WORKING PAPERS

N° 1600

February 2026

“Matching disadvantaged children to day care: Evidence from a centralized platform”

Olivier De Groote and Minyoung Rho

Toulouse
School of
Economics

Matching disadvantaged children to day care: Evidence from a centralized platform

Olivier De Groote and Minyoung Rho*

February 25, 2026

Abstract

We use data from a platform that centralizes day care matching and estimate parents' preferences and nursery priorities from rank-ordered lists and acceptance decisions. Our novel estimation approach, inspired by dynamic discrete choice models, accounts for strategic behavior. We then evaluate centralized matching policies, comparing mechanisms and assessing the effects of subsidies, increased capacity, and affirmative action. We find that affirmative action is key to increasing participation of disadvantaged children, though it can raise segregation due to strong differences in preferences.

Keywords: day care, affirmative action, segregation, centralized matching markets, CCP estimation. JEL codes: C61, D82, I24

---

*De Groote: Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France & IZA. E-mail: olivier.de-groote@tse-fr.eu. Rho: Oxera Consulting LLP, Oxford, UK. This paper benefited from helpful comments from several audiences at ASU, CMU, CREST, Duke, KU Leuven, Northwestern, Tilburg, Tinbergen Institute, TSE, UAB, UAntwerp, UPenn, UW-Madison, QMU London, IIOC 2021, EEA-ESEM 2021, IAAE 2021, EARIE 2021, EALE 2021, the Dynamic Matching and Queuing Workshop Paris 2022, LEER 2023, the Leuven summer event 2023, the Paris workshop in honor of YingHua He, and the structural microeconomics conference 2025. We thank Kristof De Witte for his help in obtaining access to the data and "Loket Kinderopvang" (kinderopvangleuven.be) for providing it. We also thank Jarn Denijs for excellent research assistance. Olivier De Groote acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010 and of the MATCHINEQ project: ANR-22-CE26-0005-01. This work has been supported by the European Research Council (ERC), through Starting Grant n. 804989 and ERG grant EDDYNCHOICE - 101162723.

1 Introduction

Many governments promote participation in formal early childhood education and care with two main objectives: increasing female labor force participation and fostering child development. Both effects are particularly strong for socio-economically disadvantaged families (Cornelissen et al., 2018; Drange and Havnes, 2019; Felfe and Lalive, 2018; Gathmann and Sass, 2018; Havnes and Mogstad, 2011). For more advantaged families, child development effects are mixed and depend critically on the quality gap between formal care and alternative arrangements (Baker et al., 2008, 2019; Fort et al., 2020). At the same time, disadvantaged families make less use of day care, generating reverse selection on gains (Cornelissen et al., 2018). Hermes et al. (2024) show that this gap is largely driven by application barriers in decentralized matching markets.

This paper empirically investigates improving the matching process in day care - particularly for disadvantaged children. Using data from a centralized platform with a decentralized matching process, we estimate preferences of families and the probability of being accepted in a nursery. We then use these estimates to evaluate counterfactual centralized matching algorithms, modifying mechanisms, prices, and capacities to assess their effects on participation, welfare, and segregation.

Our analysis uses a unique dataset from Leuven, Belgium, covering all formal day care (ages 0-2.5). The matching process is decentralized: families rank nurseries (large-scale day care centers), and nurseries may accept or reject children. Both sides can act at any time and for any chosen start date. This context is well-suited to investigate day care matching. While there is a large literature on optimal school choice policies, there is limited literature on day care as these markets impose additional challenges. First, day care is rarely centralized, so rich data on both supply and demand are often unavailable. We observe capacities, acceptance decisions, and enrollments for all formal day care options (childminders and nurseries), alongside families' nursery rankings.[1] Second, application costs are low due to income-based pricing, which makes day care broadly affordable. As a result, our data cover 88.6% of births in Leuven.[2]

We estimate household preferences, using rank-ordered lists (ROLs) from 4,922 parents for seats between mid-2013 and the end of 2016. Parents are modeled as a mixture of naive and strategic agents. Naive agents rank alternatives according to true preferences, while strategic agents use a heuristic that accounts for expected utility, combining utility with rank-specific acceptance probabilities. This distinction is empirically relevant: nurseries are 10 percentage points more likely to accept a child ranked first, creating incentives for strategic ranking.

Strategic parents are assumed to solve a sequential decision problem: the alternative they choose in rank 1 is assumed to yield expected utility (acceptance probability q, multiplied by utility). In addition, they receive the option value of lower-ranked alternatives, multiplied by $1 - q$. This continues for the next options in the ROL. By allowing q to be rank-specific and by accounting for option values, it captures the key strategic trade-offs parents face. At the same time, it does not require them to solve an arguably too complicated optimization problem of when and how to form and update their ROL in a decentralized matching setting. Our approach provides substantial computational benefits by connecting the problem to the dynamic discrete choice and Conditional Choice Probability (CCP) literature (Rust, 1987; Hotz and Miller, 1993; Arcidiacono and Miller, 2011). Acceptance probabilities are recovered directly from the data and interpreted as parental beliefs. Parents who stop ranking without securing a place choose a terminal outside option, implying finite dependence (Arcidiacono and Miller, 2011) and allowing option values to be identified in a first step. The remaining estimation resembles a static discrete choice model with a correction term, enabling the use of an EM algorithm to flexibly model unobserved

---

[1] Nurseries account for 87.5% of the local capacity.

[2] We compare births in 2013 and 2014 to avoid right-censoring issues, as our data covers registrations through 2016.

heterogeneity and missing data (Arcidiacono et al., 2025).

Identification relies on several features of the setting. Families typically rank multiple options, facilitating the identification of preference heterogeneity with unobserved types. A small number is sufficient as we observe important determinants of preferences such as travel time and sibling attendance. Continuous application over time generates rich within-nursery variation in key variables, allowing us to include group-specific nursery fixed effects to capture unobserved quality, perceived differently by advantaged and disadvantaged families. Capacity variation helps identifying the share of strategic agents, while variation in the presence of disadvantaged children identifies peer preferences as a potential cause of segregation. Finally, income-based pricing in many, but not all, nurseries allows us to estimate price elasticities.

Estimates show that location and sibling attendance are the main drivers of utility, while preferences for peer composition are small. Only 8% of families behave strategically, with little heterogeneity by socio-economic status. Acceptance probabilities reveal systematically lower acceptance rates for disadvantaged families.

To simulate matching policies, we divide each year into six enrollment periods and sequentially apply centralized algorithms proposed by Delacrétaz et al. (2023), adjusting for seat availability and endogenous nursery characteristics. We implement variants of Knapsack Deferred Acceptance (KDA) and Knapsack Top Trading Cycles (KTTC), which extend classic DA and TTC to allow parents to specify the number of care days.

Five main results emerge. First, disadvantaged children benefit from centralized mechanisms with transparent priority rules. We prioritize siblings, employment, and those living in the same neighborhood or working for an employer that funds the nursery. Still, tie-breakers matter and using travel time is beneficial for all groups.

Second, theoretical differences between KDA and KTTC matter empirically. KTTC assigns more parents to their top choice (55% vs. 49%) and increases welfare by about 400 EUR per household applying for a spot. However, most families not matched to their top choice would envy the allocation of children with lower priorities in KTTC - an issue absent in KDA.

Third, affirmative action policies have large distributional effects and are costly to compensate for. We implement a dynamic version of soft-bounds quotas for disadvantaged children: prioritizing them in a nurseries that did not meet the quota in the previous period. A 30% quota (there are 28% of disadvantaged children in the sample) increases welfare by 1900 EUR for disadvantaged families, but reduces it by 1400 EUR for advantaged families. It also reduces the rate of unmatched disadvantaged children from 34% to 22%, but increases it for advantaged children from 31% to 36%. To gain support from advantaged families, a 9% increase in income-based capacity would make them indifferent, while further improving welfare and matches for disadvantaged families and costing society 1200 EUR per household.

Fourth, affirmative action policies increase socio-economic segregation under KTTC, but not under KDA.[3] While the latter only gives disadvantaged families an opportunity in places where they are underrepresented, KTTC allows them to gain from trade with advantaged families. While this is beneficial for welfare, the preference heterogeneity between the two groups leads to an increase in segregation. Additional simulations show that location preferences drive this result more than peer preferences.

Fifth, alternative policies are costly and not always effective. Free day care mainly benefits advantaged families, while more progressive pricing has little impact given already low fees. In contrast, letting all families pay the maximum (after tax credit) income-based price of 23 EUR per day would have detrimental effects on the participation and welfare of the disadvantaged. Expanding income-based capacity increases enrollment for all groups and, despite its cost, is welfare-improving given minimal positive externalities.

---

[3] Segregation is undesired as it leads to more prejudice and discrimination in society (Rao, 2019).

Related literature Beyond the day care literature (see above), this paper contributes to two additional strands of research.

First, we contribute to the large literature on equity and efficiency concerns in the design of matching mechanisms for allocating children to educational institutions (see section 1.7 of Che et al. (2025) for a survey). Theoretical work has characterized the properties of affirmative action policies in matching markets. Abdulkadiroğlu and Sönmez (2003) introduce quotas into Deferred Acceptance (DA) and Top Trading Cycles (TTC). Subsequent work refines these designs. Hafalir et al. (2013) show that minority reserves (soft-bounds) make minorities better off in both DA and TTC. Moreover, DA with minority reserves Pareto dominates DA with majority quotas (or hard-bounds). TTC with minority reserves results in better matching for minorities than TTC with majority quotas. Ehlers et al. (2014) establish that soft bounds guarantee the existence of fair and non-wasteful assignments, while such assignments may fail to exist under hard bounds. Closest to our work, Oosterbeek et al. (2021) empirically explores the impact of affirmative action policies in the centralized secondary school market of Amsterdam, finding modest reductions in segregation, while it reduces welfare.

Our paper provides several contributions. First, we empirically investigate matching policies and affirmative action at a much younger age (0 to 2.5 years old). Therefore, we use the newly developed theoretical framework of 'knapsack' algorithms in Delacrétaz et al. (2023) to be able to account for the fact that parents differ in the days of the week they request care. Second, we illustrate how affirmative action in a TTC-style mechanism leads to more segregation when preference heterogeneity is important. Third, the unique context of day care - where parents can apply throughout the year - generates high-frequency variation in the socio-economic status of peers within institutions. This allows us to separately identify this demand-side channel of segregation while flexibly controlling for differences in perceived unobserved quality by different socio-economic groups.[4] We distinguish between residential segregation, non-residential location preferences, preference for the socio-economic status of peers, and disagreement on the (unobserved) quality of a nursery.

Second, we contribute to the literature on estimating preferences in matching markets. Rank-ordered lists (ROLs) provide rich information and allow for highly flexible preference estimation when agents are truthful (Abdulkadiroğlu et al., 2017). Accounting for strategic behavior, however, substantially complicates this. Agarwal and Somaini (2020) review existing approaches.

A first set of papers uses Gibbs sampling, following Agarwal and Somaini (2018), which becomes computationally infeasible in settings with many ranked options, as in our context with over 40 nurseries. Simulated maximum likelihood faces similar challenges, as multiple potential strategies need to be considered and many simulations are required to avoid bias. Calsamiglia et al. (2020) reduce this for several centralized mechanisms by expressing rationally optimal strategies as backward induction problems. Our approach is related but avoids simulation entirely by including rank-specific utility shocks, in addition to persistent unobserved heterogeneity through types. By modeling strategic behavior through a simple sequential heuristic with rank-specific acceptance probabilities, we obtain closed-form choice probabilities that apply more generally, including in decentralized matching environments.[5] Consistent with recent trends documented by Che et al. (2025), our approach exploits bounded rationality restrictions to make estimation feasible, complementing work on DA with ranking costs (Idoux, 2023) and binding list-size constraints (Wang et al., 2025).

The remainder of the paper is structured as follows. Section 2 describes the insti-

---

[4] Prior research has identified several drivers of segregation, including extended travel distances to preferred educational institutions (Laverde, 2021), institutional screening policies (Lee and Son, 2022; Gazmuri, 2020), individual abilities (Oosterbeek et al., 2021), and the prevailing racial or socio-economic composition of specific groups (Burgess et al., 2014; Hastings et al., 2009; Caetano and Maheshri, 2017; Laverde, 2021).

[5] Beuermann et al. (2023) also propose a simulation-free estimator for strategic ranking, but do not derive it from a specific ranking strategy and do not need to allow acceptance probabilities to vary by rank in their setting.

tutional context. Section 3 provides a summary of the data and documents differences across socio-economic groups. Section 4 presents the specification of household preferences, and Section 5 discusses identification and estimation of the model parameters. Section 6 reports the estimation results. Section 7 presents counterfactual simulations of centralized matching mechanisms, and Section 8 concludes.

2 Institutional context

Belgium is a federal country in which day care policy is decentralized and administered by its communities. We focus on Flanders, which accounts for roughly 60% of the Belgian population. We first describe the institutional context and briefly compare it to other countries, before turning to the matching system and data from the city of Leuven.[6]

2.1 Formal day care in Flanders

Children can attend free preschool from age two and a half. From the first months after birth until preschool entry, they are eligible for formal day care. All formal day care is regulated by the government agency Kind & Gezin. Care is provided either by childminders, who host children in their homes, or by large-scale nurseries, which account for the majority of seats.

Providers are free to enter the market and choose their pricing regime. Fixed-price institutions set prices freely, while income-based institutions charge government-regulated fees that depend on household income. Both types receive subsidies, but income-based providers receive higher subsidies net of parental fees. The number of income-based institutions is limited by government budget decisions. Income-based daily prices range from 5.24 to 29.09 EUR, with lower fees possible in exceptional

---

[6] This section draws on the overviews in Gaer et al. (2013); Teppers et al. (2019); Van Lancker and Vandenbroeck (2019).

cases,[7] and average 14.16 EUR.[8] Fixed-price institutions typically charge between 25 and 35 EUR per day. In addition, all parents receive a tax credit of 45% of expenses, capped at a daily price of 11.20 EUR.[9]

The decree also specifies how income-based institutions must prioritize children when capacity is constrained (Kind&Gezin, 2019).[10] Absolute priority is given to children whose parents need day care for work or education. Additional priority applies to single-parent families, low-income households, foster children, and siblings.

On an annual basis, at least 20% of enrolled children must satisfy two or more of the following criteria: (1) parents work or are in education, (2) single parenthood, (3) low income, and (4) foster care. Further subsidies are granted if at least 30% of children belong to a designated "vulnerable" group. A child is classified as vulnerable if it meets at least two of the following conditions, with at least one drawn from the last three: (1) parents work or are in education, (2) single parenthood, (3) low income, (4) low parental education, or (5) health or care-related needs.[11]

2.2 Flemish day care in an international context

Day care provision in Flanders expanded rapidly from the 1970s onward (Van Lancker and Ghysels, 2012). Participation among children under age three rose from about 20% in the 1990s to roughly 63% by 2010, well above the European Council's Barcelona target of 33%.

---

[7] Source: https://www.vlaanderen.be/kinderopvang-met-inkomenstarief , consulted on 28/10/2020.

[8] Source: https://www.vrt.be/vrtnws/nl/2019/07/02/hoeveel-kost-kinderopvang-in-uw-gemeente/ , consulted on 28/10/2020.

[9] Both public and private providers operate in the sector. Public nurseries and affiliated childminders charge income-based prices. Private providers may choose between fixed and income-based pricing. A 2012 decree (implemented from 2014) largely harmonized public and private financing, maintaining only the distinction between income-based and fixed-price institutions.

[10] Earlier rules were less explicit, though priority was generally given to single parents, low-income or low-educated parents, and children with social or pedagogical needs. Appendix A.2 analyzes the policy change.

[11] The income threshold is indexed and was 28,757 EUR per year in 2019. Parents are classified as low educated if neither holds a high school degree.

Formal day care in Belgium is cheap: parents spend around 5% of their income, compared with an OECD average of 12% (OECD, 2011). In terms of participation, Belgium ranks 7th among OECD countries with 56% of children aged 0-2 enrolled in formal care (OECD average: 35%). Teppers et al. (2019) investigate a sample from Flanders and find an even higher rate today of around 74% regular users in the age category 3 months - 3 years.[12]

2.3 The matching platform of the city of Leuven

Our data come from an online platform in Leuven, the fourth-largest city in Flanders, with just over 100,000 inhabitants. The city comprises a historic center of about 30,000 residents and surrounding suburban areas. Leuven is a university town, hosting roughly 55,000 higher-education students, of whom about 35,000 reside in the city during the academic year (not included in the population count). The presence of major academic institutions results in a population that is, on average, more highly educated and wealthier than in other Flemish cities. Nevertheless, Leuven exhibits substantial socio-economic diversity. In particular, 20% of births occur in low-income families, compared with a Flemish average of 13%.[13]

In 2011, the city of Leuven, together with local childminders, nurseries, and welfare organizations, launched a joint platform known as Loket Kinderopvang. Its most visible component was a website (kinderopvangleuven.be) designed to match supply and demand for day care services in Leuven. The initiative proved successful: despite the absence of any legal obligation, all childminders and nurseries in the city chose to organize their services through the platform.

The platform was designed around two core principles: autonomy for day care providers and freedom of choice for parents. As a result, it is substantially more

---

[12] Source OECD data: indicator PF3.2 Enrollment in day care and pre-school. While the age differs slightly between both groups, this is likely explained by regional differences too: Van Lancker and Ghysels (2012) also found a large difference within Belgium with 63% participation at the time in Flanders and only 45% for Belgium as a whole.

[13] Sources: Stad Leuven, "Omgevingsanalyse: Leuven in Cijfers", and Census 2011 (Statbel).

flexible than centralized platforms used for school choice in many countries. Families seeking a place in a nursery rank up to five options and specify their desired days, weekly schedule, and start and end dates. Families interested in childminders register online and, nine months before their desired start date, gain access to a list of available places. If a spot is available, they contact the childminder directly. When a family ranks a nursery, the nursery is notified and can view the full application on the platform. Nurseries may either make an offer or reject the application, and they differ in how quickly they process requests. Parents can accept or decline offers. If rejected, they remain on a waiting list in case a place becomes available or they can decide to replace the option with another choice.

3 Data and the difference between socio-economic groups

We first provide a descriptive overview of the data on matches and rankings and the differences by socio-economic status. Note that our analysis will focus on demand for nurseries (not childminders), but in the first subsection we give a more complete overview of formal day care use in Leuven. The appendix contains more details about the source data (which we received from "Loket Kinderopvang") and the data cleaning process.

3.1 Matching

We report statistics for all children attending day care between June 2013 and December 2016 (the matching sample), as well as for a restricted sample used in the remainder of the paper. This restricted sample includes only families for whom we reliably observe nursery ranking data (the ranking sample) and excludes families who considered only childminders. A full description of the data cleaning process is provided in the Appendix.

Table 1 reports characteristics of all formal day care options in June 2013 (see Appendix Table A11 for December 2016). For each option, we show capacity, price system, and type (childminder or nursery). Day care options vary widely in size, and most are income-based. Weighted by capacity, 88.5% of available spots are income-based, and 87.7% are in nurseries.[14]

Table 1: Day care characteristics in June 2013

	Unweighted			**Weighted by capacity**		
	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Overall						
Capacity	108	19.315	26.887	2,086	56.396	40.553
Income-based	108	0.824	0.383	2,086	0.885	0.319
Nursery	108	0.417	0.495	2,086	0.875	0.330
Nurseries only*						
Capacity	44	41.182	31.135	1,812	64.187	37.818
Income-based	44	0.773	0.424	1,812	0.898	0.302
KUL	44	0.136	0.347	1,812	0.221	0.415
Distance to city center (km)	44	2.601	1.605	1,812	2.561	1.551
Distance to east (km)	44	3.081	1.426	1,812	3.265	1.462
Distance to north (km)	44	4.822	1.932	1,812	5.159	1.830
Distance to south (km)	44	4.038	2.037	1,812	3.728	2.015
Distance to west (km)	44	3.799	1.999	1,812	3.530	2.137
Operating hours	44	10.892	0.938	1,812	11.277	1.076

Child care characteristics in June 2013 (see Appendix Table A11 for December 2016).
*We drop one nursery in the rest of the analysis because of the small number of children attending it.

For nurseries, additional characteristics are available. A dummy variable indicates "KUL" affiliation, covering KU Leuven, its university hospital (UZ Leuven), and a large spin-off (IMEC), which offer priority to employees and account for 22% of nursery spots. Location data allow us to calculate travel times from home, though parents may value other locational factors such as workplace, grandparents, or neighborhood quality. To capture this, we calculate distances to five city points: the centroids of merged neighborhoods in the city center and four points along the cardinal directions, and treat these as nursery characteristics. Finally, we observe operating hours and

---

[14] 78% of childminder spots were income-based, compared with 91% for nurseries.

report the weekday average.

Table 2 summarizes family characteristics. First, the matching and ranking samples exhibit largely similar statistics, which is important since our demand estimation uses only the ranking sample. Second, the table shows the share of families in the groups targeted by the government's priority policy: 28% fall in the 20% priority group, while 24% are in the 30% group. We also report additional family characteristics. About 25% are low income, 25% do not speak Dutch (the local language) at home, 10% of children live in single-parent households, and 9% have parents without a high school degree. The "work or study" indicator shows that most parents are employed, studying, or seeking work. Average daily income-based fees are 14 EUR (SD = 8 EUR), though these are observed only for a subset of the sample. Data are missing either because nurseries do not use the platform for bookkeeping or because children never attend an income-based institution. As these missing observations are likely non-random, they must be accounted for in the estimation procedure.

Table 2: Characteristics of the family

	Matching sample			**Ranking sample**		
	Obs	Mean	SD	Obs	Mean	SD
20% priority group	6,038	0.281		4,922	0.266	
30% priority group	6,038	0.240		4,922	0.228	
Low income	6,038	0.250		4,922	0.237	
No Dutch	6,038	0.246		4,922	0.242	
Single parent	6,038	0.098		4,922	0.086	
Low education	6,038	0.090		4,922	0.090	
Work or study	6,038	0.938		4,922	0.940	
Income-based price	2,128	13.866	7.839	1,740	14.263	7.737

Matching sample: characteristics of families that started day care in between 2013 week 26 and 2016 week 52, added by unallocated families that wanted to start in this time period.
Ranking sample: standard data cleaning (see Appendix) and only including households that rank at least one nursery.

Since the 20% and 30% priority groups are similar in size, we focus on the 20% group and refer to them as "disadvantaged" throughout the paper; all other families are labeled "advantaged".

Table 3 reports family characteristics for the matching sample by socio-economic

Table 3: Characteristics of the family by socio-economic priority group

	Advantaged			**Disadvantaged**		
	Obs	Mean	SD	Obs	Mean	SD
20% priority group	4,342	0.000		1,696	1.000	
30% priority group	4,342	0.000		1,696	0.855	
Low income	4,342	0.024		1,696	0.830	
No Dutch	4,342	0.144		1,696	0.507	
Single parent	4,342	0.004		1,696	0.338	
Low education	4,342	0.039		1,696	0.222	
Work or study	4,342	0.966		1,696	0.867	
Income-based price	1,403	17.166	6.440	725	7.481	6.188

group. By construction, the two groups differ markedly. The primary distinction is income: only 2.4% of advantaged families are low-income, compared with 86% of disadvantaged families, which is also reflected in the income-based fees they face (17 EUR vs. 7 EUR). Other indicators of disadvantage show similar gaps. While language at home is not used to define the groups, most disadvantaged families do not speak Dutch at home, compared with 14% of advantaged families. Single-parent households and parents without a high school degree are common among the disadvantaged (about one-third and one-fifth, respectively) but rare among the advantaged. Both groups have high rates of parents working or studying, though it approaches 100% for advantaged families.

Table 4 shows differences in matching outcomes. While the total number of matches is similar, disadvantaged children are slightly less likely to be matched to a nursery and, when matched, substantially less likely to receive their top-ranked choice. They also attend different nurseries. Appendix Figure A3 shows the segregation curve, which implies a Gini-coefficient of 0.37 and dissimilarity index of 0.26.[15]

---

[15] The Gini coefficient measures the area between the segregation curve and the 45° line relative to the total area under the 45° line. The dissimilarity index equals one half of the sum of absolute differences in group shares across day cares.

Table 4: Match statistics

	Overall	**Advantaged**	**Disadvantaged**
Match	0.743	0.741	0.749
with nursery	0.650	0.667	0.606
with childminder	0.094	0.074	0.144
Rank*			
1	0.546	0.586	0.424
2	0.110	0.106	0.122
3	0.071	0.065	0.089
4	0.054	0.051	0.064
5	0.044	0.041	0.051

Characteristics of families applying for day care in a slot starting 2013 week 26 - end of 2016. Total obs = 6,038, of which 4,342 advantaged and 1,696 disadvantaged.
*To calculate the share in each rank, we use only children matched to a nursery that are part of the ranking sample: total of 3,355 children allocated to nurseries, of which 2,527 advantaged and 828 disadvantaged. Rank >5 not included in table.

3.2 Ranking

When parents enter the platform, they can rank up to five alternatives and they can do so from the moment they know they are expecting.[16] Figure 1 illustrates the timing of spot requests for advantaged and disadvantaged families. Advantaged families typically request a spot earlier, while disadvantaged families show a pronounced peak just before their desired start date. Similarly, advantaged families generate high demand within the first two months after conception, whereas disadvantaged families' requests are more evenly distributed over time, with many submitting applications after the child's birth.

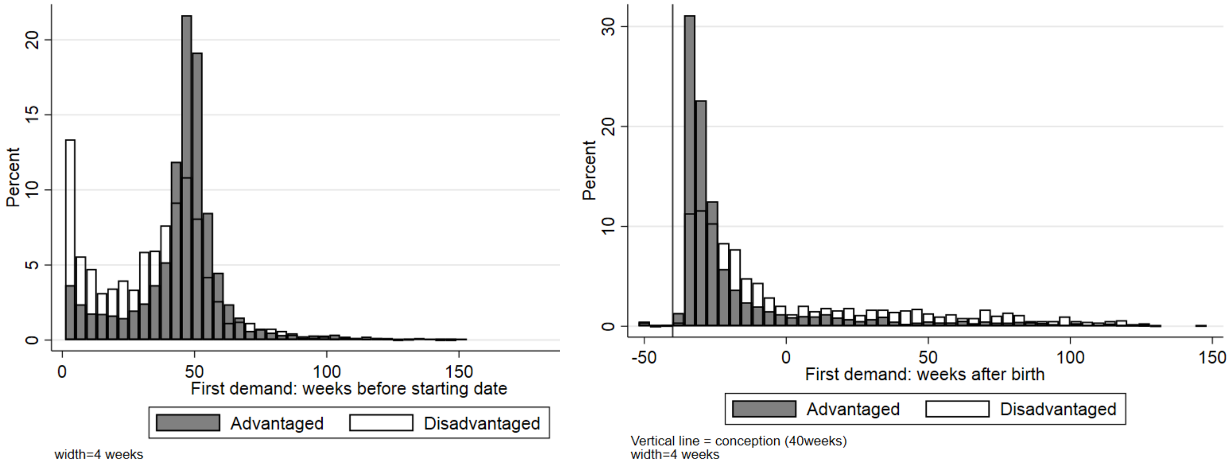
Figure 2 shows the number of ranked alternatives by each group. Most families rank the maximum of five options, but a sizable share - particularly among advantaged families - rank only one. Several families rank more than five alternatives by replacing unavailable options with new ones.

Table 5 compares the characteristics of families' top-ranked alternatives with their

---

[16] Table A12 shows that the ranking system largely corresponds with actual attendance behavior.

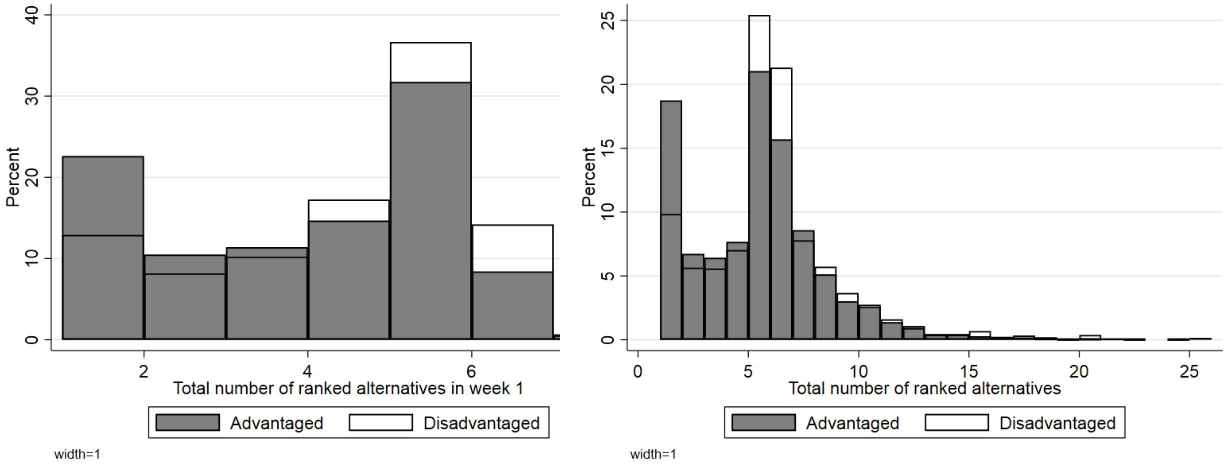
Figure 1: Timing of demand



width=4 weeks

Vertical line = conception (40weeks)
width=4 weeks

Figure 2: Number of nurseries ranked



width=1

width=1

fifth-ranked options and the average available when they apply. We include the previously discussed nursery characteristics and also factors that vary by applicant and timing. Differences reflect both preferences and, for strategic families, acceptance probabilities. Both groups strongly favor income-based nurseries, though advantaged families are relatively more likely to also consider fixed-price options, consistent with the smaller price differences they face. Both prefer shorter travel times, but disadvantaged families are particularly sensitive to bus travel and central locations, suggesting greater reliance on public transportation.[17] Finally, disadvantaged families tend to favor nurseries with a high share of disadvantaged children, whereas this factor does not influence advantaged families' rankings.

Table 5: Characteristics of ranked vs all alternatives

	Advantaged			**Disadvantaged**		
	First rank	Fifth rank	Average	First rank	Fifth rank	Average
Income-based	0.961	0.882	0.787	0.969	0.924	0.792
KUL	0.250	0.166	0.134	0.181	0.146	0.132
Distance to city center (km)	2.509	2.213	2.589	1.856	1.820	2.597
Distance to east (km)	3.278	3.011	3.077	3.142	3.082	3.069
Distance to north (km)	5.057	4.983	4.808	5.221	5.112	4.806
Distance to south (km)	3.720	3.551	4.062	3.129	3.261	4.078
Distance to west (km)	3.398	3.317	3.801	2.899	2.852	3.814
Operating hours	11.502	11.151	10.911	11.382	11.182	10.915
Share disadvantaged	0.203	0.196	0.187	0.283	0.234	0.188
Travel time (min per trip)	11.981	12.469	15.366	11.626	12.390	14.583
Time bus / time car	2.194	2.369	2.509	2.023	2.165	2.386
Enrolled children	88.381	67.569	50.104	92.265	68.543	51.205

Share disadvantaged and number of enrolled children calculated at the time of applying for a spot.

4 Household preferences

Our goal is to explore how to improve the matching of children, particularly those from low socio-economic backgrounds. To this end, we implement various centralized matching algorithms and evaluate their welfare implications. Doing so requires knowledge of household preferences. This section introduces the utility function we use, while the next section discusses how to estimate its parameters from rank-ordered lists, accounting for potential strategic behavior.

---

[17] Travel times are calculated using Google Maps based on family and nursery locations, with five minutes added to account for preparation and to avoid zeros in relative differences.

Group-specific utility functions For each family i of socio-economic group s, the indirect utility derived from attending a nursery j is denoted:

$$\begin{aligned} u_j^s(W_{ij}) =& \delta_{js} + W_{ij}\beta_s - p_{ij} && \text{if } j \neq 0 \\ u_j^s(W_{ij}) =& 0 && \text{if } j = 0 \end{aligned} \tag{1}$$

δ_{js} is a group-specific nursery fixed effect, capturing heterogeneity in the valuation of observed and unobserved characteristics of each nursery by each socio-economic group. W_{ij} is a vector of individual and nursery characteristics, arbitrarily correlated with the fixed effect, and p_{ij} is the price that family i would pay for nursery j. Opting out of the centralized matching system is denoted as $j = 0$, with its utility normalized to 0 for all i.

Heterogeneity within socio-economic groups Heterogeneity within socio-economic groups is captured by the observed and unobserved characteristics in W_{ij}. To rationalize all ranking data using a computationally efficient estimator, we will add rank-specific random shocks. Because these shocks are uncorrelated across ranks, u_{ij} must be sufficiently rich to capture (persistent) heterogeneity in preferences, ensuring reliable welfare estimates and substitution patterns. The same argument is used in the dynamic matching context by (Agarwal et al., 2021) as their shocks are also rapidly replaced by new ones. We extend this by incorporating unobserved heterogeneity directly into u_{ij}, analogous to the motivation of BLP-models that use random coefficients to avoid the restrictive substitution patterns implied by a restrictive (but computationally convenient) shock distribution (Berry et al., 1995). Specifically, we allow each agent to belong to one of two unobserved preference types (A or B) and, along with observed family characteristics, interact them flexibly with nursery characteristics.

We account for various nursery characteristics. First, we capture a general preference for attending a nursery in Leuven relative to each family's outside option. We do this by allowing an additive shift in the utility of all inside options with a term that depends on observed family characteristics, unobserved preference type, and the time remaining until the desired start date. We also control for the income-based price families would pay, as income-based institutions in other cities are part of the outside option.

Second, we estimate the importance of travel time in a flexible way. We consider the commuting time by car with the home location and allow its effect to differ by observed family characteristics and the unobserved preference type. We also include a preference for relative differences in travel time by public transportation.

Third, we capture preferences over nursery characteristics that vary across households. Because parents can apply and enroll at any time, several variables exhibit within-nursery variation.[18] Their effects are estimated via β_s, while group-specific perceived quality is captured by δ_{js}. In particular, parents consider the number of children and the share of disadvantaged children when applying. Controlling for group-specific fixed effects allows us to interpret β_s as capturing preferences for social composition, net of correlated effects arising from group-specific perceptions of unobserved quality.[19] Nursery prices vary by household due to the income-based pricing structure of most (but not all) nurseries, enabling us to control for unobserved quality through fixed effects while identifying price effects. Utility is normalized with respect to price so this identifies the scale of the utility function in euros. We also include additional factors influencing choice: whether a sibling is enrolled and whether the nursery belongs to the KUL system.

Finally, we capture preference heterogeneity over invariant nursery characteristics, both observed and unobserved. To model heterogeneity in unobserved characteristics,

---

[18] Although these characteristics are time-varying, a time-subscript is not needed in most cases as each i comes with a particular application time. However, if parents add alternatives later, we do have within-individual variation that we ignore in notation here, but we account for it in estimation by updating these variables.

[19] As explained in Manski (1993), similarities in behavior within groups may arise not only from endogenous effects (peer influence) or contextual effects (peers' characteristics), but also from correlated effects - shared, potentially unobserved, individual traits or environments.

we first cluster the 45 nurseries into four groups using k-means.[20] We interact a constant, dummy variables for three of the four clusters, and observable characteristics (distances to five locations and operating hours) with observed family characteristics and unobserved preference types. In a second-stage regression, we relate the nursery fixed effects to the same characteristics to better understand what families value.

5 Estimation

This section outlines our empirical strategy to uncover household preferences. Section 5.1 presents the estimation strategy assuming (naive) agents report preferences truthfully. Section 5.2 extends this to strategic agents, and Section 5.3 derives a likelihood function for a mixture of both types. Finally, Section 5.4 discusses identification.

5.1 Truth-telling agents

If family i is naive, i.e., reports preferences truthfully, we assume they rank options according to the choice-specific utility u_{ij} from (1) plus an individual shock ϵ_{ijr} for each choice j at each rank r. We also include a ranking cost c_s. Unlike a standard exploded logit, shocks vary not only by family i and alternative j, but also by rank r. While this does not affect estimation for naive agents, it provides substantial computational advantages when modeling strategic behavior. It also captures rank-specific effects, such as trembling-hand errors or updated information when parents add alternatives later. Persistent unobserved heterogeneity is captured by u_{ij}.

To rationalize choices at each rank, we define the agent's perceived rank-specific value function:

---

[20] We use the number of children, the share of 20% and 30% priority groups, the share of single-parent children, the share of low-income children, the share of children of parents with low education, the share of children of parents who do not speak Dutch and the share of parents working or studying, after subtracting their mean and dividing by their standard deviation.

$$V_r^{\text{s,Naive}}(X_{ir}^{\tau}, \epsilon_{ir}) = \max_{j \in \mathcal{J}_i \setminus H_{ir}} \left\{ v_{jr}^{\text{s,Naive}}(X_{ir}^{\tau}) + \epsilon_{ijr} \right\}.$$

where $X_{ir}^{\tau} \equiv \{W_i, H_{ir}\}$ is the relevant state for a family i at slot r of unobserved type τ. It consists of two elements: (1) observed and unobserved characteristics that influence utility $W_i \equiv \{W_{ij}\}_{\forall j}$ and (2) an unordered set of previous optimal choices $H_{ir} \equiv \{d_{i\rho}\}_{\rho=1}^{r-1}$. The previous optimal choices are removed from the choice set for the next slot. ϵ_{ir} is a vector of all ϵ_{ijr}. $\mathcal{J}_i$ is the set of options in the market when i applies. $v_{jr}^{\text{s}\tau\text{,Naive}}(X_{ir}^{\tau})$ is defined as the conditional value function, given by

$$v_{jr}^{\text{s,Naive}}(X_{ir}^{\tau}) = u_j^s(W_{ij}) - c_s.$$

If we assume ϵ_{ijr} to be mean-zero Extreme Value distribution type 1 (EV1) distributed with scale σ_s, we obtain logit probabilities:

$$p_{jr}^{\text{s,Naive}}(X_{ir}^{\tau}) = \frac{\exp(v_{jr}^{\text{s,Naive}}(X_{ir}^{\tau})/\sigma_s)}{\sum_{j' \in \mathcal{J}_i \setminus H_{ir}} \exp(v_{j'r}^{\text{s,Naive}}(X_{ir}^{\tau})/\sigma_s)} \tag{2}$$

If all agents were truth-telling, a likelihood function can be composed from these probabilities and gives rise to a standard exploded logit model.[21]

5.2 Strategic agents

Families may not always rank truthfully. In the decentralized matching, nurseries observe child characteristics and rank-order lists, and may prioritize families that rank them highly. Abdulkadiroğlu et al. (2017) document this for New York high schools (pre-2003), and we observe similar patterns in our setting (see results section). This creates a strategic incentive not to "waste" a top ranking on an infeasible option.

---

[21] If all agents are naive, the ranking cost c_s is not identified and should be dropped, since it enters the conditional value function of every choice alternative.

We model strategic agent i as forming rank(r)-specific beliefs about their admission probabilities in each nursery j: q_{ijr}. We leverage the computational advances in dynamic discrete choice estimation (Rust, 1987; Hotz and Miller, 1993; Arcidiacono and Miller, 2011), by treating the formation of the rank-ordered list as a dynamic problem in which agents first decide what to rank first and then go down their list. State transitions are given by beliefs on acceptance probabilities. $q_{i0r} = 1$ and in Appendix A.4, we explain how we take advantage of the acceptance decisions of the nurseries to obtain an estimate of the other q_{ijr}.[22] At each ranking slot, they choose the option with the highest perceived value. Different from a naive agent, they consider the *expected* utility of the alternative as well as the option value arising from lower-ranked alternatives.

This results in the following value functions:

$$V_r^{\text{s,Str}}\left(X_{ir}^{\tau}, q_i, \epsilon_{ir}\right) = \max_{j \in \mathcal{J}_i \setminus H_{ir}} \left\{ v_j^{s,Str}(X_{ir}^{\tau}, q_i) + \epsilon_{ijr} \right\},$$

with the conditional value function and the resulting choice probabilities now defined as

$$v_{jr}^{\text{s,Str}}\left(X_{ir}^{\tau}, q_i\right) = q_{ijr} u_j^s(W_{ij}) - c_s + (1 - q_{ijr})\, E_{\epsilon_{r+1}}\left[V_{r+1}^{\text{s,Str}}\left(X_{ir+1}, q_i, \epsilon_{ir+1}\right)\right], \tag{3}$$

$$p_{jr}^{\text{s,Str}}(X_{ir}^{\tau}, q_i) = \frac{\exp(v_{jr}^{\text{s,Str}}(X_{ir}^{\tau}, q_i))/\sigma_s)}{\sum_{j' \in \mathcal{J}_i \setminus H_{ir}} \exp(v_{j'r}^{\text{s,Str}}(X_{ir}^{\tau}, q_i)/\sigma_s)}. \tag{4}$$

5.3 Likelihood function

Valuations are parameterized with group-specific parameters (δ_s and β_s), allowing separate estimation for advantaged and disadvantaged families. Since the family's preference type τ is unknown, we estimate its distribution. The probability that a family

---

[22] They are treated as data when estimating utility. A bootstrap procedure that also re-estimates q_{ijr} in each bootstrap sample provides standard errors.

of observed type s belongs to unobserved type τ is denoted π_s^τ, with $\sum_{\tau=\{A,B\}} \pi_s^\tau = 1$. Similarly, we estimate the proportion of strategic and naive types. The proportion of the population who is strategic is denoted as $\lambda_s(S_i, \tau)$, with its counterpart being naive. S_i captures observed family characteristics (low income, single parent, low education, no Dutch, work or study) and the time difference between the first application and the requested starting date. We model this probability as a binary logit. All parameters to estimate at this stage are summarized in $\Theta = \{\delta, \beta, \pi, \lambda\}$.

The likelihood of observing the ROL d_i, given the acceptance probabilities q_i and the observed state X_i is given by equation

$$\mathcal{L}\left(d_i, X_i, q_i|\Theta\right) = \Pi_{i\in\mathcal{I}} \left[\sum_{\tau=\{A,B\}} \pi_s^\tau \left(\lambda_s(S_i,\tau) L^{\text{s,Str}}\left(d_i, X_i^\tau, q_i|\Theta\right) + \left(1 - \lambda_s(S_i,\tau)\right) L^{\text{s,Naive}}\left(d_i, X_i^\tau|\Theta\right)\right)\right]$$

with X_i^τ accounting for both observed and unobserved states,

$$L^{\text{s,Str}}\left(d_i, X_i^\tau, q_i|\Theta\right) = \Pi_{r=1}^{R_i} \Pi_{j\in\mathcal{J}_i\setminus H_{ir}} p_{jr}^{\text{s,Str}}(X_{ir}^\tau, q_i)^{I(d_{ir}=j)},$$
$$L^{\text{s,Naive}}\left(d_i, X_i^\tau|\Theta\right) = \Pi_{r=1}^{R_i} \Pi_{j\in\mathcal{J}_i\setminus H_{ir}} p_{jr}^{\text{s,Naive}}(X_{ir}^\tau)^{I(d_{ir}=j)}$$

and $p_{jr}^{\text{s,Naive}}(X_{ir}^\tau, q_i)$ and $p_{jr}^{\text{s,Str}}(X_{ir}^\tau)$ are the CCPs given by (2) and (4). Finally, $I\left(d_{ir} = j\right)$ indicates that the observed choice for i at slot r is the nursery j.

Interpreting the ROL data Parents can add alternatives later and change their ranks, and we do not always observe whether a family prefers the outside option over non-ranked alternatives. We assume that later-added alternatives are ranked below previously ranked ones, and that i ranks $j = 0$ above non-ranked alternatives only if they leave the platform without an allocation. For other cases, we make no assumptions about the ordering of non-ranked options relative to the outside option.

Missing price data In several cases, the income-based prices parents (would) pay are unobserved. While some are missing at random, we also do not observe prices when a child never attends an income-based nursery. Since this is likely non-random, we integrate over missing prices in the likelihood, similar to Arcidiacono et al. (2025) for missing college majors. In practice, this creates 12 types for each socio-economic group s: two behavioral types (naive or strategic) × two preference types (A or B) × three price types (5, 16, or 23 EUR/day). To separately identify the price types, we first estimate their population probabilities using a Heckman selection model (Heckman, 1979).[23]

Option values in a first step To avoid solving the model for strategic agents by backward induction, we obtain the option value of lower-ranked alternatives in a first stage by using CCP estimation. With the EV1 assumption, Hotz and Miller (1993) and Arcidiacono and Miller (2011) show that we can write

$$E_{\epsilon_{r+1}}\left[V_{r+1}^{\mathrm{s,Str}}\left(X_{ir+1}^{\tau}, q_i, \epsilon_{ir+1}\right)\right] = v_{j^*,r+1}^{\mathrm{s,Str}}(X_{ir}^{\tau}, q_i) - \sigma_s \ln\left[p_{j^*,r+1}^{\mathrm{s,Str}}\left(X_{i,r+1}^{\tau}, q_i\right)\right] \tag{5}$$

with j^* any option in the choice set.[24] This equation has an intuitive interpretation: the ex-ante value function is a sum of the conditional value function of any option j^*, plus a nonnegative term that adjusts for j^* not being the optimal choice. Which alternative we choose to be j^* is arbitrary, but a convenient choice is the outside option: $j^* = 0$. This gives $v_{j^*,r+1}^{\mathrm{s,Str}}(X_{ir}^{\tau}, q_i) = -c_s$, hence removing any further dependence on the future. We then obtain the following conditional value functions:

$$\begin{aligned} &v_{jr}^{s,\mathrm{Str}}\left(X_{ir}^{\tau}, q_i\right) \equiv q_{ijr} u^s(W_{ij}) - c_s + (1 - q_{ijr})\left(-c_s - \sigma_s \ln\left[p_{0,r+1}^{\mathrm{s,Str}}\left(X_{i,r+1}^{\tau}, q_i\right)\right]\right), \\ &v_{0r}^{\mathrm{s,Str}}\left(X_{ir}^{\tau}, q_i\right) \equiv -c_s. \end{aligned}$$

---

[23] An ordered probit models the three price categories, with a selection equation for missing data that includes observed family characteristics and the probability of ranking a price-missing alternative first. The latter is obtained from a conditional logit using travel time, controlling for nursery fixed effects and other non-price attributes, exploiting heterogeneity in travel times to nurseries that report prices for identification.

[24] See also Murphy (2018) for the proposed money-metric utility specification, i.e. with a price coefficient normalized to -1, but an estimated scale parameter.

Aside from rescaling flow utility by q_{ijr}, the estimator is a standard dynamic discrete choice model with unobserved types and finite dependence, which can be estimated using the EM algorithm adaptation of Arcidiacono and Miller (2011). To simplify estimation, we first fit a specification with the scale normalized to one but an estimated price coefficient. Dividing all parameters by the price coefficient then expresses utility in euros. A reduced-form estimate of the CCP for the outside option in (5) is required and is obtained in a first stage via a flexible logit on family and nursery characteristics.

5.4 Identification

Dynamic discrete choice models are identified after normalizing the utility of one option, fixing the discount factor, and specifying the distribution of the error terms (Magnac and Thesmar, 2002). Apart from common ranking costs, we normalize the utility of the outside option to zero. The discount factor is set to one for strategic agents and to zero for naive agents. Identification of the mixture exploits variables that affect acceptance probabilities - relevant only for strategic agents - but do not enter the utility function. As shown in Appendix A.4, we impose several empirically relevant exclusion restrictions, notably available capacity at the time of application and the rank position of the nursery in the family's list. We also include an applicant score that was used by nurseries on the platform. It is based on a formula set by them, using family characteristics. While this score does not constrain them, they can use it to sort applications. Note that our specification differs slightly from a standard dynamic discrete choice model in that flow utility is multiplied by the acceptance probability (see (3)). This does not create identification issues, however, since state transitions are identified outside the model using observed acceptance decisions. We also allow for unobserved preference heterogeneity by exploiting the full ROL. Since the error term ϵ_{ijr} is assumed i.i.d. across ranks, persistent unobserved differences across individuals must be captured by the preference types. For example, an individual who consistently ranks nearby institutions highly is more likely to have a strong preference for proximity.

In addition to unobserved preference types, we must integrate over unobserved prices for a substantial share of the sample. Some income-based institutions do not report prices, and for families allocated to fixed-price institutions we do not observe the income-based price they would have paid.[25] This setting is the mirror image of the standard wage-selection problem, where low wages are often unobserved (Heckman, 1979). Identification follows similar logic: we exploit preference shifters - most importantly travel time to nurseries that do or do not report income-based prices - to identify the distribution of unobserved prices.

We also identify the cost of ranking an additional alternative, using variation in the length of the ROL. Since only utility differences are identified, this cost drops out for naive families. For strategic families, however, it enters the conditional value functions through its effect on expected continuation values. These families effectively solve an optimal stopping problem, deciding when to stop adding alternatives if not accepted. Ranking an option with a low acceptance probability increases the expected need to add lower-ranked alternatives, thereby lowering expected utility. Intuitively, higher ranking costs induce strategic agents to favor nurseries with higher acceptance probabilities to avoid extending their list. Estimating this cost is also important to offset any artificial incentive to be rejected, since rejection allows agents to draw a new set of shocks.

6 Estimation results

We first present the (only) estimates required for our main simulations: household preferences. We then briefly summarize acceptance probabilities, which are needed to account for strategic parents in the data.

---

[25] Prices in fixed-price institutions are absorbed by fixed effects and therefore do not affect identification.

6.1 Household preferences

Utility parameters (β_s) Table 6 and Appendix Tables A13-A14 present the demand estimation results. We first estimate the model with a normalized scale and a price coefficient, then divide all parameters by the price coefficient to obtain the

Table 6: Demand side: estimation results (1 of 3)

	Advantaged		**Disadvantaged**	
	coef	se	coef	se
β_s in equation (1)				
Travel time (min per trip)	-6.788	(0.447)	-1.510	(0.213)
× Type B	4.185	(0.213)	-2.697	(0.205)
× No Dutch	0.300	(0.125)	0.103	(0.091)
× Low income	0.959	(0.319)	0.674	(0.126)
× Single parent			0.319	(0.077)
× Low education	0.251	(0.221)	-0.067	(0.090)
× Work or study	-0.122	(0.275)	-0.152	(0.123)
Time bus / time car	-6.370	(0.382)	-5.437	(0.485)
Enrolled children	0.033	(0.009)	0.012	(0.008)
Share disadvantaged	16.194	(2.110)	18.864	(1.812)
Sibling	34.341	(2.385)	16.602	(2.066)
KUL priority	32.152	(1.810)	19.491	(1.773)
Constant × Income-based price	2.131	(0.074)	1.459	(0.092)
Constant × Weeks until start	0.269	(0.032)	0.225	(0.024)
Scale (σ_s)	12.731	(0.627)	7.580	(0.511)
Cost to rank ($-c_s$)	-44.872	(2.297)	-26.406	(1.852)
Share type B	0.415		0.487	
Share strategic	0.069		0.103	
Nursery obs × demographics	YES		YES	
Nursery clusters × demographics	YES		YES	
Children	3,613		1,309	
Children × ranks	17,849		7,252	

Utility scaled in daily prices. Heterogeneous effect of observed and unobserved nursery characteristics can be found in Table A13 and Table A14. Group-specific nursery fixed effects included. The constant and baseline effects of operating hours and distances are obtained from a second stage regression of the nursery fixed effects. Bootstrap standard errors in parentheses, based on 50 replications.

money-metric utility specification in (1). This allows interpreting estimates as willingness to pay for each day of day care use. Standard errors are computed via a bootstrap procedure.[26]

Travel time strongly affects families' preferences. For the least sensitive unobserved types, a one-way trip reduces utility by 2.6 EUR/day for advantaged families and 1.5 EUR/day for disadvantaged families. Since parents typically make this trip twice daily, this implies an opportunity cost of time of roughly 1.3 and 0.8 EUR/minute, respectively. Other unobserved types face substantially higher costs, while observable characteristics have relatively small effects.[27] This type likely captures families who value proximity for reasons beyond time, such as neighborhood ties. Moreover, the average travel times in the data are low (a one-way trip is around 15 minutes for the average nursery available), and do not take into traffic jams or alternative modes (walking, biking). While this may affect interpreting the parameter as an opportunity cost of time, it does not affect our analysis of optimal nursery matches. We also account for differences in public transportation travel times: for parent-nursery pairs with poor connections, doubling the travel time reduces utility by 5-6 EUR/day.

We also account for two endogenously evolving nursery characteristics, though their effects are small. Adding 10 children to a nursery increases utility by only 0.1-0.3 EUR/day. Interestingly, a higher share of disadvantaged children is positively valued by both groups: a 10 percentage point increase raises utility by 1.6 EUR/day for advantaged families and 1.9 EUR/day for disadvantaged families.

Having a sibling in the facility strongly increases utility by 34 EUR/day for advantaged families and 19 EUR/day for disadvantaged ones. These values are consistent with the travel cost estimates, as a sibling reduces the number of trips by roughly

---

[26] We use 50 bootstrap samples and re-estimate the full model, including supply, demand, and auxiliary regressions. The EM algorithm is not re-run; weights are fixed, assuming no uncertainty in type estimation.

[27] We omit interactions with the single-parent dummy for advantaged families, as there are only twelve in this subsample.

half. A similar effect is observed for KUL priority. Note that while both variables also strongly predict acceptance (discussed later), our estimates here account for strategic behavior, so the effect sizes are not driven by acceptance probabilities.

We capture heterogeneity in outside options by interacting the constant with families' income-based price and the time remaining until the requested start date. Parents who face higher income-based prices (including outside Leuven) value outside options less: each additional euro reduces the attractiveness of outside alternatives by about 2 EUR/day for advantaged families and 1.5 EUR/day for disadvantaged families. Since this effect is larger than 1, it likely also reflects a higher intrinsic preference for day care among high-income families. Families who apply earlier are also less likely to choose the outside option, indicating that those with a high need ensure timely applications. People who are 10 weeks earlier, value day care more at a rate of around 2 to 3 EUR per day.

Appendix tables show the large number of estimates capturing heterogeneous preferences over fixed nursery characteristics. Appendix Table A13 shows that all families value more operating hours, although less so for parents who do not speak Dutch. For location preferences, unobserved types are very important, but also some observables matter. Non-Dutch-speaking parents prefer nurseries in the city center, while low-income households stay away from the north and south. Appendix Table A14 highlights large differences between how unobserved types value different clusters, while effects of observables are less important. We do find heterogeneity over clusters with Non-Dutch-speaking parents preferring clusters 2 and 3, and advantaged but low-income parents preferring cluster 3.

Other parameters Table 6 also reports the estimates of parameters that do not enter u_{ij}, including the scale of the shocks. Note that a standard deviation of the shocks is given by $\sigma\frac{\pi}{\sqrt{6}}$, which corresponds to 16 EUR for advantaged families, and 10 EUR for disadvantaged families. These parameters test the richness of our model.

Their relatively small magnitude - despite being used to rationalize the entire ROL after accounting for observables and unobserved types - indicates that the main choice margins are well captured by u_{ij}. Finally, we estimate a high cost of adding alternatives to the ROL, suggesting that families are reluctant to exhaust all options before turning to their outside alternative

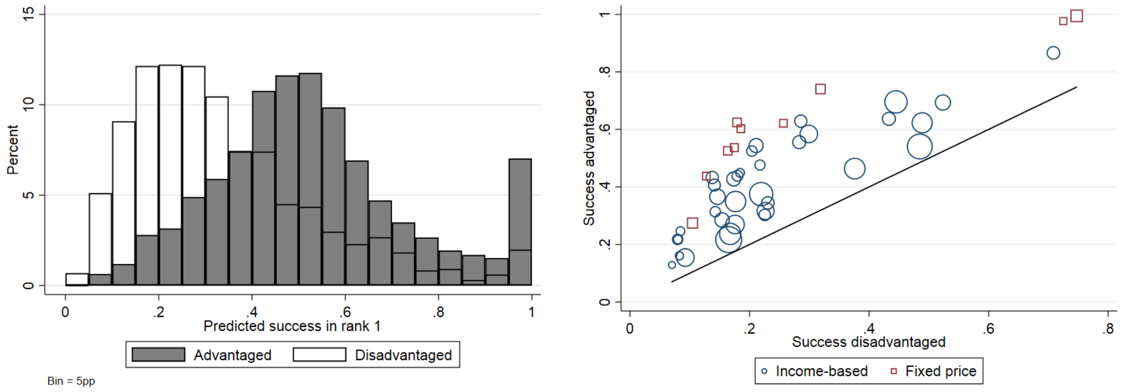
Finally, we report the share of unobserved preference types and strategic families. The unobserved types are almost evenly divided, while very few families behave strategically in either group - 7% among advantaged and 10% among disadvantaged families. Appendix Table A15 reports the average marginal effects for being a strategic type. For advantaged children, only the number of weeks remaining until the desired start date is statistically significant: more weeks decrease the likelihood of strategic behavior. This likely reflects the smaller need to act strategically when chances of acceptance are higher and parents have more time to add alternatives to their list. For disadvantaged children, the same effect holds, with additional influences: non-Dutch-speaking parents have a lower probability of being strategic (as they potentially have less access to information), while being a single parent or belonging to preference type B increases it.

6.2 Acceptance probabilities

Details on the estimation of acceptance probabilities and a discussion of the estimates are provided in Appendix A.4 and Tables A18-A19. We summarize the main insights in three steps.

First, we plot acceptance probabilities of a first-ranked alternative (q_{ij1}) for each group in the left panel of Figure 3. The two distributions differ substantially. Disadvantaged children exhibit a large mass at acceptance probabilities around 20% and very little mass above 60%. In contrast, low acceptance probabilities are uncommon among advantaged children, whose distribution has a mode close to 50%, with a substantial share of families facing success rates close to 100%.

Figure 3: Acceptance probabilities and socio-economic groups



Notes: plots of the acceptance probabilities in rank 1: q_{ij1} (see equation (6)). The left panel plots them for the agents in the sample for the alternative they apply for. The right panel plots the probability for a representative set of household and time-varying nursery characteristics for each socio-economic group in each nursery. We take characteristics that are close to the average for households that apply between 30 and 50 weeks ahead: 8 % of spots available, 22% disadvantaged children, located 12 minutes away by car, 18 weeks old at the start, and 130 weeks demanded for the five weekdays. We also omit the effects of individual demographic characteristics. We plot this for a working household that applies (the average) 40 weeks ahead, with no sibling in a nursery. Circles and squares are proportional to capacity.

Second, the right panel of Figure 3 plots the heterogeneous impact of nurseries on acceptance probabilities. It shows the probability of acceptance in each nursery for two representative (average) families who differ only in their disadvantaged status. The differences are driven by interactions of disadvantages status with observed and unobserved nursery characteristics, as reported in Appendix Table A19. We observe substantial heterogeneity, yet all nurseries exhibit higher acceptance probabilities for advantaged families. This pattern is particularly pronounced for fixed-price institutions, but even income-based nurseries - which are expected to prioritize disadvantaged children - do not fall below the 45° line.

Finally, we compute marginal effects, i.e. the change in acceptance probabilities q_{ijr} resulting from small changes in characteristics; these are reported in Appendix Table A20. Several variables have large effects on acceptance probabilities. Having a sibling at the nursery increases acceptance by about 30 percentage points, while the

KUL priority (common among advantaged families) raises it by around 15 percentage points. Capacity and rank are also key determinants: a 10 percentage point increase in available capacity raises acceptance by about 1 percentage point, and ranking a nursery first rather than second increases acceptance by roughly 13 percentage points for advantaged families and 7 percentage points for disadvantaged families, with little variation across lower ranks. Applying later and not speaking Dutch at home are both associated with lower acceptance probabilities, patterns that disproportionately affect disadvantaged families.

7 Simulations of centralized mechanisms

In this section, we present simulations aimed at increasing the participation of disadvantaged families in day care. We first compare the current system to centralized alternatives, then study affirmative action policies, and finally examine the role of prices and capacities.

For all simulations, we divide each year into six enrollment periods defined by school vacations, mirroring preschool enrollment cycles. Each family is assigned to the period closest to its observed (or requested) start date.[28] We simulate matching sequentially over 20 periods between late 2013 and late 2016, drawing parental utilities and nursery priorities. At the start, we net out the capacity taken by children excluded from the analysis. After each round, we update capacities and endogenous nursery characteristics (enrollment and the share of disadvantaged children), and allow families to incorporate continuation values when forming utilities. To allow for burn-in, we report results only for the final 12 periods. We construct complete rank-order lists by simulating from the estimated utility distribution. As all simulated mechanisms are truth-revealing (except in decomposition exercises), we abstract from

---

[28] We use actual start dates for matched children and requested dates otherwise. Actual dates capture that parents may adjust their plans after failing to find a match, which we abstract from in the analysis. We also assume no subsequent switching across nurseries, which occurs in only 10.5% of cases.

Table 7: Simulations: centralization

Mechanism:	Decentralized	KDA	KDA	KDA	KDA	KDA	KTTC
Priorities:		Data	No rank	No time	Points (R)	Points (T)	Point (T)
Match (share of families)							
Unmatched	0.32	0.35	0.35	0.35	0.35	0.34	0.32
Advantaged	0.27	0.33	0.32	0.32	0.34	0.33	0.31
Disadvantaged	0.42	0.40	0.43	0.42	0.38	0.37	0.34
Match with rank 1	0.57	0.56	0.40	0.39	0.43	0.49	0.55
Advantaged	0.61	0.60	0.42	0.41	0.44	0.50	0.56
Disadvantaged	0.45	0.46	0.33	0.34	0.39	0.45	0.52
With envy		0.13	0.07	0.02	0.00	0.01	0.42
Advantaged		0.12	0.07	0.02	0.00	0.01	0.42
Disadvantaged		0.15	0.08	0.02	0.00	0.02	0.44
Welfare (Δ/capita in 1000 EUR)							
Parental benefit			-0.70	-0.78	-0.23	0.64	1.03
Advantaged			-0.69	-0.81	-0.45	0.53	0.98
Disadvantaged			-0.72	-0.69	0.33	0.90	1.16
Government cost			0.00	0.02	0.08	0.17	0.36
Segregation (0 to 1)							
Dissimilarity	0.25	0.25	0.25	0.26	0.28	0.32	0.28
Gini	0.35	0.36	0.36	0.36	0.39	0.43	0.39

Decentralized denotes simulations from estimated choices, acceptance probabilities and compliance statistics. Other simulations use KDA or KTTC mechanisms with different priorities. Data-driven priorities are given by the acceptance probabilities, added by a noise term to match % top ranked (inverse of a logit probability with type 1 extreme value shocks with scale 0.9). "No rank" is identical to "Data", except that it uses the priorities in rank 1 for all. "No time" is identical to "No rank", except that it removes the impact of the timing of application (i.e. the available capacity and the time left before the requested entry date). "Points (R)" uses points for having a sibling at the institution (4), working or studying (1), and living in the same neighborhood (1). For KUL-owned nurseries, we replace the location preference by a preference for working at KUL. "Points (T)" is identical to "Points (R)" but replaces the tie-breaker by travel time. Envy is defined as preferring an alternative nursery in which they have a higher priority than any family matched in the same period. Welfare differences are calculated with respect to the KDA mechanism with data-driven priorities and abstracting from ranking costs. We report averages over 25 simulations for the last 12 matching periods.

strategic ranking. Lists are formed based on $u_j^s(W_{ij})$ plus an additive shock.[29]

7.1 Centralizing the matching process

Table 7 shows the impact of centralizing the matching process. The first column

---

[29] To limit simulation noise, we report averages over 25 simulations. We also do not allow for new shock draws after rejection, consistent with ranking all nurseries truthfully at a single point in time, without defining a strategy for optimally choosing which option to rank next.

of results shows the predictions from the decentralized mechanism.[30] Because it is difficult to calculate welfare in this setting without further assumptions, column 2 replicates these results using a centralized algorithm.[31] We follow He et al. (2024) who impose stability in estimation in a decentralized setting, to then justify simulating a counterfactual using a DA algorithm. Instead of using a standard DA algorithm, we revert to the "Knapsack" DA (KDA) of Delacrétaz et al. (2023), which rejects families who cannot be accommodated by the nursery.[32] To stay close to the data, we use the estimated acceptance probabilities, plus a noise term, to define priorities.[33] Comparing the first two columns shows that we manage to replicate the decentralized setting quite well for most outcomes, although we find slightly fewer matches among the advantaged families.

In column 3, we remove the rank effect by giving all families the rank 1 acceptance probability, eliminating the bonus for higher-ranked alternatives (which gives rise to strategic ranking). This does not benefit any group: match rates remain similar, but average welfare drops to 700 EUR for platform participants. The number of matches with the first-ranked alternative also decreases.

Column 4 further equalizes conditions by giving all parents acceptance probabilities as if they applied simultaneously, with the same available capacity. Results are

---

[30] We use the estimates of utility, the shock distribution, acceptance probabilities and CCPs to predict the outcomes of the status quo. We also randomly allocate a share of the parents who matched with a nursery to an outside option instead, based on the numbers of Table A12.

[31] As utility is scaled in daily prices, we can quantify changes in average parental welfare by multiplying the utility of the allocated nursery with the total number of days to attend. We proceed similarly for the public cost of a spot, given an estimate of the daily cost, obtained from policy documents of government agency Kind & Gezin for the year 2016, see https://www.kindengezin.be/sites/default/files/2021-12/subsidiebedragen-kinderopvang.xlsx, downloaded in September 2024. Under some assumptions, we find a cost of 60EUR/day, minus the (after tax credit) income-based price paid for income-based spots and 3.4EUR/day + tax credit for fixed price institutions. We verified our assumptions by contacting the government agency to compare our calculation method applied to the 2024 subsidies and received almost identical numbers for the daily price.

[32] We thank the authors for providing their code on https://github.com/nhhai196/Refugee-Resettlement, which we adapted to our setting.

[33] Priorities are set to be $\ln(q_{ijr}/(1-q_{ijr})) + e_{ijr}$ with e_{ijr} an extreme value type 1 shock with scale parameter 0.9, calibrated to match the number of first-rank matches.

similar to column 3, suggesting that earlier applications by disadvantaged families in the decentralized setting would have little effect.

Columns 2-4 are decomposition exercises that would be hard to implement in practice. If priorities depend on rank, families may misreport their preferences, and eliciting priorities matching our estimated acceptance probabilities would be difficult. The following simulations propose more realistic centralized mechanisms that are easy to implement, align with nursery and policy priorities, and are incentive-compatible for truthful ranking.[34] We apply a transparent point system: sibling at the nursery (4), working or studying (1), and living in the same neighborhood (1).[35] For KUL nurseries, the location point is replaced by a point for working at KUL.

Column 5 applies the point system with a random tie-breaker. Results are mixed: disadvantaged families see slight gains in match rates and welfare, while advantaged families lose. This likely reflects that the point system aligns less with parental preferences than the current acceptance probabilities, as both preferences and priorities are negatively affected by travel time. Column 6 uses travel time as the tie-breaker instead, which improves outcomes compared to the baseline: disadvantaged families gain 900 EUR on average (up from 300), and advantaged families gain 500 EUR instead of losing 500. Segregation rises in both cases, especially in the latter.

Finally, column 7 switches from KDA to KTTC, the "knapsack" variant of Top-Trading Cycles (TTC). KTTC is Pareto efficient and thus expected to increase welfare, but it can generate justified envy, which DA avoids. We define envy here in the traditional sense, meaning that one can be envious of someone who got a spot, also if this was due to requesting other days of the week. Still, we see very little envy in our point-based KDA mechanisms, while 42% of families suffer from this in the KTTC mechanism. However, KTTC wins the game for one of the main outcomes we

---

[34] Unlike DA, KDA can be manipulable and may need adjustment: Delacrétaz et al. (2023) propose a TKDA system, which performs slightly worse (Appendix Table A16). Our main policy proposals use a KTTC mechanism, which avoids this issue.

[35] Values inspired by Table A17, regressing nurseries' self-reported priorities on characteristics (Appendix A.4).

care about: participation by disadvantaged families. The rate of unmatched disadvantaged families is now only 34% instead of 37% in KDA. Advantaged families also gain. Both groups gain around 1000 EUR extra welfare compared to the baseline, while the KDA only achieves an average of 600 EUR. Segregation is still higher than in the baseline, but lower than in KDA.

7.2 Affirmative action

In Table 8 we consider four different affirmative action policies. In all four cases, we implement a dynamic version of a soft-bounds policy. Such a policy prioritizes disadvantaged children until a quota is reached, but still allows other children to fill remaining seats if the quota is not met (Ehlers et al., 2014). We adapt this to our sequential setting by modifying priorities: disadvantaged children are given higher priority in income-based nurseries that did not reach the quota in the previous period. Since 28% of children are disadvantaged, we set the quota at 30%. Appendix Table A16 also reports results for a 20% quota.

We use KTTC without affirmative action as a benchmark and modify priorities in column 2. This leads to a substantial reduction in the unmatched rate among disadvantaged families (22% instead of 34%) and an average welfare gain of about 1,900 EUR for them. Advantaged families lose around 1,400 EUR. Somewhat surprisingly, segregation increases slightly.

Column 3 instead uses a KDA mechanism. As before, this entails a welfare loss relative to KTTC, but it substantially reduces justified envy. In addition, segregation now decreases markedly. The contrast reflects how priorities operate across mechanisms. Under KDA, affirmative action mainly benefits disadvantaged families by assigning them to nurseries where they are underrepresented, which reduces segregation. Under KTTC, higher priorities at nurseries that advantaged families particularly value can instead be used to form cycles, allowing disadvantaged families

Table 8: Simulations: affirmative action

	Base	Affirmative action			
Mechanism:	KTTC	KTTC	KDA	KTTC	KTTC
Compensation:				+ Cap 9%	+ av price
Match (share of families)					
Unmatched	0.32	0.32	0.34	0.26	0.32
Advantaged	0.31	0.36	0.38	0.30	0.35
Disadvantaged	0.34	0.22	0.25	0.17	0.26
Match with rank 1	0.55	0.52	0.46	0.54	0.52
Advantaged	0.56	0.50	0.46	0.52	0.50
Disadvantaged	0.52	0.55	0.44	0.58	0.58
With envy	0.42	0.46	0.20	0.42	0.45
Advantaged	0.42	0.51	0.27	0.48	0.51
Disadvantaged	0.44	0.33	0.01	0.29	0.30
Welfare (Δ/capita in 1000 EUR)					
Parental benefit		-0.47	-0.84	0.76	-0.31
Advantaged		-1.40	-1.54	0.06	-0.59
Disadvantaged		1.87	0.93	2.52	0.40
Government cost		0.00	-0.20	1.24	0.00
Segregation (0 to 1)					
Dissimilarity	0.28	0.30	0.20	0.32	0.32
Gini	0.39	0.41	0.31	0.43	0.43

Simulations of KTTC and KDA mechanisms with priority rule: points for having a sibling at the institution (4), working or studying (1), and living in the same neighborhood (1). For KUL-owned nurseries, we replace the location preference by a preference for working at KUL. Travel time between the nursery and the home location is used as a tie-breaker. Affirmative action: absolute priority for disadvantaged families if their current rate is below 30 %. Pricing, capacity and affirmative action policies considered here only affect income-based nurseries. "+ cap 9%" adds 9% of capacity. "+av price" applies the after tax cut average price of 10.5 EUR per day for all. Envy is defined as preferring an alternative nursery in which they have a higher priority than any family matched in the same period. Welfare differences are calculated with respect to the base (i.e. KTTC, no change in prices or capacity and no affirmative action). We report averages over 25 simulations for the last 12 matching periods.

to access more preferred options.[36] Since both groups are always better off under KTTC, we maintain this mechanism in what follows.

Columns 4 and 5 aim to compensate advantaged families for their losses. We find that they are indifferent when income-based capacity increases by 9%, at a fiscal cost of roughly 1200 EUR per applicant. Compensation without government spending is not feasible unless pricing becomes regressive. Column 5 abolishes progressive pricing in a budget-neutral way by charging a uniform price to all families. This, however, is still insufficient to fully compensate advantaged families. Notably, disadvantaged families would nevertheless prefer this policy to the current progressive scheme without affirmative action.

7.3 Importance of prices and capacities

Table 9 simulates alternative policies for income-based institutions. In all simulations, we use the KTTC algorithm with a point system and travel-time tie-breaking.

Column 2 highlights the role of progressive pricing by requiring all families to pay the maximum income-based price. Although the after-tax price of 23 EUR per day remains well below costs and international benchmarks, it raises the unmatched rate among disadvantaged families from 34% to 49% and leads to substantial welfare losses for parents. The larger reduction in government spending underscores the distributional concerns and potential externalities motivating public intervention.

Column 3 shows that making day care free has little effect on matching outcomes but substantially increases public spending (about 2,900 EUR per applicant), with most of the gains accruing to advantaged families who currently pay higher prices. Column 4 increases progressivity by requiring families in the highest income tercile to finance free day care for those in the lowest tercile. While this redistributes welfare toward disadvantaged families, it does not meaningfully reduce their unmatched rate,

---

[36] Appendix A.5 shows that segregation would be similar if all families were assigned to their preferred nursery, and decomposes the utility channels behind this result.

even though more than 75% of them are affected.

The final two columns are the policies that were combined with affirmative action policies in the previous table. Budget-neutral flat pricing redistributes welfare from disadvantaged to advantaged families and increases the unmatched rate for disadvantaged children by 5 percentage points. Increasing capacity by 9% at the start of the simulations raises participation by 5 percentage points for advantaged families and by 6 percentage points for disadvantaged families.[37] This policy costs about 1,200 EUR per applicant, but increases parental welfare by a similar amount, implying that even modest positive externalities of day care would make capacity expansion welfare improving.

8 Conclusion

Disadvantaged children often do not enter day care, despite well-documented gains. We estimate preferences using a novel method on unique data of parents' application decisions and day care acceptance decisions, and use these estimates to study how centralized matching mechanisms can increase welfare and improve access for disadvantaged children.

Our results have clear policy implications. First, a centralized matching algorithm is feasible and benefits disadvantaged families.[38] Designing priorities to reflect key preference shifters, such as travel time, is important. Second, affirmative action policies are easy to implement in a centralized system and effectively increase participation by disadvantaged children. However, they may raise segregation and are generally disliked by advantaged families. Third, subsidized prices are crucial for disadvantaged families' participation. Since out-of-pocket fees are already low, further price progressivity yields limited gains. In contrast, expanding subsidized capacity

---

[37] Recall that results are reported for the final 12 of 20 periods.

[38] Our proposed mechanism also improves outcomes for advantaged families relative to the replicated decentralized mechanism. However, since we do not fully match their observed match rates, we interpret these gains with caution.

Table 9: Simulations: prices and capacity

	Base	Price High	Price Free	Price More prog	Price Average	Capacity Extra 9%
Match (share of families)						
Unmatched	0.32	0.34	0.32	0.32	0.33	0.26
Advantaged	0.31	0.29	0.31	0.32	0.30	0.26
Disadvantaged	0.34	0.49	0.35	0.33	0.39	0.28
Match with rank 1	0.55	0.57	0.53	0.55	0.56	0.57
Advantaged	0.56	0.58	0.54	0.56	0.56	0.58
Disadvantaged	0.52	0.57	0.51	0.51	0.54	0.55
With envy	0.42	0.31	0.47	0.43	0.41	0.38
Advantaged	0.42	0.32	0.47	0.42	0.41	0.38
Disadvantaged	0.44	0.27	0.48	0.45	0.40	0.40
Welfare (Δ/capita in 1000 EUR)						
Parental benefit		-2.19	2.54	-0.03	0.33	1.20
Advantaged		-1.70	3.19	-0.22	0.93	1.41
Disadvantaged		-3.41	0.93	0.47	-1.20	0.67
Government cost		-3.41	2.89	0.05	0.10	1.22
Segregation (0 to 1)						
Dissimilarity	0.28	0.31	0.28	0.28	0.29	0.28
Gini	0.39	0.43	0.38	0.39	0.40	0.39

Simulations of KTTC mechanisms with priority rule: points for having a sibling at the institution (4), working or studying (1), and living in the same neighborhood (1). For KUL-owned nurseries, we replace the location preference by a preference for working at KUL. Travel time between the nursery and the home location is used as a tie-breaker. Pricing and capacity policies considered here only affect income-based nurseries. "High price" set prices to the after tax cut maximum of 23.09 EUR per day. "More prog price" sets price to 0 for those who currently have a pre-tax price below 10, and increases it by their average (2.75 EUR per day) for those with a pre-tax price above 20. "Average price" are set to the after tax cut average of 10.5 EUR per day for everyone. Income-based capacity increased by 9 % in last column. Envy is defined as preferring an alternative nursery in which they have a higher priority than any family matched in the same period. Welfare differences are calculated with respect to the base (i.e. no change in prices or capacity). We report averages over 25 simulations for the last 12 matching periods.

substantially increases participation, and its public cost is justified by the welfare gains.

Further research could explore how to handle cases where parents are flexible in the timing of their demand, both within a week and across enrollment months. It would also be valuable to examine how improved matching policies affect parents' labor market outcomes and children's educational outcomes, to micro-found the estimated parental welfare gains and quantify additional benefits from positive externalities.

References

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017). The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match. *American Economic Review*, 107(12):3635–3689.

Abdulkadiroğlu, A. and Sönmez, T. (2003). School Choice: A Mechanism Design Approach. *The American Economic Review*, 93(3):729–747.

Agarwal, N., Ashlagi, I., Rees, M., Somaini, P., and Waldinger, D. (2021). Equilibrium Allocations under Alternative Waitlist Designs: Evidence from Deceased Donor Kidneys. *Econometrica*.

Agarwal, N. and Somaini, P. (2018). Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism. *Econometrica*, 86(2):391–444.

Agarwal, N. and Somaini, P. (2020). Revealed Preference Analysis of School Choice Models. *Annual Review of Economics*, 12:471–501.

Arcidiacono, P., Aucejo, E., Maurel, A., and Ransom, T. (2025). College Attrition and the Dynamics of Information Revelation. *Journal of Political Economy*, 133(1):53–110.

Arcidiacono, P. and Miller, R. A. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 79(6):1823–1867.

Baker, M., Gruber, J., and Milligan, K. (2008). Universal Child Care, Maternal Labor Supply, and Family Well-Being. *Journal of Political Economy*, 116(4):709–745.

Baker, M., Gruber, J., and Milligan, K. (2019). The Long-Run Impacts of a Universal Child Care Program. *American Economic Journal: Economic Policy*, 11(3):1–26.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 63(4):841.

Beuermann, D. W., Jackson, C. K., Navarro-Sola, L., and Pardo, F. (2023). What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output. *The Review of Economic Studies*, 90(1):65–101.

Burgess, S., Greaves, E., Vignoles, A., and Wilson, D. (2014). What Parents Want: School Preferences and School Choice. *The Economic Journal*.

Caetano, G. and Maheshri, V. (2017). School segregation and the identification of tipping behavior. *Journal of Public Economics*, 148:115–135.

Calsamiglia, C., Fu, C., and Güell, M. (2020). Structural Estimation of a Model of School Choices: the Boston Mechanism vs. Its Alternatives. *Journal of Political Economy*, 128(2).

Che, Y.-K., Grenet, J., and He, Y. (2025). Allocating Students to Schools: Theory, Methods, and Empirical Insights. In *Handbook of the Economics of Matching*, volume 2, pages 307–407. Elsevier.

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2018). Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance. *Journal of Political Economy*, 126(6):2356–2409.

Delacrétaz, D., Kominers, S. D., and Teytelboym, A. (2023). Matching Mechanisms for Refugee Resettlement. *American Economic Review*, 113(10).

Drange, N. and Havnes, T. (2019). Early childcare and cognitive development: Evidence from an assignment lottery. *Journal of Labor Economics*, 37(2).

Ehlers, L., Hafalir, I. E., Yenmez, M. B., and Yildirim, M. A. (2014). School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic Theory*, 153:648–683.

Felfe, C. and Lalive, R. (2018). Does early child care affect children's development? *Journal of Public Economics*, 159:33–53.

Fort, M., Ichino, A., and Zanella, G. (2020). Cognitive and Noncognitive Costs of Day Care at Age 0–2 for Children in Advantaged Families. *Journal of Political Economy*, 128(1):158–205.

Gaer, E. V., Gijselinckx, C., and Hedebouw, G. (2013). Het gebruik van opvang voor kinderen jonger dan 3 jaar in het Vlaamse Gewest. Technical Report 12, Steunpunt Welzijn, Volksgezondheid en Gezin.

Gathmann, C. and Sass, B. (2018). Taxing Childcare: Effects on Childcare Choices, Family Labor Supply, and Children. *Journal of Labor Economics*, 36(3):665–709.

Gazmuri, A. M. (2020). School Segregation in the Presence of Student Sorting and Cream-Skimming: Evidence from a School Voucher Reform. *Working paper*.

Hafalir, I. E., Yenmez, M. B., and Yildirim, M. A. (2013). Effective affirmative action in school choice: Effective affirmative action in school choice. *Theoretical Economics*, 8(2):325–363.

Hastings, J., Kane, T., and Staiger, D. (2009). Heterogeneous preferences and the efficacy of public school choice. *NBER Working Paper*, 2145.

Havnes, T. and Mogstad, M. (2011). No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes. *American Economic Journal: Economic Policy*, 3(2):97–129.

He, Y., Sinha, S., and Sun, X. (2024). Identification and Estimation in Many-to-One Two-Sided Matching Without Transfers. *Econometrica*, 92(3):749–774.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1):153.

Hermes, H., Lergetporer, P., Peter, F., and Wiederhold, S. (2024). Application barriers and the socioeconomic gap in child care enrollment. *Journal of the European Economic Association*, (forthcoming).

Hotz, V. J. and Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497–529.

Idoux, C. (2023). Integrating New York City Schools: The Role of Admission Criteria and Family Preferences. *Working paper*.

Kind&Gezin (2019). *Toegankelijkheid en diversiteit: Voorrangsgroepen en opnamebeleid*.

Laverde, M. (2021). Unequal Assignments to Public Schools and the Limits of School Choice. *Working paper*.

Lee, J. and Son, S. J. (2022). Distributional Impacts of Centralized School Choice. *Working paper*.

Magnac, T. and Thesmar, D. (2002). Identifying dynamic discrete decision processes. *Econometrica*, 70(2):801–816.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531–542.

Murphy, A. D. (2018). A Dynamic Model of Housing Supply. *American Economic Journal: Economic Policy*.

OECD (2011). *Doing Better for Families*.

Oosterbeek, H., Sóvágó, S., and van der Klaauw, B. (2021). Preference heterogeneity and school segregation. *Journal of Public Economics*, 197:104400.

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review*, 109(3):774–809.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55(5):999–1033.

Teppers, E., Schepers, W., and Van Regenmortel, T. (2019). Het gebruik en de behoefte aan kinderopvang voor baby's en peuters jonger dan 3 jaar in het Vlaamse Gewest. Technical Report 25, Steunpunt Welzijn, Volksgezondheid en Gezin.

Van Lancker, W. and Ghysels, J. (2012). Who benefits? The social distribution of subsidized childcare in Sweden and Flanders. *Acta Sociologica*, 55(2):125–142.

Van Lancker, W. and Vandenbroeck, M. (2019). De verdeling van de kinderopvang in Vlaanderen en in de centrumsteden: spanning tussen de economische en sociale functie van kinderopvang. Technical report, Centrum voor Sociologisch Onderzoek, KU Leuven and UGent.

Wang, A., Wang, S., and Ye, X. (2025). Option value neglect: Evidence from centralized college admission. *Working paper*.

A Appendix

A.1 Data cleaning

A.1.1 Datasets received

In march 2017, we received three datasets from the platform that manages day care allocation in the city of Leuven, kinderopvang.be. The dataset was created by an IT consulting firm that manages the platform. The data were collected with the purpose of managing the platform, we obtained a snapshot in March 2017, but also historical data that was collected in the preceding years. Note that the data contains both day care data, as well as after-nursery day care for toddlers. In the data cleaning, we will restrict this to day care (before age 2.5).

- **bestand_akdv.csv**

This file lists the rankings of families over day care centers. Each line corresponds to an action by the family (id = dossiernr), such as adding an alternative to the rank-ordered list, or the day care ("opvanginitiatief"), such as offering a spot. This dataset also contains characteristics of the family, the requested (half)days of day care and the time span for which they want it, as well as a timestamp ("aanmaakdatum") for each entry. Time stamps range from November 2011 until March 2017.

- **bestand_inschrijving.csv**

This file lists the enrollment of children over day care centers and childminders. Each child can go to more than one place, therefore each line corresponds to a child (id=dossier_nr) - day care ("opvanginitiatief") combination. It also contains the time span and days of the week the day care is used and for some day cares the price that is paid. Starting dates range from January 2003 (ignoring likely mistakes) until October 2019. End dates range from March 2004 until January 2027.

- **bestand_opvanglocaties.csv**

This file contains information of each day care such as address, name, id and capacity at different moments in time. We also supplement it with information recovered from website kinderopvangleuven.be.

- **priorityscores.xlsx**

We added an excel file using print screens taken in November 2023 and sent by Loket Kinderopvang Leuven of the priority scores that are used by some nurseries to sort applications.

A.1.2 Data cleaning

For the purpose of this paper, we restrict our attention to a subset of the data. As the platform gradually started in 2011 and we received a snapshot at the beginning of 2017, we restrict attention to familiesthat want to start day care between week 26 of 2013 and week 52 of 2016. We also ignore requests made before week 26 of 2012. Since the enrollment data also includes children allocated before the start of the system, we can use information on all allocated children during this time period too. Table A10 summarizes the data cleaning and sample selection process.

We restrict attention to day cares that target <2.5 year olds. To do this we calculate the number of children starting in each day care that are younger and older than age 2 at the moment they start. If less than 10% is younger than age 2, we assume it is not a day care. For day care centers we observe this information and it confirms the use of this cutoff.

We restrict attention to children living within 20km of the centroid of the city of Leuven. We only consider day care centers that are considered by at least 50 families over the relevant sample. We restrict attention to first allocations and first accepted offers.

For the purpose of creating day care characteristics, such as % disadvantaged children, we did not drop observations and used all available information.

Table A10: Data selection

	Total number of children
All rankings before data cleaning	11959
Drop information that is not about a ranked alternative	11496
Drop childminders	11460
Dropping if requested starting date before 2013w26	8841
Dropping if requested starting date after 2016w52	5864
Drop observations with negative requested duration	5858
Drop orderings after first allocation	5704
Drop orderings after requested starting date	5679
Drop if more than 20km from centroid Leuven	5383
Drop if request made before 2012w26	5348
Drop if the highest rank is no longer included because of previous drops	5272
Drop if starting before or on the day of ranking the alternative	5119
Drop entire observation if request day care that we omitted	5110
Drop if missing observables (not possible)	5110
Drop if gaps in ranking	5011
Drop if unable to find travel time	4922

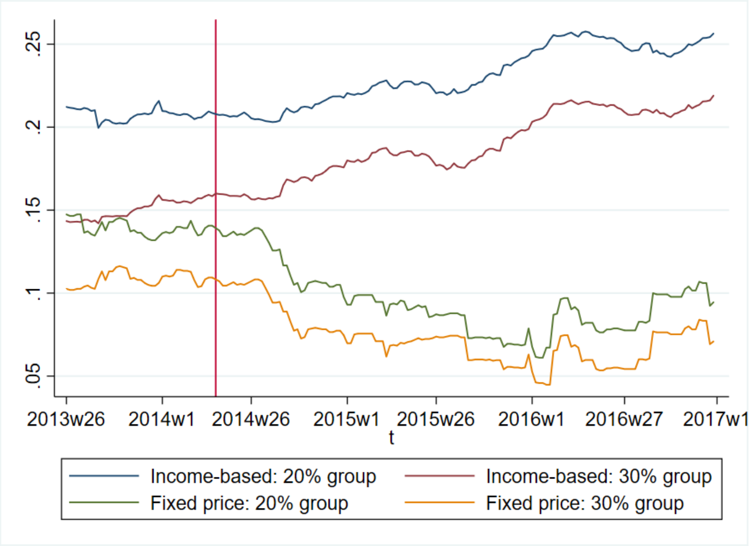
A.2 Impact of the 2014 policy change

The 2014 policy change (see section 2.1) aimed to increase the participation of disadvantaged children by requiring income-based institutions to prioritize them. Figure A1 shows that the share of disadvantaged groups increases after the policy change for both types of priority groups. Income-based institutions saw an increase, while fixed-price institutions saw a decrease. Note that almost 90 % of the capacity is in income-based institutions, showing this is not just a compositional shift. However, we should be cautious in interpreting this result as we cannot exclude other changes over time. Because of equilibrium effects, the SUTVA assumption is violated for fixed-price institutions so we cannot use them for a difference-in-differences estimator.[39] It is also unclear how the policy was enforced as nurseries are not fined if they can argue why compliance was difficult, e.g. due to a lack of demand from priority groups.

---

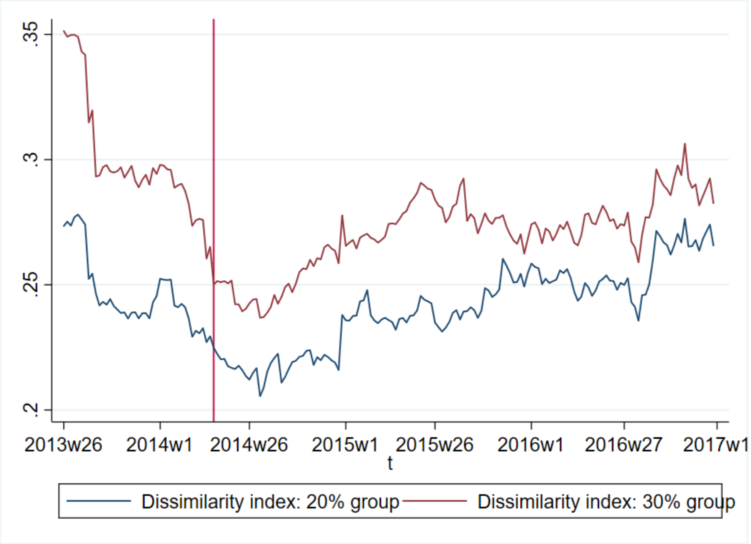
[39] We also noticed that the share of children of different groups finding a match (not shown here) remains constant over time, casting doubt on the policy having an impact on the acceptance behavior of nurseries.

Figure A1: Share of priority groups in day care



Note: the 20% and 30% priority groups refer to definitions of disadvantaged groups defined by government policy, explained in section 2.1. The vertical line denotes the start of the new priority policy.

Figure A2: Dissimilarity index by priority groups in day care



Note: the 20% and 30% priority groups refer to definitions of disadvantaged groups defined by government policy, explained in section 2.1. The vertical line denotes the start of the new priority policy. The dissimilarity index takes the sum of the absolute difference in the relative share of each group within each day care and divides it by 2. Higher numbers denote more segregation.

Figure A2 shows the dissimilarity index to describe how segregation differs after the policy change. Interestingly, the increase in attendance of disadvantaged children goes along with an increase in segregation. This goes against the common intuition in school choice. Offering better opportunities for disadvantaged children is expected to give them access to better schools that are usually dominated by advantaged children. This figure suggests this is not necessarily the case for day care. Again, caution is advised as other things might have changed over time. For example, we see that segregation was decreasing before the policy change, so it is possible there is a downward trend over time which would lead to an underestimation of the effects of the policy change based on this figure.

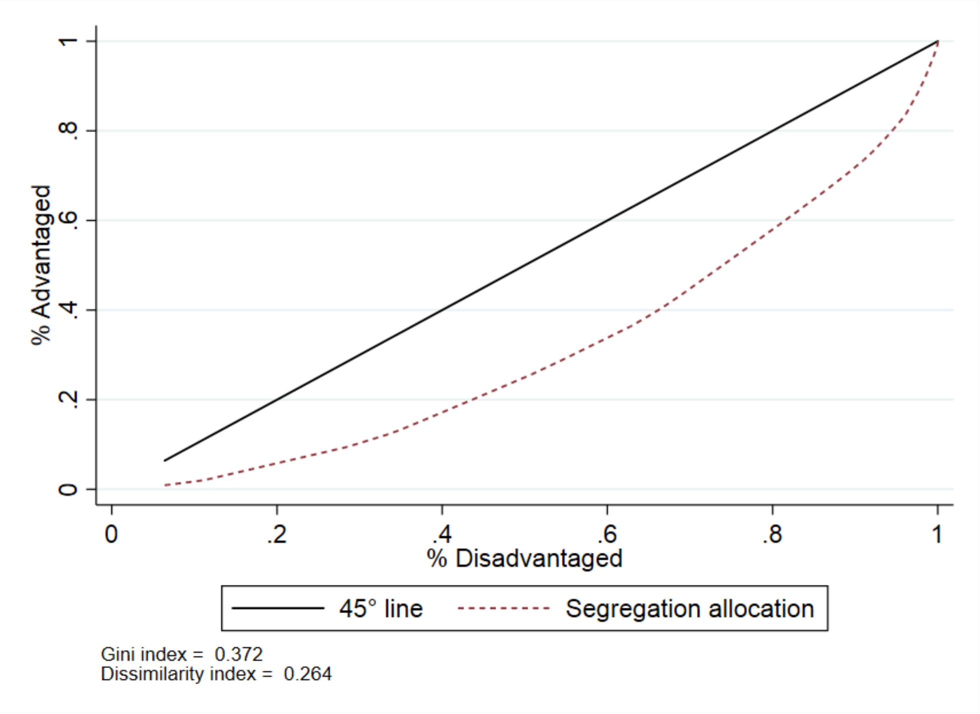
A.3 Appendix Tables and Figures

Table A11: Day care characteristics in December 2016

	Unweighted			**Weighted by capacity**		
	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Overall						
Capacity	108	19.796	32.743	2,138	73.452	63.165
Income-based	108	0.870	0.337	2,138	0.917	0.276
Nursery	108	0.370	0.485	2,138	0.870	0.336
Nurseries only						
Capacity	40	46.525	42.147	1,861	83.752	61.355
Income-based	40	0.825	0.385	1,861	0.927	0.260
KUL	40	0.125	0.335	1,861	0.215	0.411
Distance to city center (km)	40	2.632	1.663	1,861	2.404	1.459
Distance to east (km)	40	3.029	1.482	1,861	3.258	1.436
Distance to north (km)	40	4.787	1.905	1,861	5.119	1.691
Distance to south (km)	40	4.157	2.001	1,861	3.637	1.883
Distance to west (km)	40	3.868	2.039	1,861	3.313	2.051
Operating hours	40	10.961	0.887	1,861	11.474	1.167

Child care characteristics in December 2016.

Figure A3: Segregation in day care



Notes: Nurseries are ranked from lowest to highest share of advantaged children. The x-axis shows the cumulative share of disadvantaged children and the y-axis the cumulative share of advantaged children. Perfect integration corresponds to the 45° line. The figure indicates segregation: day cares serving 40% of disadvantaged children account for slightly less than 20% of advantaged children. This pattern is reflected in a Gini coefficient of 0.37 and a dissimilarity index of 0.26. The Gini coefficient measures the area between the segregation curve and the 45° line relative to the total area under the 45° line. The dissimilarity index equals one half of the sum of absolute differences in group shares across day cares.

Table A12: Match statistics

	Overall	**Advantaged**	**Disadvantaged**
Final allocation			
Nursery	0.682	0.699	0.633
Childminder	0.069	0.052	0.115
None	0.249	0.249	0.252
Behavior after acceptance nursery (3762 children)			
Nursery: highest offer	0.752	0.764	0.714
Nursery: lower offer	0.044	0.045	0.041
Nursery: no offer	0.051	0.049	0.058
Childminder	0.036	0.030	0.056
Leave	0.116	0.112	0.131
Behavior after no offers nursery (1160 children)			
Childminder	0.174	0.135	0.246
Leave	0.682	0.768	0.522
Nursery: no offer	0.144	0.097	0.232

This table shows how well the results of the ranking system correspond with actual attendance behavior. The large majority of families comply with the system. After acceptance by a nursery, 75% of families attend their highest accepted alternative. 3.6% instead chooses a childminder and 12% does not end up going anywhere. In 4.4% of cases, the family does not follow the rank order and goes to a lower-ranked, accepted alternative, while in 5.1% of cases they go to a place they did not rank. Differences between advantaged and disadvantaged groups are small. Out of the families who are not accepted by any nursery, the majority do not go to day care in Leuven, but we see a more substantial number of disadvantaged children still being allocated to a childminder or nursery.

Table A13: Demand side: estimation results (2 of 3)

	Advantaged		**Disadvantaged**	
	coef	se	coef	se
β_s in equation (1)				
Operating hours	5.558	(1.061)	4.126	(0.827)
× Type B	0.771	(0.279)	-1.131	(0.255)
× No Dutch	-1.969	(0.617)	-0.604	(0.301)
× Low income	-0.146	(1.116)	-0.172	(0.428)
× Single parent			0.100	(0.383)
× Low education	0.979	(0.988)	0.047	(0.347)
× Work or study	0.064	(0.954)	-0.382	(0.495)
Distance to center (km)	-0.163	(2.388)	-9.887	(1.800)
× Type B	-7.929	(0.884)	12.396	(1.327)
× No Dutch	-6.772	(1.319)	-3.026	(0.731)
× Low income	-4.603	(3.535)	-1.211	(1.015)
× Single parent			-2.502	(0.790)
× Low education	-2.426	(2.074)	-4.942	(0.950)
× Work or study	1.216	(2.259)	3.288	(1.062)
Distance to north (km)	-6.242	(1.079)	2.619	(1.132)
× Type B	5.564	(0.424)	-2.956	(0.518)
× No Dutch	3.651	(0.715)	0.821	(0.492)
× Low income	4.431	(1.617)	0.874	(0.553)
× Single parent			0.780	(0.495)
× Low education	0.839	(1.075)	0.786	(0.471)
× Work or study	1.121	(1.150)	-1.481	(1.010)
Distance to east (km)	3.339	(1.391)	5.625	(1.051)
× Type B	2.358	(0.452)	-4.202	(0.623)
× No Dutch	2.251	(0.854)	0.762	(0.443)
× Low income	-1.455	(1.935)	-1.056	(0.633)
× Single parent			0.670	(0.430)
× Low education	2.093	(1.553)	1.219	(0.606)
× Work or study	-0.590	(1.312)	-1.303	(0.674)
Distance to south (km)	-7.478	(1.272)	2.052	(1.251)
× Type B	5.204	(0.507)	-3.802	(0.608)
× No Dutch	3.000	(0.703)	0.137	(0.537)
× Low income	4.809	(1.764)	0.028	(0.672)
× Single parent			0.915	(0.535)
× Low education	0.014	(1.023)	1.942	(0.451)
× Work or study	0.705	(1.288)	-1.079	(1.002)
Distance to west (km)	5.976	(1.638)	3.771	(1.300)
× Type B	-1.669	(0.462)	-3.464	(0.651)
× No Dutch	0.510	(0.991)	0.861	(0.582)
× Low income	-1.413	(2.185)	-0.216	(0.806)
× Single parent			0.888	(0.605)
× Low education	0.893	(1.549)	1.442	(0.610)
× Work or study	-0.620	(1.487)	-0.737	(0.746)

Utility scaled in daily prices. Other results can be found in Table 6 and A14. Group-specific nursery fixed effects included. The constant and baseline effects of operating hours and distances are obtained from a second stage regression of the nursery fixed effects. Bootstrap standard errors in parentheses, based on 50 replications.

Table A14: Demand side: estimation results (3 of 3)

	Advantaged		**Disadvantaged**	
	coef	se	coef	se
β_s in equation (1)				
Constant	25.519	(16.765)	-68.796	(13.324)
Type B	-83.947	(6.161)	73.627	(8.233)
No Dutch	-6.040	(9.830)	-0.521	(4.312)
Low income	-24.978	(18.406)	-2.479	(6.337)
Single parent	6.307	(10.336)	-10.977	(6.270)
Low education	-21.353	(14.278)	-7.391	(5.793)
Work or study	-7.935	(17.231)	19.410	(10.103)
Cluster 2	-9.185	(2.482)	-6.993	(1.473)
× Type B	2.519	(0.837)	-1.903	(0.622)
× No Dutch	4.507	(1.061)	1.435	(0.787)
× Low income	3.534	(2.312)	2.472	(0.968)
× Single parent			1.510	(0.847)
× Low education	-1.360	(1.888)	1.888	(0.826)
× Work or study	-3.940	(2.354)	-2.420	(0.908)
Cluster 3	1.525	(2.051)	2.900	(1.276)
× Type B	-1.180	(0.537)	-1.208	(0.547)
× No Dutch	-1.221	(0.957)	-0.815	(0.725)
× Low income	3.162	(1.677)	-0.987	(0.741)
× Single parent			0.102	(0.739)
× Low education	-1.043	(1.239)	-0.815	(0.736)
× Work or study	0.952	(1.835)	-0.923	(0.815)
Cluster 4	4.302	(1.993)	3.580	(1.737)
× Type B	1.245	(0.645)	-3.976	(0.641)
× No Dutch	-0.469	(1.344)	0.375	(0.753)
× Low income	-0.131	(2.608)	1.404	(0.857)
× Single parent			-0.607	(0.859)
× Low education	4.877	(2.245)	-0.559	(0.851)
× Work or study	-2.062	(1.964)	-1.398	(1.425)

Utility scaled in daily prices. Other results can be found in Table 6 and A13. Group-specific nursery fixed effects included. The constant and baseline effects of operating hours and distances are obtained from a second stage regression of the nursery fixed effects. Bootstrap standard errors in parentheses, based on 50 replications.

Table A15: Average marginal effects of logit predicting probability to be strategic

	(1) Advantaged	(2) Disadvantaged
Type B	0.010	0.057
	(0.009)	(0.017)
Weeks left	-0.002	-0.001
	(0.000)	(0.000)
No Dutch	-0.012	-0.043
	(0.011)	(0.017)
Low income	-0.018	0.023
	(0.023)	(0.020)
Low education	-0.020	-0.010
	(0.019)	(0.020)
Work or study	0.029	0.027
	(0.016)	(0.023)
Single parent		0.041
		(0.020)

Standard errors in parentheses

Table A16: Simulations: appendix

Mechanism: AA threshold:	KTTC	KTTC 30%	KTTC 20%	KDA	TKDA
Match (share of families)					
Unmatched	0.32	0.32	0.32	0.34	0.35
Advantaged	0.31	0.36	0.34	0.33	0.34
Disadvantaged	0.34	0.22	0.28	0.37	0.37
Match with rank 1	0.55	0.52	0.53	0.49	0.48
Advantaged	0.56	0.50	0.53	0.50	0.49
Disadvantaged	0.52	0.55	0.54	0.45	0.44
With envy	0.42	0.46	0.44	0.01	0.00
Advantaged	0.42	0.51	0.47	0.01	0.00
Disadvantaged	0.44	0.33	0.38	0.02	0.00
Welfare (Δ/capita in 1000 EUR)					
Parental benefit		-0.47	-0.23	-0.39	-0.49
Advantaged		-1.40	-0.71	-0.44	-0.55
Disadvantaged		1.87	0.98	-0.26	-0.32
Government cost		0.00	-0.02	-0.19	-0.27
Segregation (0 to 1)					
Dissimilarity	0.28	0.30	0.28	0.32	0.32
Gini	0.39	0.41	0.38	0.43	0.43

Simulations of KTTC, KDA and TKDA mechanisms with priority rule: points for having a sibling at the institution (4), working or studying (1), and living in the same neighborhood (1). For KUL-owned nurseries, we replace the location preference by a preference for working at KUL. Travel time between the nursery and the home location is used as a tie-breaker. Affirmative action (AA): absolute priority for disadvantaged families if their current rate is below 20 or 30 %. Envy is defined as preferring an alternative nursery in which they have a higher priority than any family matched in the same period. Welfare differences are calculated with respect to the base (i.e. KTTC and no affirmative action). We report averages over 25 simulations for the last 12 matching periods.

A.4 Acceptance probabilities

This appendix discusses the estimation and results of the acceptance probabilities q_{ijr}, which serve as beliefs for strategic agents in our demand model.

A.4.1 Duration model

When parents rank a nursery, nurseries can respond by providing an offer, or by rejecting them. Note that parents can stay on a waiting list when the nursery rejects them but no longer considers them if they accept another offer (or leave). The data on acceptance should therefore be interpreted as survival data with right-censoring. Another reason to consider censoring is that parents need day care by a specified date.

Let q_{ijr} be the probability for family i to receive an offer from alternative j in rank r. We assume that the time it takes to be accepted follows a Weibull distribution. q_{ijr} is then given by the counterpart of the survival function at the time of the slot:

$$q_{ijr} = 1 - \exp(-\lambda_{ijr}\iota_i^p). \tag{6}$$

We specify $\lambda_{ijr} = \exp(Z_{ijr}\alpha)$ with Z_{ijr} a vector of family and nursery characteristics, and dummy variables for the rank and nursery. Note that nursery characteristics can change over time. We use them at the time the family i ranks them. ι_i denotes the number of weeks between the application and the starting date. α and p are parameters to estimate.[40]

We observe the decisions made by nurseries when they receive an application and use this to estimate acceptance probabilities q_{ijr}, given by equation (6), and estimated via a Weibull duration model. We include a rich list of variables, most of which are also included in the demand model. Several are expected to have an important impact.

---

[40] Note that households can add alternatives to their ROL at a later time, thereby requesting spots at different times. This will change the relevant time lag and the time-varying nursery characteristics. We abstract from this in the notation here but allow for it in estimation.

Many nurseries announce that they prioritize siblings, and KUL facilities prioritize their own employees. Importantly, we also include several variables that are excluded from demand. This improves the prediction of q_{ijr} and identifies the share of strategic families in the sample, without relying on arbitrary functional form assumptions:

Exclusion restrictions

Points Each nursery can give points for observable family characteristics. When they sort families to allocate, they can do this based on the points they have. It could therefore be reflective of the preferences of a specific nursery. To make the points comparable across nurseries, we normalize them to be mean 0 and standard deviation 1 within each nursery. We include them for the 24 out of 45 nurseries that make use of this and add a dummy for making use of the points system (in specifications without fixed effects). Table A17 shows a linear regression of the normalized points on the characteristics used. We see that sibling, same neighborhood and the KUL priority are often attributed a large number of points, but also the disadvantaged category is used, suggesting that nurseries take into account the priority policy.

Available capacity We calculate the available capacity for each requested day at the time a family i applies for a spot. To know the relevant available capacity for each day requested in nursery j, we subtract the enrolled children that applied before i from the total capacity of each specific day. We use the share available and we then take the minimum value over all days requested as a variable to include.

Timing We include the total number of days requested and dummy variables for each day of the week requested (omitting Monday), each month of the year and each year. We also include the age of the child at entry, the total weeks demanded and a polynomial for the time left until the starting date.

Table A17: Points used by nurseries (standardized)

Variables	coef	se
Sibling	3.942	(0.608)
KUL priority	0.550	(0.095)
Disadvantaged	0.699	(0.189)
30% priority group	-0.045	(0.058)
Work or study	0.084	(0.033)
Low income	0.072	(0.045)
Single parent	0.030	(0.078)
At least 5 days per week	0.031	(0.049)
Same neighborhood	0.663	(0.189)
Rank = 1	0.153	(0.121)
Rank = 2	0.060	(0.051)
Constant	-0.589	(0.103)
Observations	15,334	
R-squared	0.563	

Standard errors clustered within nursery.

Rank As nurseries can see their rank, we include dummy variables for rank 1 to 5 and one for 6 or higher. This is to account for the fact that nurseries might prefer to be better ranked, as was for example the case for schools in New York when it was using a similar decentralized matching system (Abdulkadiroğlu et al., 2017).

Policy target We calculate how much an income-based nursery is currently below its yearly target of 20% disadvantaged children for each year a family requests. We set it to 0 if the end of the year still takes more than 6 months or if they are above the target.

Discontinuities in location In the demand model, we use travel time by car and the relative difference by public transportation. For the acceptance probabilities, we still control for travel time by car but we also account for discontinuities by including a dummy for living in the official neighborhood of the nursery as this is also a criterion used for the points.

A.4.2 Estimates

Table A18 and A19 show the estimates of α in three specifications. The final one is the richest and is also used in the paper to calculate q_{ijr}. Marginal effects on q_{ijr} are discussed in the main text and can be found in Table A20. The signs in the estimation table can be interpreted as an increase (+) or decrease (-) in the hazard to receive an acceptance. For dummy variables, it is also easy to interpret the magnitude of the effect by $\exp(\alpha)$, which denotes the hazard ratio of an agent with dummy = 1, compared to 0. It also shows an estimate of $\ln p$. As this is negative in each specification, it shows that the hazard of being accepted decreases over time.[41]

The first specification includes only a dummy for being disadvantaged. We find a negative impact corresponding to a hazard rate of 85% of that of advantaged families ($exp(-0.160) = 0.85$). The second specification adds the variables that are excluded from demand but expected to influence acceptance probabilities. Indeed, we see positive effects of points, available capacity and higher-ranked nurseries. Especially the latter is important to highlight: the hazard ratio is only 30% for an alternative ranked second instead of first. Surprisingly, the effect of the 20% priority policy goes in the opposite direction but is also imprecisely estimated.

The final specification adds the variables also used in the demand model and interacts all with the disadvantaged status. We add nursery fixed effects in this specification, but let heterogeneity by disadvantaged status go through observables and clusters as some nurseries send very few acceptances to this group (clusters were defined in the paper when discussing heterogeneity on the demand side). The main text provides a more detailed analysis of the quantitative importance of different variables by using a marginal effects approach, but it is useful here to highlight the impact on estimates of adding more controls. The effect of points becomes negligible once

---

[41] Note however that it is partly counter-balanced by the tendency to prioritize urgent requests, captured by a polynomial in months between the date of demand and requested starting date. It is important to keep in mind that we condition on available capacity (which decreases when the deadline approaches) and that these are effects on the hazard, not on the predicted acceptance probability which also directly depends (positively) on the time available (see ι in (6)).

Table A18: Supply side: estimation results (part 1 of 2)

	(1)		(2)		(3)	
	coef	se	coef	se	coef	se
Disadvantaged	-0.160	(0.036)	-0.421	(0.043)	0.000	(0.927)
Points			0.250	(0.010)	0.045	(0.022)
× disadvantaged					-0.038	(0.038)
% capacity			0.568	(0.075)	1.068	(0.119)
× disadvantaged					-0.048	(0.210)
Weeks demanded			0.001	(0.001)	0.002	(0.001)
× disadvantaged					-0.001	(0.002)
Age at start			-0.004	(0.001)	-0.002	(0.001)
× disadvantaged					0.001	(0.002)
Days per week			0.031	(0.051)	0.140	(0.056)
× disadvantaged					-0.426	(0.136)
Tuesday			-0.098	(0.086)	-0.171	(0.095)
× disadvantaged					0.539	(0.232)
Wednesday			-0.029	(0.068)	-0.077	(0.076)
× disadvantaged					0.371	(0.177)
Thursday			-0.093	(0.074)	-0.135	(0.082)
× disadvantaged					0.083	(0.188)
Friday			0.226	(0.071)	0.128	(0.078)
× disadvantaged					0.509	(0.189)
Saturday			0.492	(0.247)	0.925	(0.309)
× disadvantaged					-0.826	(0.548)
Below target priority policy			0.091	(0.045)	0.050	(0.057)
× disadvantaged			-0.163	(0.087)	-0.085	(0.113)
Same neighborhood			0.052	(0.044)	0.108	(0.054)
× disadvantaged					-0.062	(0.114)
Rank: 2			-1.213	(0.046)	-1.257	(0.055)
× disadvantaged					0.493	(0.108)
Rank: 3			-1.516	(0.054)	-1.562	(0.065)
× disadvantaged					0.537	(0.124)
Rank: 4			-1.653	(0.060)	-1.696	(0.072)
× disadvantaged					0.433	(0.139)
Rank: 5			-1.681	(0.065)	-1.811	(0.079)
× disadvantaged					0.668	(0.145)
Rank: >5			-1.344	(0.050)	-1.438	(0.059)
× disadvantaged					0.327	(0.119)
Ln(p)	-0.513	(0.014)	-0.332	(0.014)	-0.270	(0.013)
Constant	-3.551	(0.033)	-0.738	(0.164)	FE	
Observations	23,596		23,596		23,596	
Polynomial months remaining	NO		YES		YES	
Month of year FE	NO		YES		YES	
Year FE	NO		YES		YES	
Family controls	NO		NO		YES	
Nursery controls	NO		NO		YES	
Nursery FE	NO		NO		YES	

Specification (3) shows the estimates of α in equation (6). Family and nursery controls can be found in Table A19. Standard errors in parentheses.

Table A19: Supply side: estimation results (part 2 of 2)

	(3)	
	coef	se
Share disadvantaged	0.824	(0.355)
× disadvantaged	-0.526	(0.558)
Income-based × disadvantaged	0.079	(0.197)
Distance to north (km) × disadvantaged	-0.157	(0.056)
Distance to east (km) × disadvantaged	-0.032	(0.071)
Distance to south (km) × disadvantaged	-0.249	(0.069)
Distance to west (km) × disadvantaged	0.085	(0.095)
Hours × disadvantaged	0.094	(0.058)
Cluster 2 × disadvantaged	0.321	(0.205)
Cluster 3 × disadvantaged	-0.438	(0.137)
Cluster 4 × disadvantaged	-0.737	(0.188)
Travel time (min per trip)	-0.022	(0.004)
× disadvantaged	0.002	(0.009)
Sibling	1.498	(0.086)
× disadvantaged	0.275	(0.193)
KUL priority	0.929	(0.098)
× disadvantaged	0.098	(0.148)
Low income	0.006	(0.124)
× disadvantaged	0.179	(0.163)
No Dutch	-0.340	(0.057)
× disadvantaged	0.070	(0.090)
Single parent	-0.109	(0.364)
× disadvantaged	0.228	(0.372)
Low education	0.045	(0.091)
× disadvantaged	0.078	(0.121)
Work or study	0.038	(0.110)

Specification (3) shows the estimates of α in equation (6). Other estimates can be found in Table A18. Standard errors in parentheses.

we account for the characteristics that define them, meaning that nursery-specific differences are minimal. Available capacity now has a more important effect in this improved specification and the rank is still (equally) important for advantaged families. This also holds for disadvantaged families, but they benefit less from the available capacity (although the interaction effect is imprecise) and from a better rank. The 20% target no longer has the opposite effect but results are now insignificant for both groups, suggesting little impact of the current priority policy. The main effects coming from variables that also enter demand (Table A19) are in line with our expectations. For advantaged children, having a sibling increases the hazard by a factor 4 (=exp(1.493)) and being employed by KUL increases it by a factor 2.6 at their institutions. For disadvantaged families, the effects are even slightly larger but not in a statistically significant way.

Table A20: Acceptance probabilities: average marginal effects

	Advantaged				Disadvantaged			
	coef	se	mean	sd	coef	se	mean	sd
Points	0.006	(0.003)	-0.112	(0.759)	0.001	(0.003)	0.299	(0.930)
% capacity	0.133	(0.015)	0.006	(0.252)	0.112	(0.026)	-0.046	(0.248)
Weeks demanded	0.000	(0.000)	121.316	(33.013)	0.000	(0.000)	112.961	(38.753)
Age at start	-0.000	(0.000)	27.135	(23.507)	-0.000	(0.000)	35.129	(28.940)
Days per week	0.017	(0.008)	4.335	(0.896)	-0.031	(0.015)	4.546	(0.827)
Tuesday	-0.022	(0.011)	0.915		0.036	(0.020)	0.942	
Wednesday	-0.010	(0.010)	0.841		0.030	(0.016)	0.862	
Thursday	-0.017	(0.011)	0.895		-0.006	(0.017)	0.924	
Friday	0.016	(0.009)	0.819		0.059	(0.015)	0.873	
Saturday	0.152	(0.083)	0.003		0.011	(0.055)	0.008	
Below target priority policy	0.006	(0.008)	0.258	(0.438)	-0.004	(0.011)	0.255	(0.436)
Same neighborhood	0.014	(0.008)	0.119		0.005	(0.012)	0.111	
Rank 2 instead of 1	-0.126	(0.005)	0.169		-0.072	(0.008)	0.168	
Rank 3 instead of 1	-0.144	(0.004)	0.156		-0.090	(0.008)	0.157	
Rank 4 instead of 1	-0.149	(0.004)	0.140		-0.103	(0.007)	0.150	
Rank 5 instead of 1	-0.150	(0.004)	0.123		-0.094	(0.007)	0.135	
Rank >5 instead of 1	-0.143	(0.005)	0.197		-0.098	(0.008)	0.199	
11m ahead vs 12m ahead	-0.000	(0.000)	0.167		-0.000	(0.000)	0.107	
10m ahead vs 12m ahead	-0.003	(0.002)	0.131		-0.005	(0.002)	0.086	
9m ahead vs 12m ahead	-0.007	(0.003)	0.063		-0.010	(0.003)	0.075	
8m ahead vs 12m ahead	-0.012	(0.004)	0.044		-0.016	(0.005)	0.076	
7m ahead vs 12m ahead	-0.017	(0.005)	0.039		-0.022	(0.007)	0.080	
6m ahead vs 12m ahead	-0.019	(0.007)	0.027		-0.026	(0.009)	0.044	
5m ahead vs 12m ahead	-0.018	(0.009)	0.022		-0.027	(0.010)	0.032	
4m ahead vs 12m ahead	-0.011	(0.011)	0.025		-0.024	(0.012)	0.049	
3m ahead vs 12m ahead	0.003	(0.012)	0.019		-0.015	(0.014)	0.044	
2m ahead vs 12m ahead	0.024	(0.013)	0.025		-0.001	(0.015)	0.059	
1m ahead vs 12m ahead	0.034	(0.013)	0.038		0.005	(0.015)	0.135	
Share disadvantaged	0.103	(0.044)	0.199	(0.121)	0.033	(0.061)	0.250	(0.156)
Sibling	0.280	(0.021)	0.022		0.331	(0.038)	0.015	
Low income	0.001	(0.015)	0.030		0.019	(0.010)	0.822	
No Dutch	-0.039	(0.006)	0.170		-0.030	(0.008)	0.513	
Single parent	-0.013	(0.044)	0.006		0.013	(0.009)	0.356	
Low education	0.006	(0.012)	0.046		0.014	(0.008)	0.216	
KUL priority	0.143	(0.014)	0.135		0.152	(0.027)	0.059	
Work or study	0.005	(0.012)	0.958		-0.007	(0.010)	0.868	
30% priority group	-0.021	(0.014)	0.000		-0.021	(0.015)	0.872	

Average marginal effects for q_{ijr} are calculated as follows: for dummy variables, we take the difference between predicted probability when the dummy is set to 1 and when the dummy is set to 0. For other variables we take the difference when 0.01 is added and divide it by 0.01. For timing variables, we take into account their impact on the hazard through the polynomial that enters Z_{ijr}, as well as the impact of longer exposure through ι (see equation (6)). Standard errors calculated using a bootstrap procedure. Mean is the mean over all $i - j$ combinations that are ranked. Note that the average available capacity is negative for disadvantaged families. This is possible because reported capacity is only a proxy for actual available capacity: nurseries may anticipate absences due to illness or non-enrollment after acceptance.

A.5 Affirmative action in KTTC and decomposing segregation

As explained in the text, segregation increases in KTTC affirmative action policies when preferences of both groups are very different. In this appendix, we first show that demand rather than supply (acceptance) is crucial to explain segregation. To do this, we simulate demand in a world without capacity constraints. This allows us to isolate demand and see its contribution to segregation. Figure A4a plots the segregation curve for both the actual data and the simulations under full capacity. The two overlap almost everywhere, suggesting a strong impact of the demand-side in explaining segregation. We subsequently replace the utility of different channels with their sample average to consider their role in explaining segregation. Figure A4b shows that segregation is mainly explained by two types of geographic concerns. First, there is the home location (captured by travel time to home by car, and the relative difference by bus). Since both groups care a lot about being close to home, residential segregation is strongly reflected in day care segregation. Second, there are heterogeneous preferences over location characteristics of the nursery, capturing other location concerns that could reflect their work location or grandparents.[42]

---

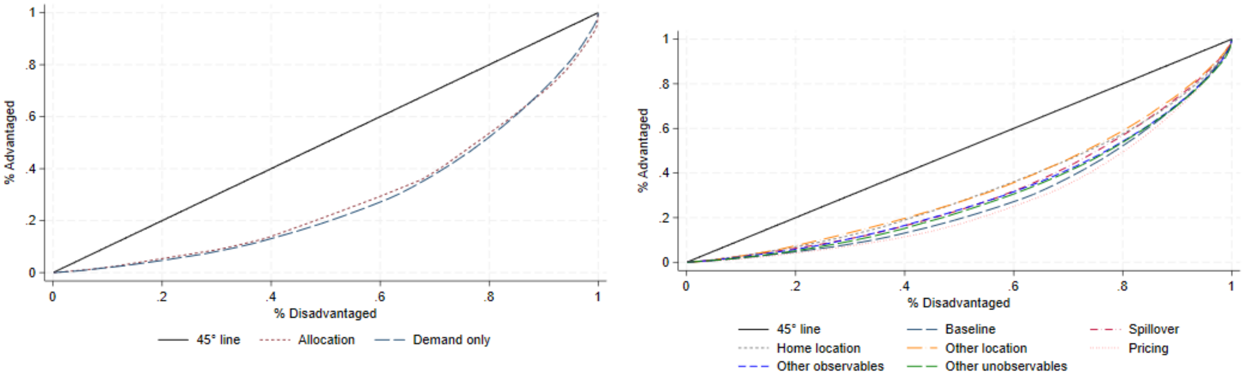
[42] Appendix Table A21 summarizes all results of the simulations.

Table A21: Simulations: segregation

	Baseline	Spillover	Home	Other location	Pricing	Other obs	Other unobs
Match (share of families)							
Unmatched	0.04	0.04	0.12	0.11	0.05	0.06	0.19
Advantaged	0.04	0.04	0.08	0.06	0.05	0.07	0.26
Disadvantaged	0.04	0.05	0.23	0.26	0.06	0.03	0.00
Segregation (0 to 1)							
Dissimilarity	0.33	0.28	0.24	0.25	0.36	0.29	0.30
Gini	0.44	0.38	0.35	0.34	0.47	0.40	0.41

Simulations of first ranked alternative under different manipulations of utility: replacing an individual effect by its average.

Figure A4: Segregation curves in simulations



(a) Comparing data (allocation) to a scenario where everyone can enter their most preferred alternative (demand only).

(b) Demand decompositions*.

*The demand decompositions allow everyone to enter their most preferred alternative when the utility of a particular (set of) variable(s) is replaced by the average in the sample. See Table A21 for the corresponding statistics.