Mixing in Early Childhood

Vincent Boucher[1] Semih Tumen[2] Michael Vlassopoulos[3]
Jackline Wahba[4] Yves Zenou[5]

November 2020

[1]Laval University
[2]TED University
[3]University of Southampton
[4]University of Southampton
[5]Monash University

Introduction

- In recent years, an unprecedented number of people around the world have been forced to leave their home.
- As of the end of 2019, the number of forcibly displaced individuals is above 70 million worldwide, around 26 million of which are refugees (UN Refugee Agency).
- Beyond the obvious humanitarian implication of displacement, this has raised major challenges for hosting countries in terms of integrating the refugee populations.

Motivation

- The integration of refugee children into the education system is important for their academic outcomes as well as their social and emotional development and well-being.
- The success (or lack of) integration in schools is important for the future labor market and social integration potential of the refugee children.
- Integration can also affect the attitudes of native children toward refugees and social cohesion in host countries.
- This paper addresses these issues in the context of the integration of Syrian refugee children in Turkey.

This paper (Random field experiment)

- Does mixing refugee children with host children in early childhood educational settings enhance the integration process?
- We embed a randomized field experiment into the design of a Summer Camp early childhood program in Turkey.
- We randomly allocate children to form groups with high, medium, and low refugee concentration.
- Estimate the causal impact of refugee concentration on outcomes of Syrian and Turkish children

This paper (Model)

- Develop a new model of friendship formation that embeds two main features that shape the network formation:
- preference for forming friendships with members of one's own group (homophily),
- congestion in the friendship formation process.
- This model generates comparative statics predictions, on how friendship formation varies with the composition of the group, that are consistent with the patterns we see in the data.

This paper (Structural estimation and counterfactual policies)

- Structurally estimate the model.
- Assess the quantitative relevance of the mechanisms identified by our theory.
- Perform counterfactual policy experiments.

Outcomes and Findings (Random field experiment)

Outcomes:

- prosocial behavior (dictator game).
- interethnic friendships
- language skills (Syrian children)

Findings:

1. Turkish children: In-group favoritism in prosocial behavior & Contact leads to more interethnic friendships.
2. Syrian children: No evidence of ethnic bias in prosocial behavior & Contact leads to more interethnic friendships (weaker effect).
3. Ethnic composition of group does affect (negatively) the language skill acquisition of Syrian children.

Outcomes and Findings (Model and structural estimation)

Outcomes:

- prosocial behavior (social effort).
- interethnic friendships (link formation).
- Homophily

Findings:

1. Syrian children socialize less than Turkish children.
2. Syrian children have a stronger taste for conformity than Turkish kids.
3. Interethnic links: Syrian-Turkish and Turkish-Syrian, are less likely to be formed than homogeneous links.
4. Network formation positively affected by the language skills of Syrian children.

Outcomes and Findings (Counterfactual policies)

Findings:

1. Inbreeding homophily (share of own-group friends relative to the size of one's group) has an inverse u-shape, i.e., the impact of contact on friendship formation is non-monotonic, so more contact can backfire.

2. Making Syrian kids fluent in Turkish is equivalent to reducing ethnic biases by more than 50%.

3. Thus, contact alone does not guarantee improved inter-group outcomes between refugees and natives in educational settings.

Key Insights

- Mixing of majority and minority children has a more nuanced effect on friendship formation than previously understood.
- Exposure to non-coethnic children has a non-monotonic (inverse U-shaped) effect on inbreeding homophily.
- Programs aiming at integrating refugee children should design policies that incorporate social cohesion, pro-sociability, language skills and interethnic exposure issues as early as the pre-school level.

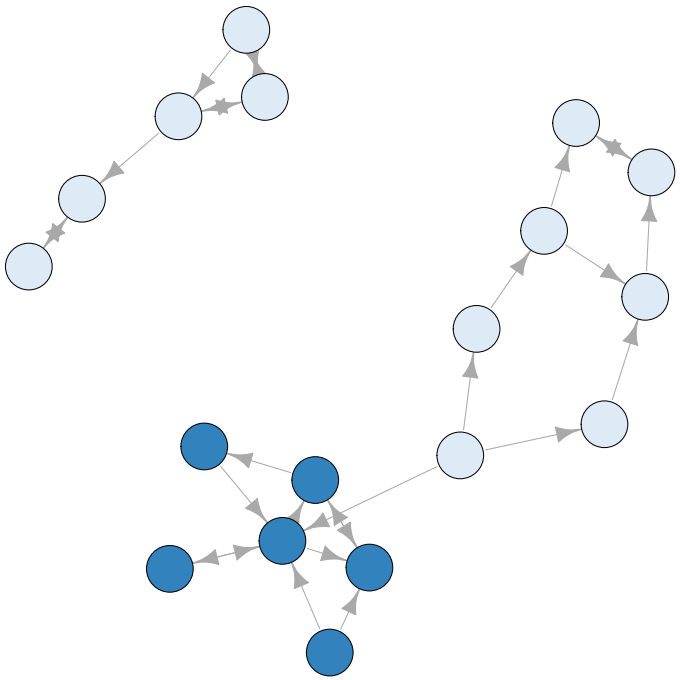


Figure: Friendship network in classroom #33 (n=18, share of Syrians = 0.33). Each node represent a child. Colored nodes are Syrians. Segregation

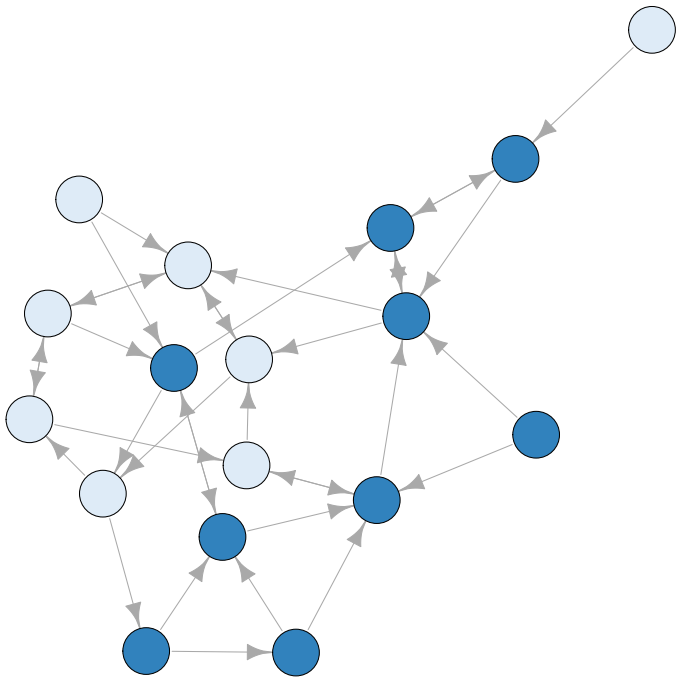


Figure: Friendship network in classroom #29 (n=17, share of Syrians = 0.53). Each node represent a child. Colored nodes are Syrians. Integration

Testing the contact hypothesis in educational settings

- Random assignment to a non-coethnic peer decreases prejudice and discrimination (Van Laar et al., 2005; Boisjoly et al., 2006; Carrell et al., 2019; Corno et al., 2019, etc.).

Our contribution

- Tests contact theory for pre-school children of age 5 (early childhood).
- Contact happens between refugee and native children.
- Structural estimation and counterfactual policies.
- Show a non-monotonic (inverse U-shaped) effect on inbreeding homophily.

Network formation

- Many papers testing network formation, e.g. Graham (2017), Mele (2017).
- Few study have network formation and inter-ethnic groups (Boucher, 2015; Currarini et al., 2010, 2009; Mele, 2020)

Our contribution

- Provide a field experiment with a structural model.
- Test network formation for early childhood.
- Counterfactual policies.

Formation of Preferences in Early Childhood

- Childhood is a formative period for the development of social preferences (Sutter et al., 2019) and that educational interventions in this period can help shape them (Cappelen et al., 2020; Kosse et al., 2020).
- Few papers examine in-group favoritism among children (Fehr et al., 2008, 2013; Angerer et al., 2016). Find more generous behavior within the same social group.

Our contribution

- Study inter-ethnic relationships. Study pro-social behavior and network formation and the role of homophily and language skills.
- We show that in-group favoritism is displayed only by children of the majority group.

Assimilation and integration of ethnic minorities

- Quality of immigrant cohorts (Borjas, 1985), country of origin (Borjas, 1987, 1992), name change (Fryer and Levitt , 2005; Arai and Thoursie, 2009), ethnic concentration (Edin et al., 2003; Lazear, 1999) impact assimilation.
- Intermarriage rate (Meng and Gregory, 2005), English skill (Chiswick and Miller, 1995, 1996; Dustmann and Fabbri, 2003) impact assimilation.

Our contribution

- Study assimilation for early childhood.
- Focus on inter-ethnic friendship relations and language skills.
- Show how language fluency reduces preference bias.

Road Map

1. Background
2. Field Setting & Experimental Design
3. Reduced-Form Results
4. Model & Structural Estimation
5. Counterfactual policies

Background

- It has been almost a decade since the Syrian conflict started generating huge refugee waves (7 million).
- Syrian refugees have been hosted largely by neighbouring countries: Turkey, Lebanon, and Jordan.
- There is neither a massive return movement nor systematic repatriation efforts—mostly due to lack of an inclusive political environment and political stability in Syria.
- The social integration of Syrian refugees is a major policy issue for host countries.

Background

- Turkey, has received 3.6 million Syrian refugees since the start of the Syrian Civil War. About 1 million are school-aged children (5-7); about 64% were enrolled in 2019-20.
- Initially, education services were provided in refugee camps and Temporary Education Centers (TECs), which served as a second-shift education in regular public school buildings.
- By 2016, the system shifted to fully integrate Syrian children to the Turkish public education system
- The focus has also shifted toward the early childhood age group.

The 2019 early childhood summer school program in Turkey

- 9-week program (July-August 2019) implemented in 26 provinces and 500 schools.
- Target group: 5 year-old Turkish and Syrian children, who were going to start primary school in the Fall semester of the 2019/20 academic year.
- 37,153 students—19,110 (around 52%) refugees.
- Curriculum is prepared by Ministry of National Education (MoNE)—UNICEF also supported.
- Around 2,000 MoNE teachers are assigned; and paid from the MoNE budget.
- Nutritious snacks are provided to incentivize participation of disadvantaged children.

Goals of the Program

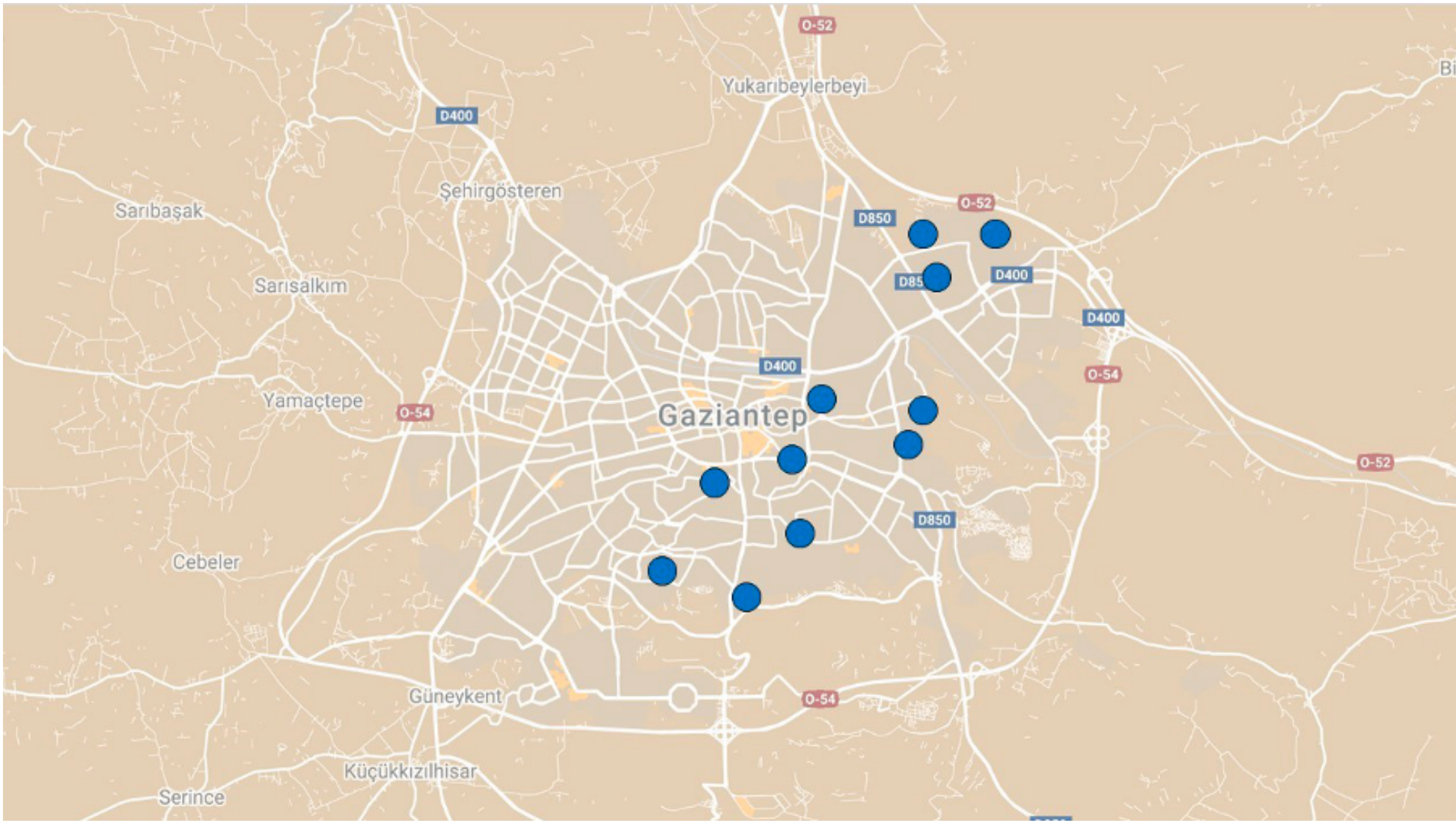
- Goal #1: provide early childhood education to disadvantaged kids.
- Goal #2: improve social integration by bringing together Turkish and Syrian kids of pre-school age.
- Goal #3: achieve educational integration by adapting new-entrant Syrian kids to the Turkish education system.
- Goal #4: improve Turkish language skills & familiarize Syrian kids with the Turkish culture.
- Goal #5: prepare Turkish children and their parents to a multi-cultural education environment.

Experimental design: schools

- The field experiment is implemented in Gaziantep, which has a large refugee population. It is also a major implementation hub for the program.
- Gaziantep is a city in the Southeastern Anatolia Region of Turkey, about 100 km north of Aleppo. Population estimated to be close to 2 million of which about half million are Syrian refugees.
- The program is advertised in refugee-intensive areas by outreach activities of MoNE and partners in the field.
- A total of 4,634 five-year-olds (2,188 refugees and 2,446 natives) were registered to the program in Gaziantep in 121 schools.
- We identified 46 schools with almost 50 percent refugee concentration. Then, we randomly selected 11 schools to implement our experiment.

Location of Schools

Location of Schools



Experimental design: sample

- In those 11 schools, students were randomly assigned to 36 groups before the program began.
- The groups vary in size from 10 to 24, with an average size of 16.9. There is a balanced mix of classrooms with high, medium, and low refugee shares.
- Switching between groups and/or schools was not allowed in our sample.
- Overall, 604 students completed the program—328 refugees and 276 natives, 310 males and 294 females. The attrition rate was 5 percent.

Outcomes

- Collected twice: in the beginning and at the end of the program.
- Turkish language skills: measured on a six-category scale as instructed in the program.
- Friendship network: students are asked to report their best friends (up to 5) in the classroom and rank them.
- Pro-social behavior: measured by a Sticker Game (version of a Dictator Game), which asks the number of stickers (out of 10 the student owns) the student would share with an anonymous classmate in the classroom.

We also collected

- Information about parents (education, occupation, place of origin and date of arrival for Syrians)

Summary statistics

Summary statistics: Outcome variables (means)

	All sample			Syrian			Turkish		
	Base	**End**	**Diff**	**Base**	**End**	**Diff**	**Base**	**End**	**Diff**
Turkish language skills	-	-	-	2.33	4.02	1.69	-	-	-
Stickers given	3.84	4.18	0.33	3.65	4.17	0.52	4.07	4.19	0.11
# of friends	2.40	2.72	0.32	2.51	2.82	0.31	2.27	2.59	0.32
# of Syrian friends	1.42	1.56	0.15	2.02	2.21	0.19	0.70	0.79	0.09
# of Turkish friends	0.99	1.15	0.17	0.49	0.61	0.12	1.58	1.80	0.23
Number of obs.		604			328			276	

Summary statistics

- Both Turkish and Syrian children make more friends during the course of the summer camp.
- Syrian children report slightly more total number of friends (2.8 vs 2.6).
- Syrian children have a slightly higher share of friends who are own ethnicity (80% vs 74%) than Turkish children.
- Language skills: there is a significant improvement (72%) during the course of the summer camp.
- Sticker game: increase in sharing for both Syrian and Turkish kids.

Empirical strategy

The main regression equation is specified as follows:

$$y_{ig}^2 = \alpha_s + \beta q_g + \gamma y_{ig}^1 + \lambda_1{}' x_{ig} + \lambda_2{}' z_{ig}^p + \epsilon_{ig}, \tag{1}$$

- y_{ig}^2 is the outcome of child i in group g measured in the endline,
- q_g is the share of (Syrian) refugees in group $g = 1, \cdots 36$,
- y_{ig}^1 is the same outcome measured at baseline,
- α_s is a school fixed effect,
- x_{ic} and z_{ic}^p are children (gender, ethnicity, number of siblings) and parent (education, employment status, ethnic composition of friends and neihgborhood; for Syrian parents we also control for region of origin in Syria, year of arrival in Turkey, and Turkish language skills) specific controls.
- Cluster standard errors at the classroom level.

Empirical strategy

- To examine possible nonlinear effects of the refugee share on the various outcomes, we also estimate a specification of the following form:

$$y_{ig}^{2} = \alpha_s + \sum_{q=2}^{3} \beta_q t_{qg} + \gamma y_{ig}^{1} + \theta' x_{ig} + \lambda' z_{ig}^{p} + \epsilon_{ig}, \tag{2}$$

where t_{qg} are indicator variables that denote the tercile of the refugee share distribution.

Prosocial Behavior: Sticker game (All kids)

Table: Stickers given to random classmate (All sample)

	[1]	[2]	[3]
Panel A - Linear Specification			
Refugee share	-0.022***	-0.019***	-0.024***
	(0.006)	(0.006)	(0.005)
Sticker Baseline		0.223***	0.226***
		(0.037)	(0.036)
Panel B - Tercile Specification			
Refugee share (Med)	-0.889***	-0.798***	-0.872***
	(0.285)	(0.233)	(0.228)
Refugee share (High)	-1.209***	-1.020***	-1.360***
	(0.384)	(0.336)	(0.323)
Sticker Baseline		0.224***	0.227***
		(0.036)	(0.035)
School fixed effects	Yes	Yes	Yes
Controls	No	No	Yes
# of obs.	604	604	604

Prosocial Behavior: Sticker game (All kids)

- We find a negative effect of the share of refugees on the sharing that is statistically significant (linear specifications).
- The effect is mildly nonlinear, as the negative effect of being in a High Refugee share group (relative to being in a Small Refugee share group) is larger (in absolute value).

Prosocial Behavior: Sticker game (Turkish kids)

Table: Stickers given to random classmate (Turkish kids)

	[1]	[2]	[3]
Panel A - Linear Specification			
Refugee share	-0.035***	-0.030***	-0.030***
	(0.007)	(0.006)	(0.007)
Sticker Baseline		0.209**	0.190**
		(0.079)	(0.085)
Panel B - Tercile Specification			
Refugee share (Med)	-0.958***	-0.865***	-0.936***
	(0.302)	(0.267)	(0.251)
Refugee share (High)	-2.124***	-1.845***	-1.896***
	(0.555)	(0.512)	(0.595)
Sticker Baseline		0.214***	0.194**
		(0.078)	(0.084)
School fixed effects	Yes	Yes	Yes
Controls	No	No	Yes
# of obs.	276	276	276

Prosocial Behavior: Sticker game (Syrian kids)

Table: Stickers given to random classmate (Syrian kids)

	[1]	[2]	[3]
Panel A - Linear Specification			
Refugee share	-0.012	-0.013	-0.018*
	(0.009)	(0.007)	(0.010)
Sticker Baseline		0.240***	0.235***
		(0.046)	(0.045)
Panel B - Tercile Specification			
Refugee share (Med)	-0.821**	-0.793**	-0.531
	(0.390)	(0.321)	(0.376)
Refugee share (High)	-0.760	-0.723	-0.884*
	(0.494)	(0.428)	(0.517)
Sticker Baseline		0.238***	0.232***
		(0.045)	(0.043)
School fixed effects	Yes	Yes	Yes
Controls	No	No	Yes
# of obs.	328	328	328

Prosocial Behavior: Sticker game (Turkish versus Syrian kids)

- The negative effect is mainly driven by Turkish kids.
- Both the linear and the tercile specifications: Negative and statistically significant effect of refugee share on the sharing of Turkish children.
- The coefficient of the high refugee share is more negative than that of the medium share group.
- Magnitude (column 3): A Turkish kid in a high refugee share group gives 1.9 stickers less than a kid in a low refugee share group, which is 45 percent of what Turkish kids give on average.

- For the Syrian kids: the effect is small, even not statistical significance. This indicates that Syrian kids do not ethnically discriminate in sharing.

Friendship regressions (Turkish kids)

	# of Turkish Friends			Share of Turkish Friends		
	[1]	[2]	[3]	[4]	[5]	[6]
Panel A - Linear Specification						
Refugee share	-0.014**	-0.011**	-0.012**	-0.007***	-0.006***	-0.007***
	(0.006)	(0.005)	(0.005)	(0.002)	(0.002)	(0.002)
Turkish friends Wave 1		0.216***	0.215***			
		(0.059)	(0.062)			
Share of Turkish friends Wave 1					0.189***	0.196***
					(0.054)	(0.060)
Panel B - Tercile Specification						
Refugee share (Med)	-0.487**	-0.452**	-0.442**	-0.192***	-0.171***	-0.207***
	(0.189)	(0.171)	(0.167)	(0.065)	(0.054)	(0.059)
Refugee share (High)	-0.864*	-0.692	-0.629	-0.248*	-0.194	-0.228*
	(0.489)	(0.417)	(0.400)	(0.143)	(0.123)	(0.113)
Turkish friends Wave 1		0.218***	0.220***			
		(0.055)	(0.060)			
Share of Turkish friends Wave 1					0.210***	0.212***
					(0.054)	(0.060)
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	No	No	Yes
# of obs.	276	276	276	276	276	276

Table: Standard errors are clustered at classroom level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Friendship regressions (Turkish kids)

- Turkish children (linear specification): Negative effect of exposure to refugee children, on both outcomes.
- Magnitude (column 3): A one standard deviation increase in refugee share leads to a 0.24 decrease in the number of Turkish friends (the mean number of Turkish friends for Turkish kids is 1.8),
- Magnitude (column 6): A one standard deviation increase in refugee share leads to a 0.14 decrease in the share of Turkish friends (the mean share of Turkish friends for Turkish kids is 0.70).

Friendship regressions (Syrian kids)

	# of Syrian Friends			Share of Syrian Friends		
	[1]	[2]	[3]	[4]	[5]	[6]
Panel A - Linear Specification						
Refugee share	0.022**	0.012	0.011	0.004*	0.003	0.003
	(0.009)	(0.008)	(0.008)	(0.002)	(0.002)	(0.002)
Syrian friends Wave 1		0.320***	0.312**			
		(0.068)	(0.073)			
Share of Syrian friends Wave 1					0.212***	0.210***
					(0.051)	(0.057)
Panel B - Tercile Specification						
Refugee share (Med)	0.714**	0.510**	0.420**	0.196**	0.155*	0.147**
	(0.295)	(0.215)	(0.204)	(0.088)	(0.078)	(0.065)
Refugee share (High)	0.578	0.058	-0.014	0.216**	0.137	0.133
	(0.423)	(0.289)	(0.300)	(0.105)	(0.102)	(0.094)
Syrian friends Wave 1		0.346***	0.332***			
		(0.066)	(0.071)			
Share of Syrian friends Wave 1					0.216***	0.217***
					(0.050)	(0.057)
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	No	No	Yes
# of obs.	328	328	328	328	328	328

Table: Standard errors are clustered at classroom level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Friendship regressions (Syrian kids)

- Syrian children: The share of refugees in the group exerts a positive but weaker effect on the number and share of own-ethnicity friends.
- Linear specification: No statistically significant effect for either outcome.
- In the tercile specification, we find some evidence of a statistically significant difference comparing medium to low refugee share groups.
- Being in the medium refugee share group is associated with an increase in the number of Syrian friends of 0.4, when on average Syrian kids report 2.2 Syrian friends.

Friendship regressions

- Analysis of friendship networks: the ethnic composition of one's Summer camp group does influence the ethnic composition of one's friendship network (contact hypothesis). Mainly for Turkish kids.
- Next figures plot the homophily index (share of own friends) against the size of the refugee group in the class, for Syrian and Turkish children.

Friendship Formation

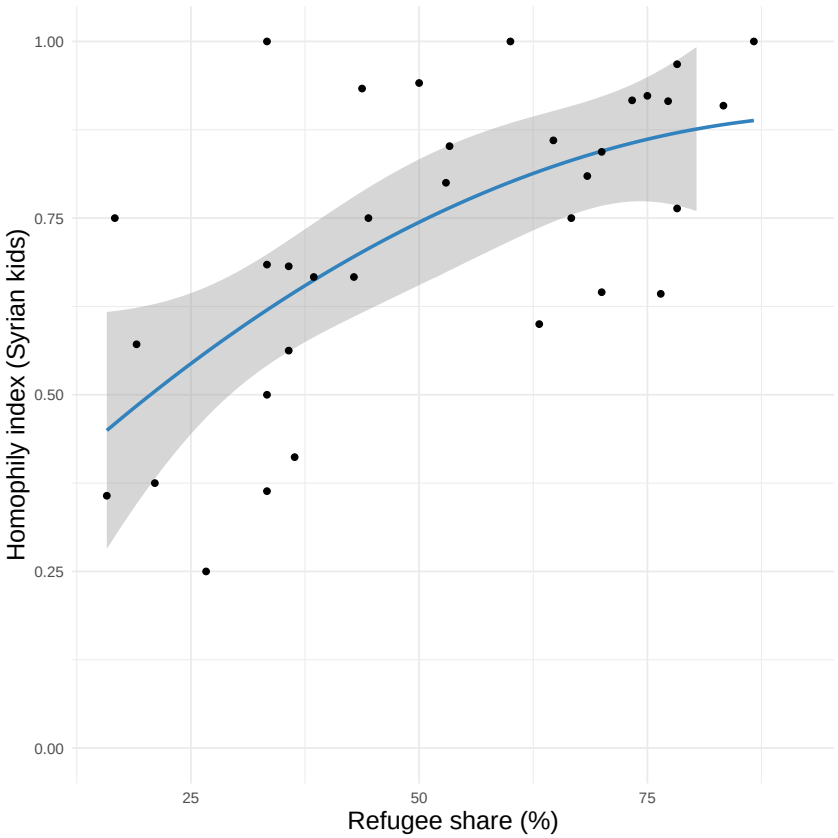


Figure: Homophily index by refugee share (Syrian kids).

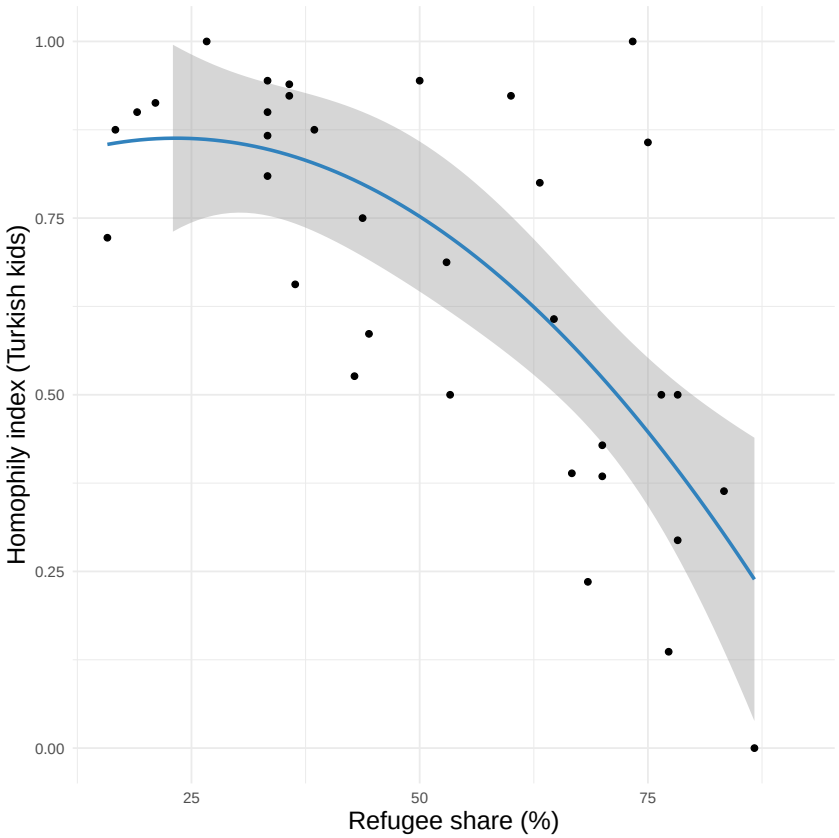


Figure: Homophily index by refugee share (Turkish kids).

Language skill regressions

- We next investigate whether there are differences in language skill acquisition for Syrian children, depending on the share of refugees in their group.

- Outcome: binary variable defined on whether the kid has high or low skills.
- Language skills are assessed by the teacher, and are reported in a scale from 1 to 6.
- We define as high skill children who are assessed to be 4 or higher (about 60% of the sample in the endline, and 25% in the baseline).

Language skill regressions

Table: Language (Syrian kids)

	[1]	[2]	[3]
Panel A - Linear Specification			
Refugee share	-0.005***	-0.005***	-0.004**
	(0.001)	(0.001)	(0.002)
Language Baseline		0.420***	0.392***
		(0.053)	(0.051)
Panel B - Tercile Specification			
Refugee share (Medium)	-0.114	-0.126*	-0.146
	(0.086)	(0.072)	(0.095)
Refugee share (High)	-0.352***	-0.367***	-0.299***
	(0.084)	(0.067)	(0.092)
Language Baseline		0.420***	0.395***
		(0.054)	(0.052)
School fixed effects	Yes	Yes	Yes
Controls	No	No	Yes
# of obs.	328	328	328

Language skill regressions

- Positive and robust association between level of language skills at baseline and endline.
- The effect of refugee share is negative and statistically significant.
- The tercile specification indicates that the effect is non-linear, as there is a difference in language skills for children in high refugee share relative to those with medium or low shares.
- The high share group is 30 percentage points less likely to be high skill than those in the low share group, which indicates a significant reduction over the baseline of 60%.

Summary Field Experiment

- Pro-social behavior: Sticker game
  - Turkish kids give less the higher is the fraction of Syrian kids in their group.
  - A Turkish kid in a high refugee share group gives 1.9 stickers less than a kid in a low refugee share group.
  - In terms of sharing/giving, Syrian kids are not affected by the fraction of Syrian kids in their group.
- Friendship formation
  - Turkish kids are less friends with other Turkish kids the higher is the fraction of Syrian kids in their group.
  - A one standard deviation increase in refugee share leads to a 0.24 and 0.14 decrease in the number and fraction of Turkish friends.
  - Syrian kids are not affected by the fraction of Syrian kids in their group.
- Language skills
  - The higher is the fraction of Syrian kids in a classroom, the lower is the development of Turkish language skills among the refugee children.

Theory

- Population of n students in a classroom, n^T Turkish, n^S Syrian, with $n^T + n^S = n$.
- For a student i, we denote by $e(i) = S, T$ the ethnicity i.
- Share of Syrian children in the classroom: $q = n^S/n$, so $1 - q$ is the share of Turkish children.
- Simple network formation model featuring two key driving forces: preference biases and congestion.
- Congestion implies that children's number of friends does not mechanically increase with the classroom size.

Network formation

- The probability p^{ij} that child i befriend child j is given by:

$$p^{ij} = \frac{\delta^{ij} s^i s^j}{\sum_{k \in T^i(j)} s^k}. \tag{3}$$

- δ^{ij} is the preference bias of i regarding j. The set $T^i(j)$ reflects congestion.
- Congestion: set of children (excluding i but including j), that are comparable with j from i's point of view (e.g. children of the same ethnicity or gender as j).
- The higher is i's socialization effort $s^i \in [0, 1]$, the more likely a link p^{ij} will be formed between i and j.
- Congestion effect $\sum_{k \in T^i(j)} s^k$: the impact of j's socialization effort is reduced when the socialization efforts of children comparable to j (from the point of view of i) increases.

Socialization choice

- Utility of child i, choosing socialization effort s^i:

$$u^i = b^i s^i - \frac{1}{2}\left(s^i\right)^2 - \frac{\kappa^i}{2}\left(s^i - \overline{s}^{-i}\right)^2 \tag{4}$$

- b^i is the ex ante heterogeneity (observable characteristics such as gender, ethnicity, parents' education, etc.) of individual i.
- $\overline{s}^{-i}$ is the average socialization effort in the classroom *leaving out* i,
- $\kappa^i > 0$ (taste for conformity), ethnic specific (κ^T and κ^S).

Socialization choice: Equilibrium

First-order condition for child i:

$$s^i = (1 - \phi^i) b^i + \phi^i \overline{s}^{-i} \tag{5}$$

where $\phi^i := \kappa^i / \left(1 + \kappa^i\right)$.

Socialization choice

- Assume that all Turkish students have the same characteristics and all Syrian students have the same characteristics, i.e., and $b^i = b^S$ for all Syrian children and $b^i = b^T$ for all Turkish children.

- Only two socialization effort levels: s^{T*} and s^{S*}.

- Social norm in the classroom: $\bar{s} = qs^S + (1-q)s^T$, where $q = n^S/n$.
- For $e = S, T$,

$$\bar{s}^e = \bar{s} - \frac{s^e}{n},$$

- First-order condition:

$$s^e = (1-\phi^e)b^e + \phi^e\left(qs^S + (1-q)s^T - \frac{s^e}{n}\right).$$

Socialization choice

- When n is large,

$$s^{S*} = \frac{\left[1-\phi^T(1-q)\right]\left(1-\phi^S\right)}{1-\phi^T-q\left(\phi^S-\phi^T\right)}b^S + \frac{\phi^S(1-q)\left(1-\phi^T\right)}{1-\phi^T-q\left(\phi^S-\phi^T\right)}b^T,$$

$$s^{T*} = \frac{\left(1-\phi^S q\right)\left(1-\phi^T\right)}{1-\phi^T-q\left(\phi^S-\phi^T\right)}b^T + \frac{\phi^T q\left(1-\phi^S\right)}{1-\phi^T-q\left(\phi^S-\phi^T\right)}b^S,$$

- When $\phi^S = 1$, $s^{S*} = \overline{s}^* = b^T$, which is independent of q.
- When $\phi^T = 1$, $s^{T*} = \overline{s}^* = b^S$, which is independent of q.
- When $\phi^S = 0$, $s^{S*} = b^S$, which is independent of q.
- When $\phi^T = 0$, $s^{T*} = b^T$, which is independent of q.

Friendship links

- Assume that $T^i(j)$ includes all children with the same ethnicity as j (except i) and that δ^{ij} only depends on the children's ethnicity.

$$
\begin{aligned}
p^{TT*} &= \frac{\delta^{TT} s^{T*}}{n\left(1-q-1/n\right)}; \\
p^{TS*} &= \frac{\delta^{TS} s^{T*}}{nq}; \\
p^{SS*} &= \frac{\delta^{SS} s^{S*}}{n\left(q-1/n\right)}; \\
p^{ST*} &= \frac{\delta^{ST} s^{S*}}{n\left(1-q\right)}.
\end{aligned}
$$

- For example, consider the probability of friendship between two Turkish kids, i.e. p^{TT}.
- When a Turkish kid wants to form links with other Turkish kids, there are exactly $n(1-q)-1$ other Turkish kids in the classroom and, therefore, the congestion for this link formation is equal to $[n(1-q)-1]\, s^T$.

Comparative statics

Proposition (Comparative Statics)

The following testable implications hold:

1. *When n is large, s^{S*} and s^{T*} decrease with q iff $b^S \leq b^T$;*
2. *If $b^S \leq b^T$, then p^{SS*} and p^{TS*} decrease with q*
3. *If $b^S \geq b^T$, then p^{TT*} and p^{ST*} increase with q*

- Reduced form estimation: The number of stickers (imperfect measure of socialization effort) given to a random classmate is decreasing in the share of Syrian kids (particularly for Turkish kids).
- Model: This is because $b^S < b^T$ (for Turkish kids) and ϕ^S close to 1 for Syrian kids.
- Implies that Syrian kids are less prone to socializing.
- Classrooms with higher share of Syrian kids therefore have smaller socialization norms, which in turns implies a smaller equilibrium socialization for everyone.

Homophily

- Homophily = share of a own ethnicity links, for each ethnic group. For ethnicity $e = S, T$,

$$H^e = \frac{\sum_{ij:e(i),e(j)=e} g_{ij}}{\sum_{ij:e(i)=e} g_{ij}}, \tag{6}$$

- where $g_{ij} = 1$ if i and j are friends and zero otherwise.

- For example, for Syrian kids, this is the number of Syrian-Syrian friendship relations divided by the number of Syrian-Syrian and Syrian-Turkish friendships.
- Because link formation is random, *ex-ante* homophily indices H^S and H^T are random variables for which is hard to describe the behavior as q changes.
- We can derive analytical approximations of the expected values $\mathbb{E}\left[H^S\right]$ and $\mathbb{E}\left[H^T\right]$.

Homophily

Proposition

For any $q \in \left(\frac{2}{n}, \frac{n-2}{n}\right)$*, using a second order Taylor expansion, we have:*

$$\mathbb{E}\left[H^S\right] = \frac{\delta^{SS}}{\delta^{SS}+\delta^{ST}} - \frac{\delta^{SS}\delta^{ST}}{(\delta^{SS}+\delta^{ST})^3 n^2}\left[\frac{\delta^{ST}}{q(1-q)} - \frac{\delta^{SS}}{q(q-1/n)}\right].$$

$$\mathbb{E}\left[H^T\right] = \frac{\delta^{TT}}{\delta^{TT}+\delta^{TS}} - \frac{\delta^{TT}\delta^{TS}}{\left(\delta^{TT}+\delta^{TS}\right)^3 n^2}\left[\frac{\delta^{TS}}{q(1-q)} - \frac{\delta^{TT}}{(1-q)(1-q-1/n)}\right],$$

where the expansion is done around $\frac{\delta^{SS}}{\delta^{SS}+\delta^{ST}}$ *for* H^S *and around* $\frac{\delta^{TT}}{\delta^{TT}+\delta^{TS}}$ *for* H^T*. This implies that:*

$$\frac{\partial \mathbb{E}\left[H^S\right]}{\partial q} > 0 \Leftrightarrow \frac{\delta^{SS}}{\delta^{ST}} < \frac{(1-2q)\left(q-1/n\right)^2}{\left(1-q\right)^2\left(2q-1/n\right)} \approx \frac{(1-2q)q}{2(1-q)^2}$$

$$\frac{\partial \mathbb{E}\left[H^T\right]}{\partial q} < 0 \Leftrightarrow \frac{\delta^{TT}}{\delta^{TS}} < \frac{(2q-1)\left(1-q-1/n\right)^2}{q^2(2(1-q)-1/n)} \approx \frac{(2q-1)(1-q)}{2q^2}$$

Structural estimation: Empirical Strategy

- First, we estimate a model of socialization choices using the share of stickers given in the baseline as a proxy for the socialization choices.
- Second, we estimate the network formation model using the realized network in the endline, controlling for the network in the baseline.
- To ensure a valid inference of the resulting two-stage estimator, the estimations are performed jointly, following Arellano and Meghir (1992).

Structural estimation: Children's socialization choices

- Denote by r the classroom with $r = 1, ..., \bar{r}$, where $\bar{r} = 36$ is the number of classrooms in our dataset.
- From the theoretical model, the socialization choices for all i and r are given by:

$$s_r^i = b_r^i[d_r^i(1-\phi^S)+(1-d_r^i)(1-\phi^T)]+d_r^i\frac{\phi^S}{n_r-1}\sum_{j\neq i}s_r^j+(1-d_r^i)\frac{\phi^T}{n_r-1}\sum_{j\neq i}s_r^j,$$

- where d_r^i is a binary variable equal to 1 if i is Syrian and 0 otherwise.
- We assume that $b_r^i = \mathbf{x}_r^i\boldsymbol{\beta} + \epsilon_r^i$, where $\mathbf{x}_r^i$ is a vector of characteristics of child i in classroom r (e.g. gender, school dummy, etc.), and where $\epsilon_r^i \sim N(0, \sigma^2)$.

Structural estimation: Children's socialization choices

- Since we do not observe the children's socialization effort s_r^i, we assume that

$$s_r^i = \tilde{s}_r^i + \eta_r^i,$$

- where $\tilde{s}_r^i$ is the fraction of stickers that i gave to a random kid in class r during the baseline,
- where η_r^i is the approximation error.
- Importantly, we make no parametric assumption on η_r^i and only assume that it is exogenous, i.e. $\mathbb{E}[\eta_r^i|\mathbf{X}] = 0$ for all i.
- We instrument the endogenous variable $\frac{1}{n_r-1}\sum_{j\neq i} s_r^j$ (the average socialization levels of the $n_r - 1$ other kids in the classroom) using $\frac{1}{n_r-1}\sum_{j\neq i} \mathbf{x}_r^j$ (the average characteristics of the $n_r - 1$ other kids in the classroom).

Structural estimation: Formation of the friendship network

- For any $j \neq i$ and r, the probability that i and j are friends during the endline is given by:

$$p_r^{ij} = \frac{\delta_r^{ij} s_r^i s_r^j}{\sum_{k \in T_r^i(j)} s_r^k} \tag{7}$$

- $\delta_r^{ij} = \Phi(\mathbf{z}_r^{ij}\gamma)$, where Φ standard normal distribution,
- $\mathbf{z}_r^{ij}$ vector of characteristics of the pair ij (e.g. if i and j are of the same ethnicity, or if they were friends during the baseline),
- γ is vector of unknown parameters.
- $T_r^i(j)$ is the set of all students (except i) who have the same ethnicity, gender and link status in the baseline as j, in classroom r.

Structural estimation: Results

	Estimator	Std. Err.
Panel A: Socialization effort ($\boldsymbol{\beta}$,ϕ^S,ϕ^T)		
Syrian	-0.741	(0.065)
Fluent in Turkish (Syrians)	0.140	(0.084)
ϕ^S	0.955	(0.003)
ϕ^T	0.694	(0.001)
Panel B: Network formation ($\boldsymbol{\gamma}$, marginal effects on δ_r^{ij})		
Syrian-Syrian pair	0.127	(0.017)
Syrian-Turkish pair	-0.212	(0.004)
Turkish-Syrian pair	-0.218	(0.013)
Fluent in Turkish (Syrians)	0.115	(0.013)
Linked in wave 1	0.235	(0.102)

Notes: Socialization choices control for: Gender, composition of the parents' friends and neighborhood, arrival year (Syrians), region in Syria (Syrians), Turkish skills (Syrians) and School fixed effects. Network formation controls for Gender, composition of the parents' neighborhood and School fixed effects. The estimates for the network formation panel correspond to the average marginal effects on δ^{ij}. That is, $(1/N_2)\sum_{i,j}[\Phi(\sum_{k'\neq k} z^{ij}_{r,k'}\gamma_{k'} + \gamma_k) - \Phi(\sum_{k'\neq k} z^{ij}_{r,k'}\gamma_{k'})]$ for the binary variable k. Standard errors are simulated using 500 draws of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \phi^S, \phi^T, \sigma)$ using its estimated asymptotic distribution. This is done in order to account for the covariance between the estimated values for $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ and ϕ.

Structural estimation: Results

- Syrian children socialize less than Turkish children.
- (model interpretation) Since Syrians socialize less, everything else equal, classrooms with higher shares of Syrians have lower socialization norms, which leads all children in those groups to socialize less.
- Syrian children have a stronger taste for conformism and Turkish children (i.e., $\phi^S = 0.955 > \phi^T = 0.694$).
- (model interpretation) This implies that the equilibrium socialization choices of Syrians are less impacted by the share of Syrians in the classroom. Consistent with reduced form results.

Structural estimation: Results

- Impact of the children's characteristics on the formation of the friendship network.
- Interethnic links: Syrian-Turkish and Turkish-Syrian, are less likely to be formed than homogeneous links (Syrian-Syrian or Turkish-Turkish links).
- Network formation positively affected by the language skills of Syrian children.
- Persistence: Friendship in the baseline strongly predicts friendship in the endline.

Structural estimation: Results

- Note that the reported numbers are not the marginal effects on the probability of a link, but rather marginal effects on the preference bias δ_r^{ij}.
- Inter-ethnic pairs have a value of δ^{ij} that is roughly 35 percentage points lower than that of same-ethnic pairs.
- Turkish kids are slightly more biased toward Syrian kids than vice-versa.
- Syrian children who are fluent in Turkish have a value of δ^{ij} that is 15 pp higher than those who are not.
- Pairs that were linked in the baseline have a 19 pp higher δ^{ij} than those that were not.

Structural estimation: Counterfactual policies

Homophily

- Impact of an exogenous change in the refugee share on the homophily index (share of same-ethnicity friends) as well as on inbreeding homophily.
- Let q^e be the share of children of ethnicity $e = T, S$ in the classroom, inbreeding homophily is defined as:

$$IH^e = (H^e - q^e)/(1 - q^e)$$

- When $H^e > q^e$, there is positive or inbreeding homophily, i.e., $IH^e > 0$.
- When $H^e < q^e$, there is negative or inbreeding heterophily, i.e., $IH^e < 0$.
- Promote integration of Syrian children: $IH^e = 0$ (i.e., $H^e = q^e$).

Counterfactual policies: Homophily

- In the data, we have only 36 classes, Major advantage of the structural estimation is that it allows us to simulate a much larger number of classes for the analysis of homophily to be more meaningful.

- In order to exogeneously vary q, we reshuffled children to classes using the following procedure (the procedure is repeated 500 times).
- (1) We randomly select classrooms until their combined capacity reached the number of Syrian children in our dataset.
- (2) We randomly select children and attribute them to classrooms, putting a weight of α on classrooms of their own ethnicity, and a weight of $1 - \alpha$ on the other classrooms (subject to capacity constraints).
- Then, by varying α, we can exogeneously change the expected share of refugee children in any given classroom, keeping everything else constant.

Counterfactual policies: Homophily

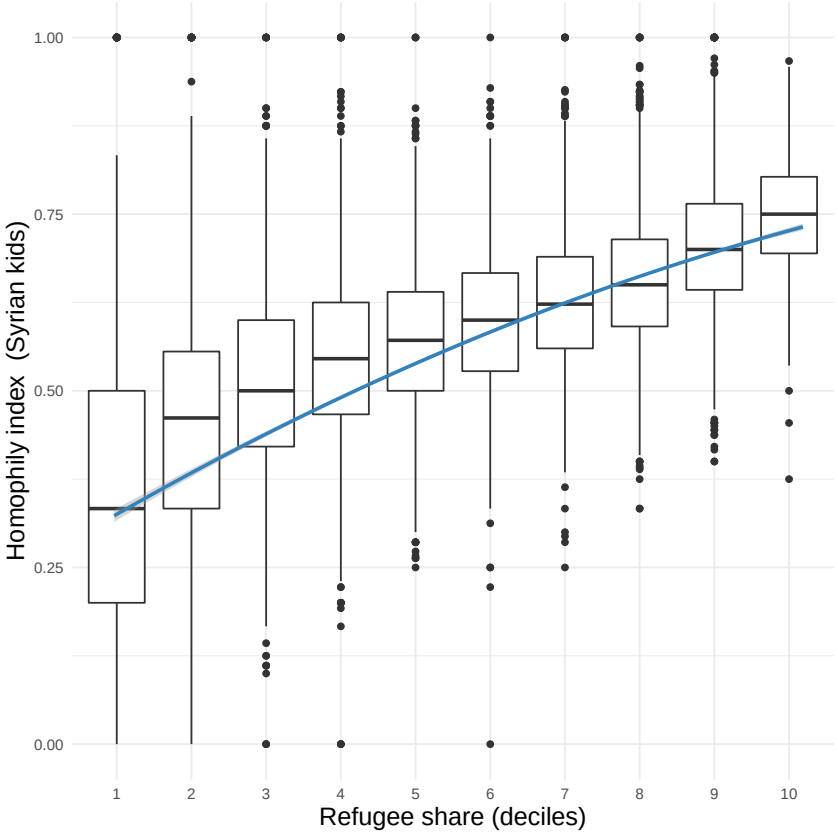


Figure: Simulated homophily index by refugee share (Syrian kids).

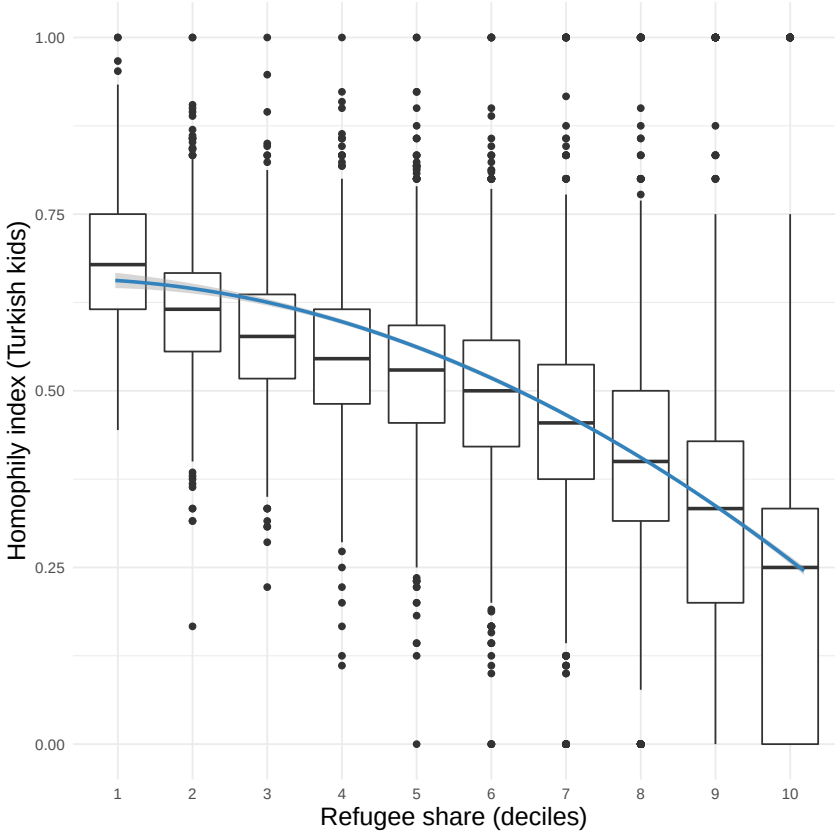


Figure: Simulated homophily index by refugee share (Turkish kids).

Data: Homophily

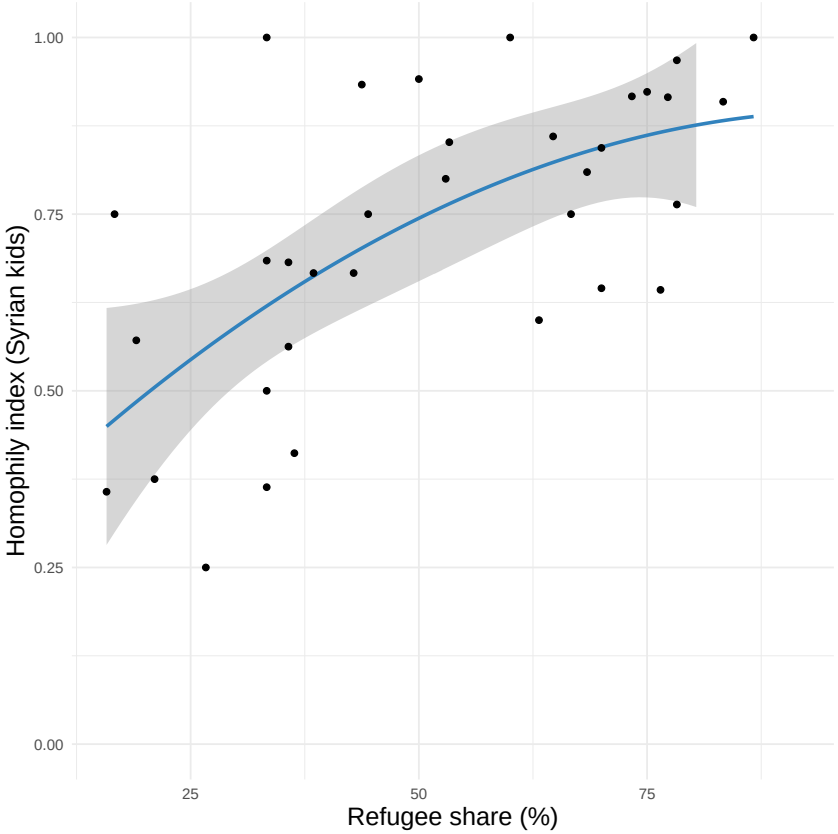


Figure: Homophily index by refugee share in the data (Syrian kids).

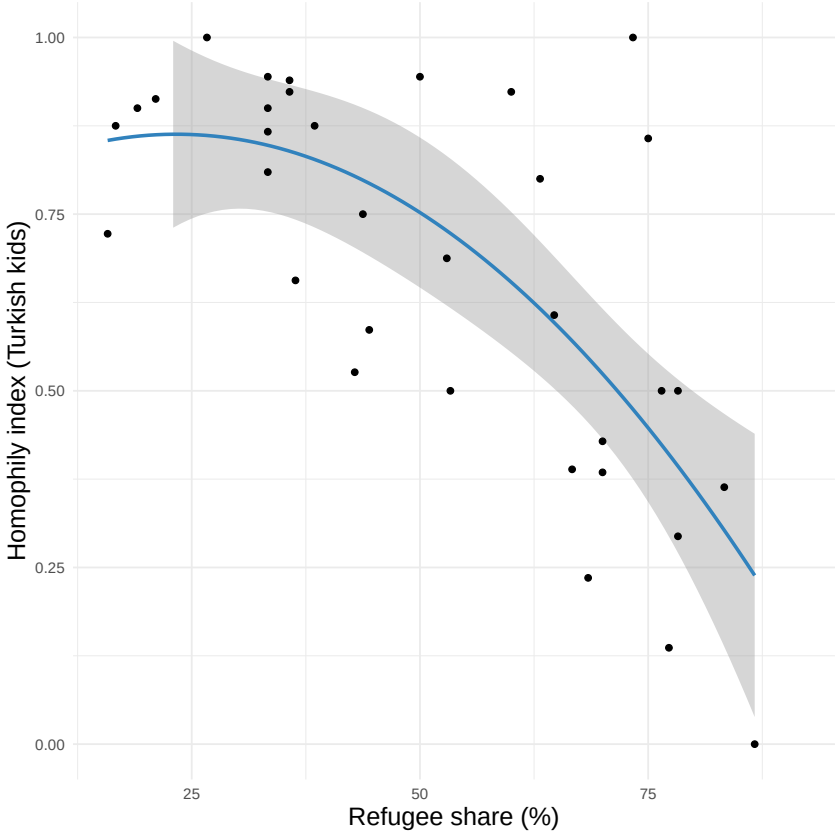


Figure: Homophily index by refugee share in the data (Turkish kids).

Counterfactual policies: Homophily

- Interplay between preference bias and congestion.
- Consider Syrian kids: Since they are postitively biased toward other Syrian kids, homophily increases with the share of Syrian kids in the classroom.
- However, the rate at which homophily increases is decreasing. This is due to congestion.
- As there are more and more Syrian kids in the classroom, they compete among themselves for friendship.
- Then, except when the share of Syrian kids is very low, homophily is not increasing as fast as the classroom share.
- We need to study inbreeding homophily.

Counterfactual policies: Inbreeding Homophily

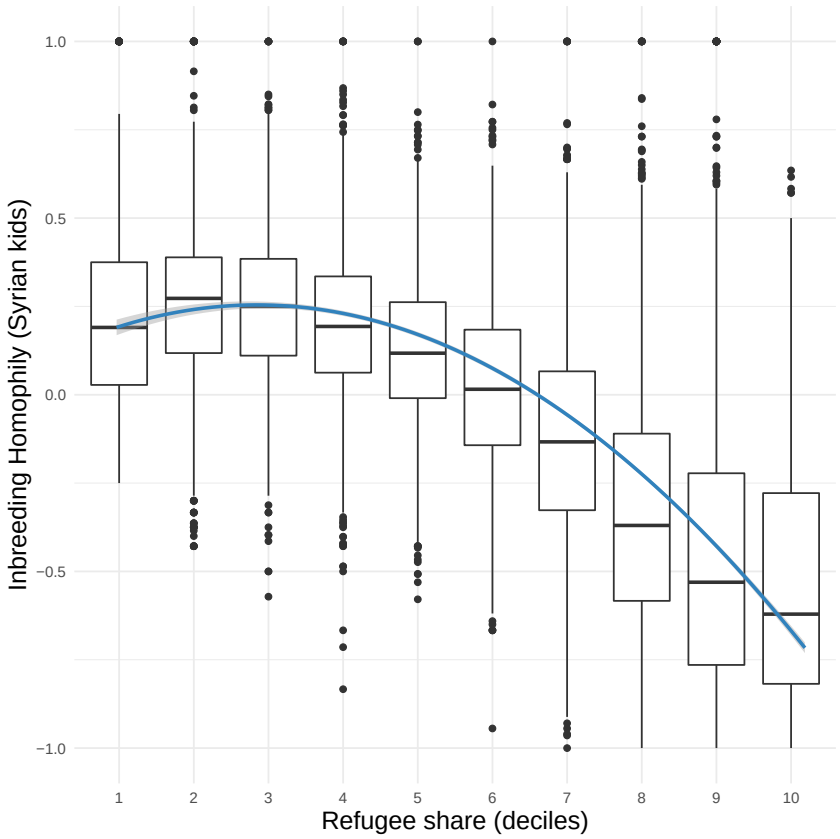


Figure: Simulated inbreeding homophily by refugee share (Syrian kids).

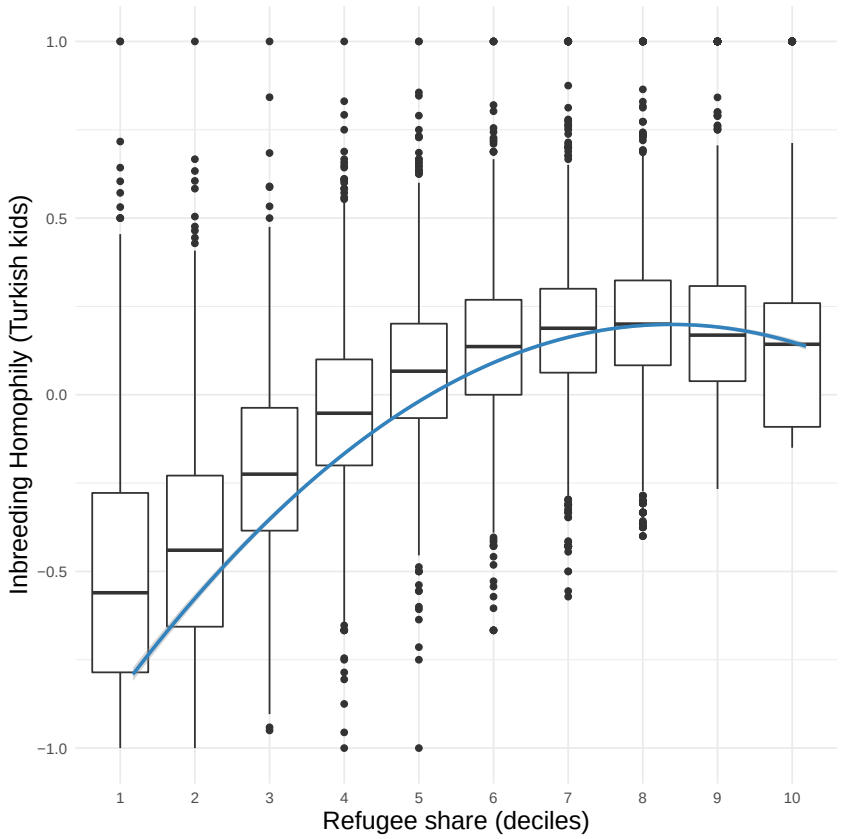


Figure: Simulated inbreeding homophily by refugee share (Turkish kids).

Data: Inbreeding Homophily

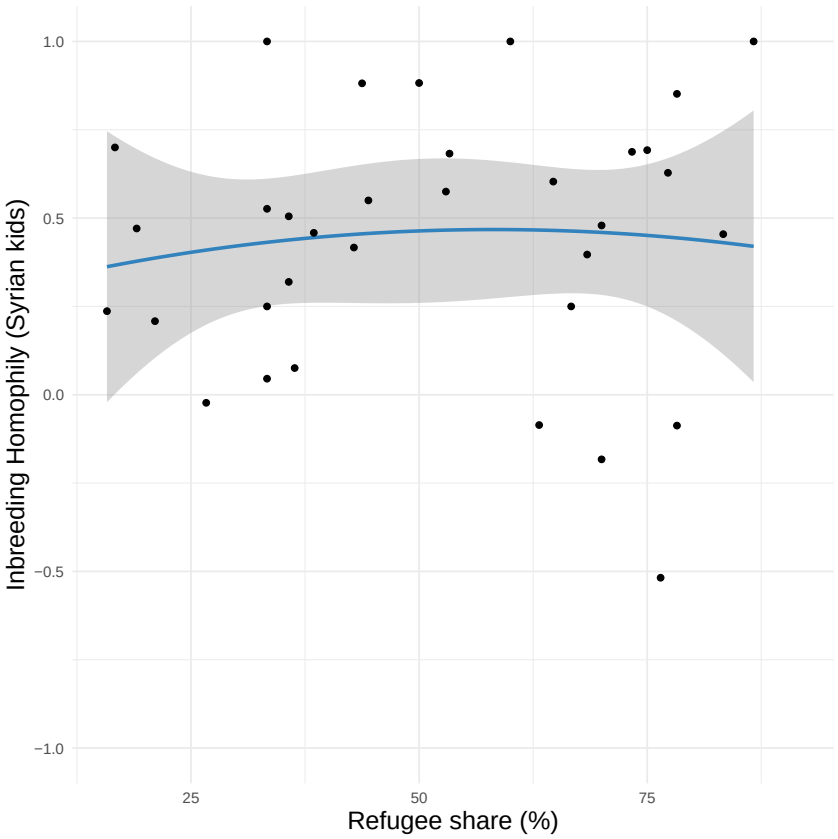


Figure: Inbreeding homophily by refugee share in the data (Syrian kids).

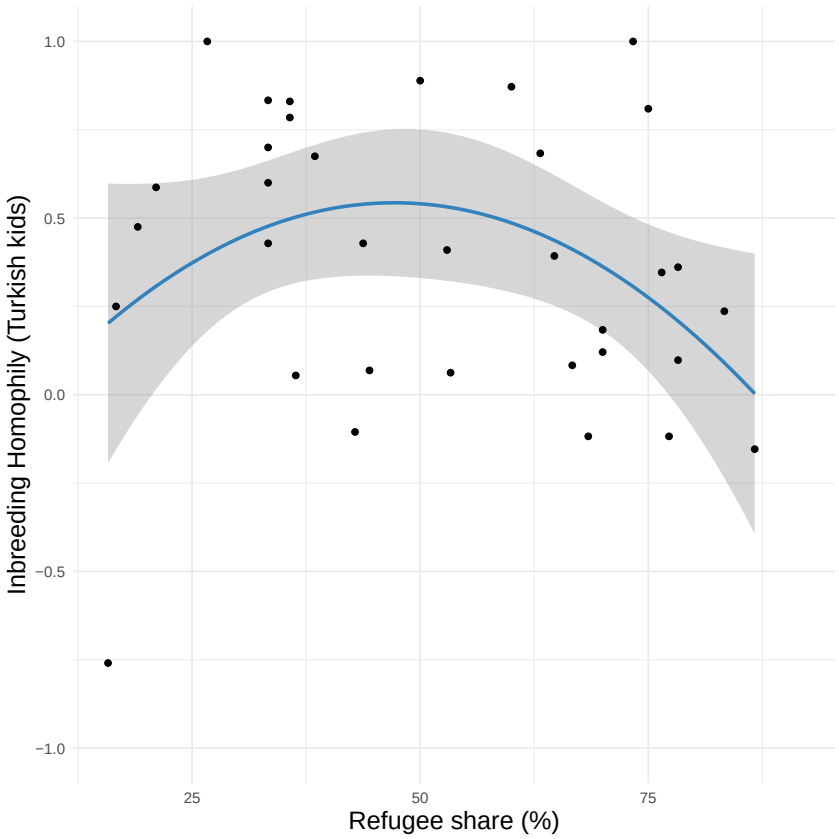


Figure: Inbreeding homophily by refugee share in the data (Turkish kids).

Counterfactual policies: Inbreeding Homophily

- Inbreeding homophily U-shape relationship with share of own ethnicity.
- Interplay between preference bias and congestion.
- $IH^S = 0$ (i.e., $H^S = q^S = q$) when q^S is in the 7th decile.
- $IH^T = 0$ (i.e., $H^T = q^T = 1 - q$) when q^S is in the $5 - 6^{th}$ decile.

Counterfactual policies: Preference bias

- We vary the strength of the ethnic-based preference biases.
- We look at two alternative specifications: one in which there is no ethnic bias, and one in which the ethnic biases are double that of the baseline level.

Counterfactual policies: Preference bias

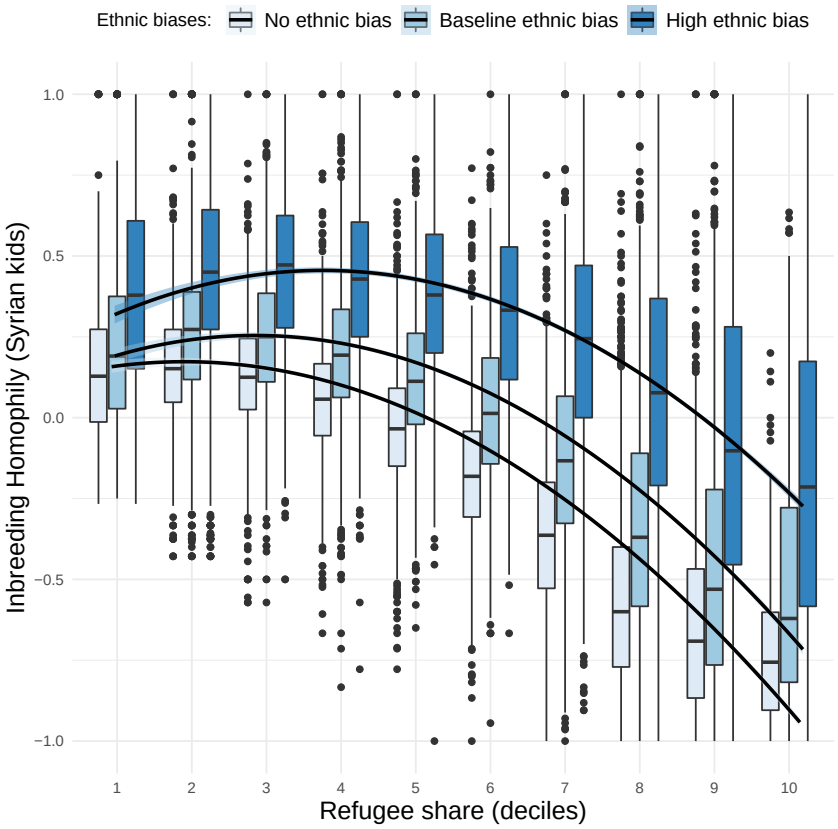


Figure: Simulated inbreeding homophily by refugee share (Syrian kids): Changes in preference biases.

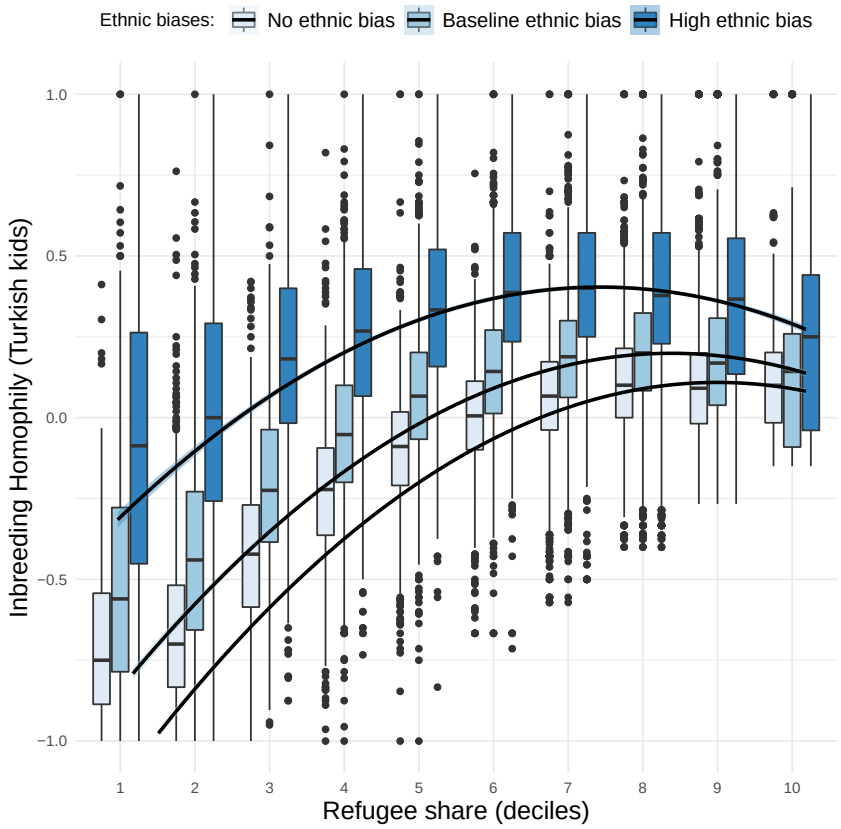


Figure: Simulated inbreeding homophily by refugee share (Turkish kids): Changes in preference biases.

Counterfactual policies: Preference bias

- As the preference biases increase, the impact of congestion in the network formation process becomes relatively less important.
- The equilibrium network exhibits very high levels of inbreeding homophily, even when the share of own-ethnicity children is high.
- Syrian (baseline): $IH^S = 0$ (i.e., $H^S = q^S$): q^S in 7th decile.
- Syrian (no ethnic bias): $IH^S = 0$ (i.e., $H^S = q^S$): q^S in 6th decile.
- Syrian (double ethnic bias): $IH^S = 0$ (i.e., $H^S = q^S$): q^S in 9th decile.

Counterfactual policies: Congestion

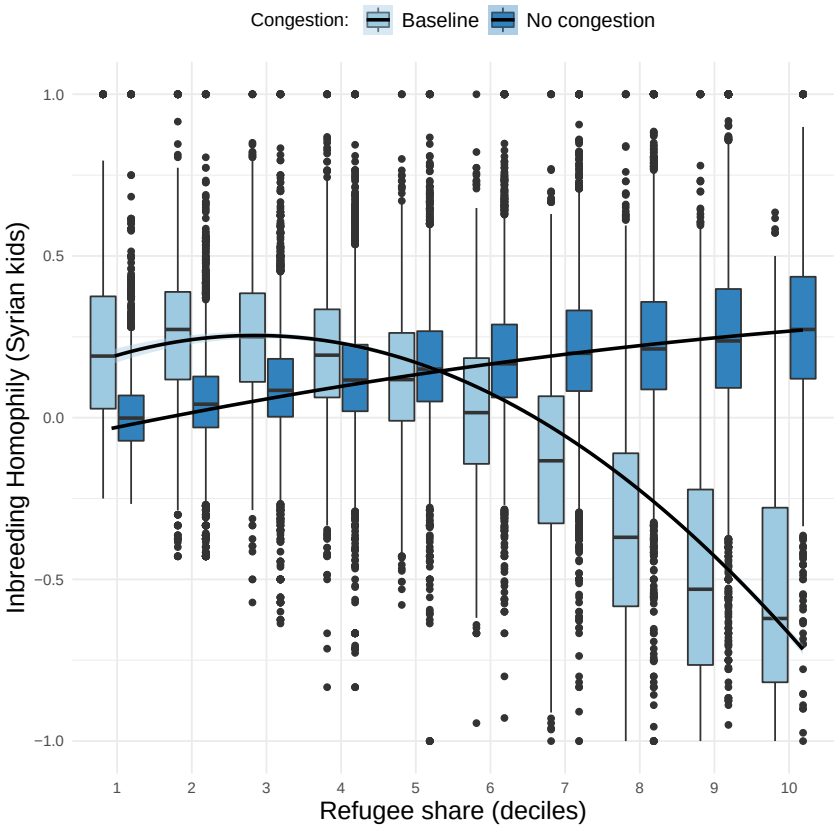


Figure: Simulated inbreeding homophily by refugee share (Syrian kids): Changes in congestion.

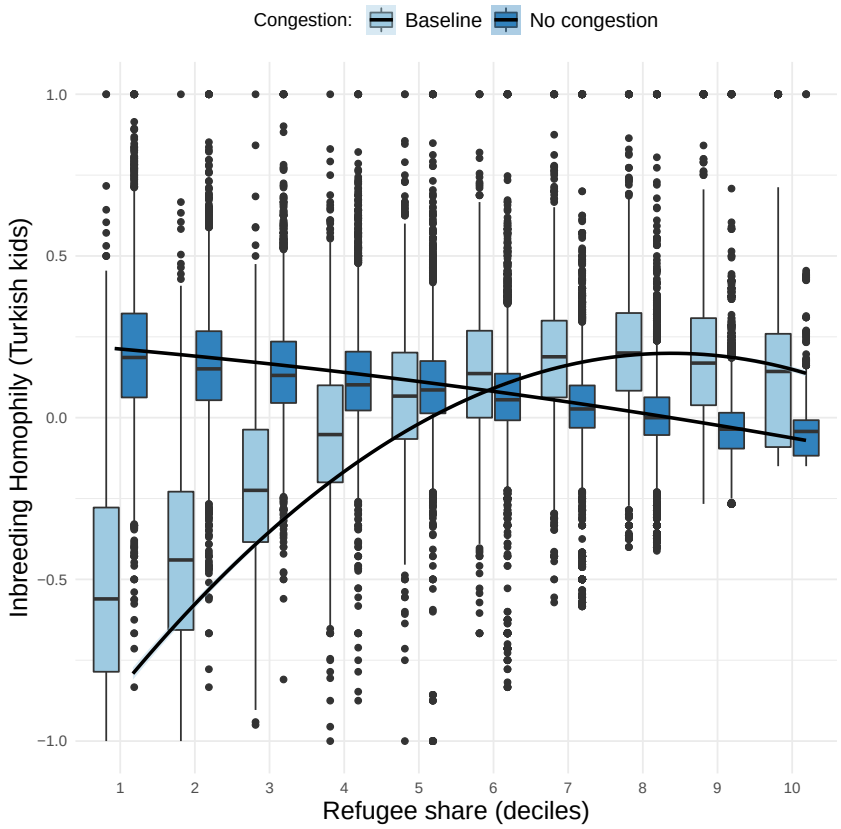


Figure: Simulated inbreeding homophily by refugee share (Turkish kids): Changes in congestion.

Counterfactual policies: Congestion

- We remove the congestion effect.
- Without congestion, inbreeding homophily is monotonically increasing with the share of own-ethnicity children in the classroom.
- Consider for example Syrian kids. Because of preference biases, links with other Syrian kids are more likely so when the share of Syrian kids increases, the (expected) number of friends Syrian kids have also increases.
- Because these additional friendships are more likely to be with other Syrian kids, this leads homophily to increase more than in the baseline, leading to a monotonically increasing inbreeding homophily index.

Counterfactual policies: Language

- We have seen that Turkish language skills of Syrian kids play an important role in the formation of inter-ethnic friendship relations.
- We now examine the impact of an intervention that would improve the Turkish language fluency of Syrian children on homophily.

Counterfactual policies: Language

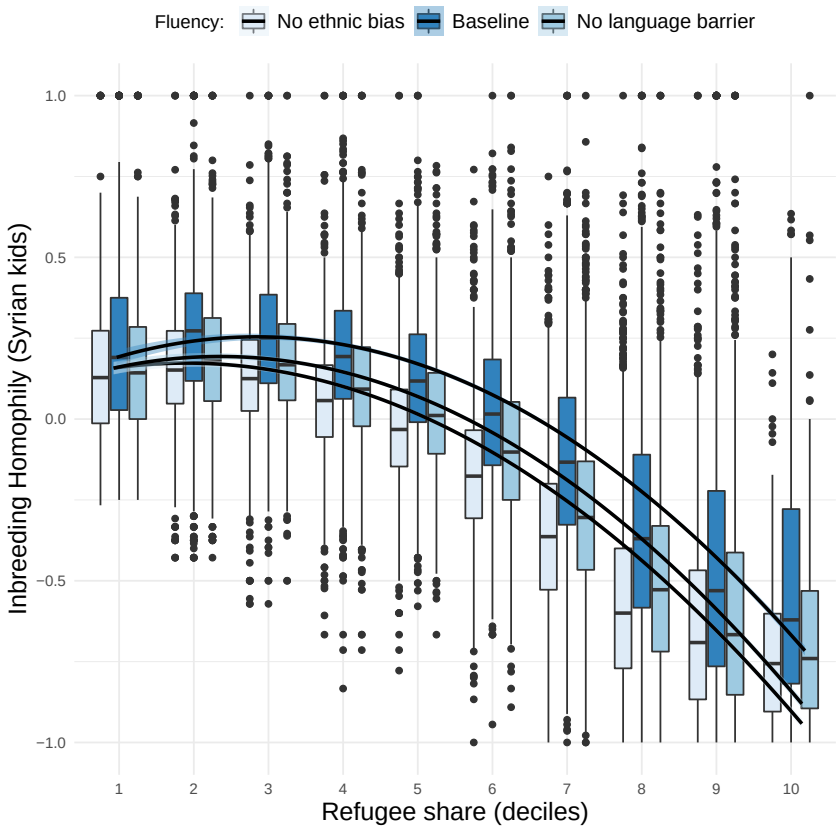


Figure: Simulated inbreeding homophily by refugee share (Syrian kids): Preference biases vs Syrian fluency.

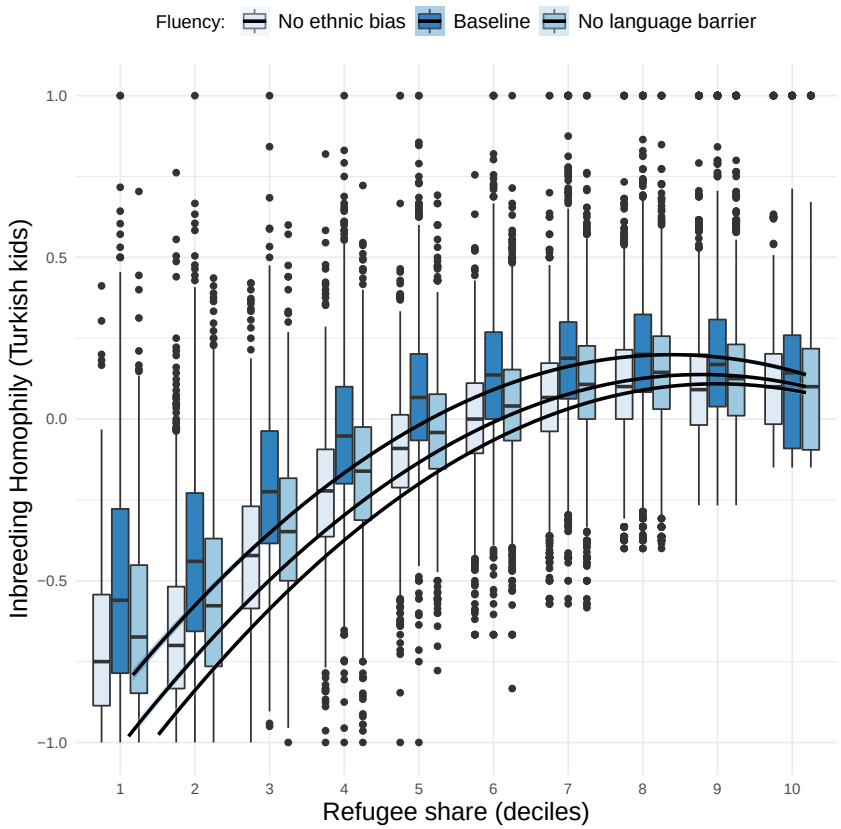


Figure: Simulated inbreeding homophily by refugee share (Turkish kids): Preference biases vs Syrian fluency.

Counterfactual policies: Language

- We compare the baseline specification with a situation without any preference bias, and a situation in which every Syrian kid is fluent.
- The resulting inbreeding homophily of making every Syrian kid fluent is roughly half way between the baseline curve and the one with no ethnic bias.
- This suggests that the policy would have the equivalent effect as reducing ethnic biases by more than 50%.
- This confirms the strong impact of language and the importance of investing in language acquisition for the integration of young children.

Counterfactual policies: Language versus ethnic bias

- Syrian (baseline): $IH^S = 0$ (i.e., $H^S = q^S$): q^S in 7th decile.
- Syrian (no ethnic bias): $IH^S = 0$ (i.e., $H^S = q^S$): q^S in 6th decile.
- Syrian (no language barrier): $IH^S = 0$ (i.e., $H^S = q^S$): q^S in 6th decile.

Conclusion

- We evaluate the impact of an early childhood Summer Camp program organized in July-August 2019 in 26 provinces,
- targeting disadvantaged 5-year-old refugee and native kids with no earlier access to any form of formal early-childhood education.
- During the Summer Camp, we implemented a randomized field experiment.
- Goal: examining the effect of random exposure to different ethnicity on various outcomes including own-group preference, friendship formation, and language skills.

Conclusion (Field experiment)

- We find that increased exposure to children of the other ethnicity leads to some increase in the formation of inter-ethnic friendships, consistent with the contact hypothesis.
- We also find a significant improvement in Turkish language skills for Syrian children across all groups, and that this improvement is related to in-class refugee concentration.
- We find that Turkish children exhibit in-group favoritism in prosocial behavior, while Syrian kids do not discriminate on the basis of ethnicity.

Conclusion (Structural estimation)

- Our structural estimation results suggest that contact alone does not guarantee improved inter-group outcomes between refugees and natives in educational settings.
- We show that the impact of contact on friendship formation is non-monotonic, so more contact can backfire.
- Programs aiming at integrating refugee children should design policies that incorporate social cohesion, prosociality, language skills and interethnic exposure issues as early as the pre-school level.