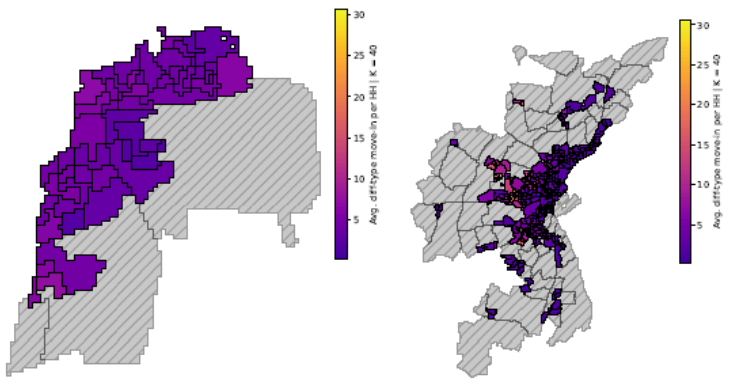
Note: The figure show the variation in receiving a new-non Western neighbors within the 40 closest parcels for native households. Municipal borders correspond to the ones imposed by "Kommunalreformen" in 2007. Household types are split up in to three types, see section 1.1 for more details.

Spatial Patterns of New Different-Type Neighbors

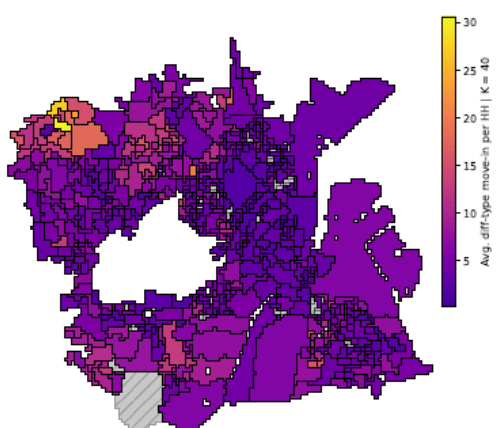
Within-city variation:

- Some Copenhagen neighborhoods: 30+ new non-Western neighbors
- Other Copenhagen neighborhoods: <2 new non-Western neighbors

Figure 4: Incidence of new different-type neighbors at the neighborhood level (1985-2020)



(a) Ishøj (b) Aarhus



(c) Copenhagen

Summary Statistics

3 key observations from summary statistics:

1. Treated households show higher mobility: 23-24% vs. 19-20% for "control" households
2. Treated native households have lower wealth (48,500 DKK vs. 81,000 DKK) and income
3. Non-Western households are better educated on average (by ~2 years)

Table 1: Summary statistics

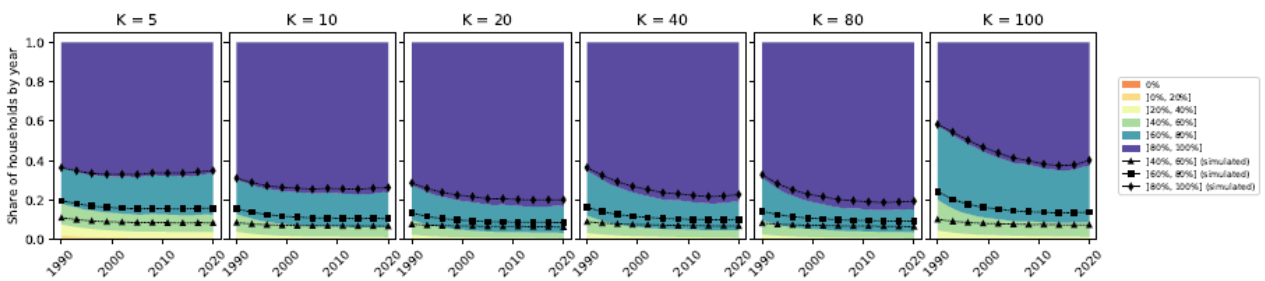
	Native households						Non-Western households					
	All		Nearest		Close		All		Nearest		Close	
Household characteristics												
Move within 2 years	17.45	(37.95)	23.19	(42.21)	19.74	(39.80)	19.14	(39.34)	24.01	(42.71)	19.34	(39.50)
Real inc. (1000s) DKK	344.11	(120.22)	337.16	(119.58)	341.07	(120.29)	313.13	(111.14)	317.37	(114.96)	314.62	(112.27)
Real net wealth (1000s) DKK	81.03	(202.68)	48.51	(185.87)	62.08	(193.44)	45.52	(163.75)	41.34	(160.52)	46.95	(165.29)
Employed	0.86	(0.34)	0.83	(0.37)	0.84	(0.37)	0.83	(0.38)	0.84	(0.37)	0.83	(0.37)
Years of education	10.15	(6.24)	10.03	(6.26)	10.13	(6.25)	12.12	(5.34)	12.13	(5.47)	12.10	(5.37)
Distance to neighbor			3.20	(4.29)	33.15	(20.69)			2.81	(3.70)	30.60	(18.15)
Household size	1.81	(1.06)	1.63	(0.94)	1.70	(0.99)	2.14	(1.29)	1.96	(1.20)	2.13	(1.27)
Oldest household member	44.25	(8.93)	43.51	(9.03)	44.18	(9.00)	43.37	(8.71)	41.84	(8.61)	43.26	(8.69)
Neighborhood characteristics												
Population density	7418.23	(7644.95)	8969.51	(8151.09)	8387.60	(7945.02)	9198.14	(8155.82)	9667.16	(8521.05)	9331.84	(8326.29)
Native share	0.89	(0.10)	0.81	(0.14)	0.83	(0.13)	0.76	(0.18)	0.78	(0.16)	0.78	(0.16)
Non-Western share	0.08	(0.09)	0.15	(0.14)	0.13	(0.12)	0.19	(0.17)	0.17	(0.15)	0.18	(0.16)
Real income (median)	322.85	(32.94)	318.41	(35.49)	320.60	(34.61)	317.40	(35.96)	319.73	(36.69)	318.51	(35.72)
Real net wealth (median)	42.74	(55.89)	27.58	(44.34)	32.01	(48.11)	27.87	(44.55)	27.54	(42.75)	27.90	(43.95)
N	33,496,551		761,842		4,671,062		3,330,242		414,103		1,443,373	

Note: This table shows presents summary statistics for households "at-risk" of receiving a different-type neighbor. Standard deviations in parenthesis. Income and wealth are equivalised to facilitate comparison between households of different size and composition. The *All* column denotes quarter-by-year observation for the sample of household defined in section 3.6. The *Nearest* ("treated") and *Close* ("control") columns denote instances, where a household experienced a new different-type among their $K \in [1, 2, 3]$ nearest neighbors or close neighbors ($K \in [4, 5, ..., 40]$).

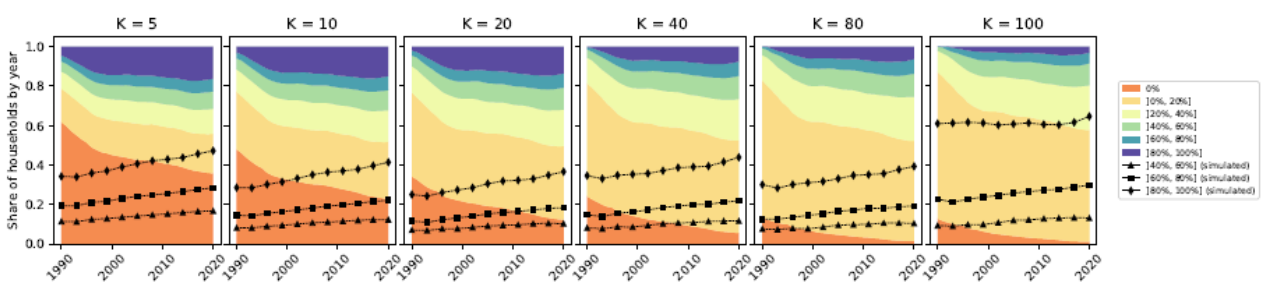
Residential sorting over time

- Given demographic context, how has residential sorting developed over time?
- **Natives**:
  - 1990 [K=100 nearest]: 40 percent had between 80-100 same-type neighbors.
  - 2020 [K=100 nearest]: 60 percent had between 80-100 same-type neighbors.
- *Quick and dirty* counterfactual simulation with fixed 1990-distribution show little-to-no change in sorting patterns.
- **Schelling behavior?**

Figure 5: Same type neighbor by K-nearest proximity (1990-2020)



(a) Native households



(b) Non-Western households

Results

Main results: Assymetric Schelling behavior

Table 4: Estimates of Schelling behavior (native households)

	Move within 2 years (=100)					
	(1)	(2)	(3)	(4)	(5)	(6)
New diff neighbor $k_{nearest}$ v k_{near}	0.357***	0.368***	0.354***	0.328***	0.309***	0.320***
	(0.084)	(0.084)	(0.081)	(0.081)	(0.081)	(0.081)
N	5,365,811	5,365,811	5,365,811	5,365,811	5,365,811	5,365,811
Neighborhood-by-quarter FE	X	X	X	X	X	X
Mean of dependent variable	20.23	20.23	20.23	20.23	20.23	20.23
Number of neighborhoods	3444	3444	3444	3444	3444	3444
Income		X	X	X		X
Wealth					X	X
Tenure			X	X	X	X
Age				X	X	X

* p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors (in parenthesis) are clustered at the neighborhood level. The table reports the estimate of $\beta_1 - \beta_2$ from equation 6. Table A.3 contains the complete set of coefficients.

- Increase moving propensity by ~0.3 percentage points when receiving a new non-Western neighbor
- 1.6% increase relative to baseline exit rate
- Effect stable across specifications
- Robust to controls for income, wealth, age, tenure

Table 5: Estimates of Schelling behavior (non-Western households)

	Move within 2 years (=100)					
	(1)	(2)	(3)	(4)	(5)	(6)
New diff neighbor $k_{nearest}$ v k_{near}	0.063	0.062	0.155	0.097	0.097	0.096
	(0.129)	(0.129)	(0.127)	(0.126)	(0.126)	(0.126)
N	1,795,109	1,795,109	1,795,109	1,795,109	1,795,109	1,795,109
Neighborhood-by-quarter FE	X	X	X	X	X	X
Mean of dependent variable	19.96	19.96	19.96	19.96	19.96	19.96
Number of neighborhoods	3332	3332	3332	3332	3332	3332
Income		X	X	X		X
Wealth					X	X
Tenure			X	X	X	X
Age				X	X	X

* p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors (in parenthesis) are clustered at the neighborhood level. The table reports the estimate of $\beta_1 - \beta_2$ from equation 6. Table A.4 contains the complete set of coefficients.

- Show substantially smaller response: 0.06-0.1 percentage points
- ~0.5% relative to baseline exit rate
- Not statistically significant
- Suggests they are unaffected by identity of new native neighbors

Heterogeneity by Socioeconomic Status

SES definitions:

- **Low SES**: Income < 200,000 DKK, outside labor market or $\leq$ 11 years of education
- **High SES**: Income $\geq$600,000 DKK, employed full-time or $\geq$ 18 years of education

Key findings:

Table 6: Estimates of Schelling behavior (native households) by SES

	Move within 2 years (=100)					
	SES: Low	SES: High	SES: Low v Low	SES: Low v High	SES: High v High	SES: High v Low
	(1)	(2)	(3)	(4)	(5)	(6)
New diff neighbor $k_{nearest}$ v k_{near}	0.557***	0.287	0.558***	0.275	-1.050	0.036
	(0.090)	(0.297)	(0.105)	(0.453)	(0.887)	(0.442)
N	5,883,637	609,652	3,614,630	156,497	55,008	310,061
Neighborhood-by-quarter FE	X	X	X	X	X	X
Mean of dependent variable	18.89	20.75	17.83	17.01	23.65	20.22
Number of neighborhoods	3451	3450	3451	3248	2688	3446
Income		X			X	X
Tenure	X	X	X	X	X	X
Age	X	X	X	X	X	X

* p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors (in parenthesis) are clustered at the neighborhood level. The table reports the estimate of $\beta_1 - \beta_2$ from equation 6. Table A.5 contains the complete set of coefficients.

Table 7: Estimates of Schelling behavior (non-Western households) by SES

	Move within 2 years (=100)					
	SES: Low	SES: High	SES: Low v Low	SES: Low v High	SES: High v High	SES: High v Low
	(1)	(2)	(3)	(4)	(5)	(6)
New diff neighbor $k_{nearest}$ v k_{near}	0.008	-1.452*	0.020	0.491	-1.654	-1.551
	(0.115)	(0.624)	(0.126)	(0.552)	(1.381)	(0.803)
N	1,984,581	169,445	1,609,988	127,667	37,413	117,717
Neighborhood-by-quarter FE	X	X	X	X	X	X
Mean of dependent variable	18.33	25.58	18.26	18.12	28.08	26.20
Number of neighborhoods	3447	3135	3444	3361	2383	3087
Income		X			X	X
Tenure	X	X	X	X	X	X
Age	X	X	X	X	X	X

* p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors (in parenthesis) are clustered at the neighborhood level. The table reports the estimate of $\beta_1 - \beta_2$ from equation 6. Table A.6 contains the complete set of coefficients.

- Schelling behavior primarily driven by low-SES native households responding to low-SES non-Western household: ~0.56 percentage points or ~2.8% increase from baseline exit rate
- Very rare for low-SES native households to receive high-SES non-Western neighbors and vice versa

Danish findings vs. (Bayer *et al.*, 2022) U.S. results:

Context	Magnitude
Denmark (Native)	1.6% above baseline
Denmark (Non-Western)	0.5% above baseline
U.S. (White)	4% above baseline
U.S. (Black)	6% above baseline

Possible explanations for differences:

- Institutional variation in housing market and integration policies
- Different neighborhood contexts (urban/dense vs. suburban)
- Stronger "type 1 & 2" (institutional & economical) segregation?

Conclusion

Key Findings

1. Native Danish households increase moving propensity by 1.6% when receiving non-Western neighbors
2. Non-Western households show no significant response to new native neighbors
3. Heterogeneity by SES: Low-SES native households responding to low-SES non-Western neighbors show strongest effects (2.8%)
5. Magnitude in Denmark (1.6%) more modest than in U.S. context (4-6%)

Extensions

- Do native households respond to new Western neighbors?
- How much are native households willing to pay in premium to live in a more homogenous neighborhood?
  - Variation?
- Those who show Schelling behavior, where do they move to?

Thank you for your attention!

Questions?

References

Bayer, P. *et al.* (2022) "Distinguishing Causes of Neighborhood Racial Change: A Nearest Neighbor Design," *Social Science Research Network* [Preprint]. Available at: https://doi.org/10.3386/w30487.

Schelling, T.C. (1971) "Dynamic models of segregation," *Journal of mathematical sociology*, 1(2), pp. 143–186.

Appendix

Balance tests

$$X_{i,j,t} = \varphi_1 I[e', k = n_{\text{nearest}}] + \varphi_2 I[e', k = n_{\text{near}}] + \varphi_3 I[e', k = n_{\text{close}}] + \omega_{j,t} + \varepsilon_{i,j,t}$$

Table 2: Balance test (native)

	Income (1,000)	Net wealth (1,000)	Oldest HH member (years)	Tenure (days)	Employed	Educ. length (years)	HH size
New diff neighbor $k_{nearest}$ v k_{near}	-1.408***	-1.903***	-0.043*	1.936	-0.002**	-0.035*	-0.009***
	(0.254)	(0.402)	(0.020)	(5.999)	(0.001)	(0.014)	(0.002)
N	5,365,811	5,365,811	5,365,811	5,365,811	5,365,811	5,365,811	5,365,811
Neighborhood-by-quarter FE	X	X	X	X	X	X	X
Mean of dependent variable	341.26	60.61	44.05	2646.87	0.84	10.11	1.70
Number of neighborhoods	3444	3444	3444	3444	3444	3444	3444

* p < 0.05, ** p < 0.01, *** p < 0.001 Standard errors (in parenthesis) are clustered at the neighborhood level. The table reports the estimate of $\phi_1 - \phi_2$ from equation 11. Table A.1 contains the complete set of coefficients.

Table 3: Balance test (non-Western)

	Income (1,000)	Net wealth (1,000)	Oldest HH member (years)	Tenure (days)	Employed	Educ. length (years)	HH size
New diff neighbor $k_{nearest}$ v k_{near}	-0.065	0.171	-0.061*	31.109***	0.001	-0.011	0.001
	(0.347)	(0.522)	(0.028)	(7.656)	(0.001)	(0.018)	(0.004)
N	1,795,109	1,795,109	1,795,109	1,795,109	1,795,109	1,795,109	1,795,109
Neighborhood-by-quarter FE	X	X	X	X	X	X	X
Mean of dependent variable	315.64	46.50	42.94	2298.84	0.84	12.11	2.11
Number of neighborhoods	3332	3332	3332	3332	3332	3332	3332

* p < 0.05, ** p < 0.01, *** p < 0.001 Standard errors (in parenthesis) are clustered at the neighborhood level. The table reports the estimate of $\phi_1 - \phi_2$ from equation 11. Table A.1 contains the complete set of coefficients.

Robustness Checks

Alternative specifications:

- Combining all control distances into a single category
- Varying distance thresholds for nearest neighbors
- Different neighborhood definitions

Results remain consistent across specifications:

- Spatial decay of effects provides additional support for Schelling mechanism
- Moving response decreases monotonically with distance to new different-type neighbors
- Effects primarily concentrated within 25 meters

The Schelling Model Simulation

Simple agent-based model:

- Agents of two types randomly allocated on grid
- Agents move if share of different-type neighbors exceeds tolerance threshold
- Even with modest tolerance thresholds, segregation emerges
- This visualization demonstrates how small individual preferences

can lead to significant macro-level segregation patterns