I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17547

Matching Disadvantaged Children to Day Care: Evidence from a Centralized Platform

Olivier De Groote
Minyoung Rho

DECEMBER 2024

DISCUSSION PAPER SERIES

IZA DP No. 17547

Matching Disadvantaged Children to Day Care: Evidence from a Centralized Platform

Olivier De Groote
Toulouse School of Economics, University of Toulouse Capitole and IZA

Minyoung Rho
Oxera Consulting LLP

DECEMBER 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Matching Disadvantaged Children to Day Care: Evidence from a Centralized Platform*

We use data from a platform that centralizes a day care matching process. We estimate parents' preferences and nursery priorities by analyzing parents' rank-ordered lists and nurseries' acceptance decisions. We account for strategic behavior by using a novel estimation approach inspired by the dynamic discrete choice framework. We use the estimates to evaluate centralized matching policies tailored to the day care setting. We compare mechanisms and assess the effects of subsidies, increased capacity, and affirmative action. We find that affirmative action policies are crucial for boosting the participation of disadvantaged children, though they increase segregation due to location-based preferences.

JEL Classification: C61, D82, I24

Keywords: day care, affirmative action, segregation, centralized matching markets, CCP estimation

Corresponding author:
Olivier De Groote
Toulouse School of Economics
1, Esplanade de l'Université
31080 Toulouse Cedex 06
France

E-mail: olivier.de-groote@tse-fr.eu

* This paper benefited from helpful comments from several audiences at ASU, CMU, CREST, Duke, KU Leuven, Northwestern, Tilburg, Tinbergen Institute, TSE, UAB, UAntwerp, UPenn, IIOC 2021, EEA-ESEM 2021, IAAE 2021, EARIE 2021, EALE 2021, the Dynamic Matching and Queuing Workshop Paris 2022, LEER 2023, and the Leuven summer event 2023. 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.

1 Introduction

Many governments promote participation in formal early childhood education and care services with two primary objectives: (1) enhancing female labor force participation and (2) fostering child development. Both outcomes are particularly pronounced among socio-economically disadvantaged groups (Cornelissen et al., 2018; Drange and Havnes, 2019; Felfe and Lalive, 2018; Gathmann and Sass, 2018; Havnes and Mogstad, 2011). For other segments of the population, the effects on child development are more mixed, as they heavily depend on the quality differential between formal day care and the alternative care options available to parents (Baker et al., 2008, 2019; Fort et al., 2020). At the same time, disadvantaged families tend to make less use of day care, giving rise to a reverse selection on gains (Cornelissen et al., 2018). Hermes et al. (2024) find that this difference in take-up is mainly driven by the application barriers disadvantaged families face in decentralized matching markets.

This paper empirically investigates improving the matching process for children to day care, particularly those from disadvantaged socio-economic backgrounds. Using data from a centralized platform with a decentralized matching process, we estimate the priorities of day care institutions and family preferences. These estimates are used to evaluate various centralized matching algorithms. We modify mechanisms, prices, and capacities to assess their effects on participation, welfare, and segregation.

We use a unique dataset from a platform that organizes all formal day care (ages 0-3) in Leuven, Belgium. The matching process is decentralized: families rank nurseries (large-scale day care centers), and nurseries can accept or reject children. Both sides can act at any time and for any chosen start date. Our data and context are particularly well suited to investigate these questions. While there is a large literature on optimal school choice policies, day care imposes additional challenges. First, day care is rarely centralized, and rich data on both supply and demand are often unavailable. Our dataset includes capacities, acceptance decisions, and enrollments for all formal day care options (childminders and nurseries), alongside families' nursery rankings.[1] Second, in many countries, high costs deter families from applying, leading to underrepresentation in the data. In Belgium, however, income-based pricing makes day care broadly affordable, with costs as low as a few euros per day for low-income families. This likely explains why our data represents 88.6% of births in Leuven.[2]

Using these data, we first present evidence from a policy reform that prioritized disadvantaged children in income-based institutions. While the reform increased their enrollment, it also heightened segregation. We do not draw causal conclusions from this as there could be equilibrium effects and confounders over time. Instead, our results are obtained by evaluating alternative policies through counterfactual simulations. For this, we need reliable estimates of preferences.

We estimate household preferences, using rank-ordered lists from 4,922 parents seeking day care seats between mid-2013 and the end of 2016. Our approach accounts for both naive agents, who rank based on true preferences, and strategic agents, who consider acceptance probabilities. For naive agents, rankings are modeled using a rank-ordered logit. For strategic agents, we adapt the dynamic discrete choice framework of Rust (1987), allowing agents to sequentially evaluate their options based on acceptance probabilities inferred from nursery decisions. We proceed in two steps: first, we estimate state transitions (acceptance probabilities); second, we use Conditional Choice Probability (CCP) methods to estimate a flexible utility function, without having to solve the model during estimation (Arcidiacono and Miller, 2011; Hotz and Miller, 1993).

The identification of flexible patterns in preferences and strategic behavior is made possible by our rich data and setting. First, rankings reveal multiple preferred options for most families, making it easy to identify preference heterogeneity through unobserved types. A small number of types is sufficient, as we also observe key factors like the locations of parents and nurseries and sibling attendance. Second, parents

[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.

can apply for a spot at any time, allowing us to exploit significant within-nursery variation in capacity and the proportion of disadvantaged children. The variation in capacity helps identify the share of strategic agents, while the presence of disadvantaged children captures spillover effects, potentially reinforcing segregation (Caetano and Maheshri, 2017). We control for nursery fixed effects specific to each socio-economic group, addressing "correlated effects" in evaluating peer effects (Manski, 1993). Finally, income-based pricing varies by individual, enabling us to estimate price elasticities of demand while controlling for unobserved differences in day care quality using fixed effects.

Our estimates show that location characteristics and having a sibling at a nursery are the primary drivers of utility differences, while spillover effects are minimal. Acceptance probabilities indicate that disadvantaged families face lower chances of acceptance, partly due to applying later. Additionally, we find strategic incentives: ranking an alternative first increases the probability of acceptance by about 10 percentage points. However, only 8% of families rank strategically, with little heterogeneity by socio-economic status.

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 different variants of the Knapsack Top Trading Cycles (KTTC) and Knapsack Deferred Acceptance (KDA) algorithms, which are similar to the classic TTC and DA algorithms but allow parents to specify the days they need care.

These simulations reveal five main findings. First, disadvantaged children benefit from centralized algorithms with transparent priority rules. We implement a rule prioritizing families with a sibling at the nursery, those employed, and those living in the same neighborhood or working for an employer that funds the nursery. We find that tie-breakers matter, with travel time being beneficial for all groups. Centralized mechanisms are particularly advantageous because disadvantaged families often apply too late in a decentralized system and are less likely to be selected by nurseries, even when applying at the same time.

Second, the theoretical trade-offs between KDA and KTTC are empirically relevant. In the KTTC algorithm, more parents are matched with their top choice (55% vs. 49%), leading to higher welfare (an additional 400 EUR per household). However, most families not matched to their favorite alternative would envy the allocation of children with lower priorities, a concern not present in the KDA algorithm.

Third, affirmative action policies create significant distributional effects and are costly to compensate. We implement a dynamic version of soft-bounds quotas for disadvantaged children, prioritizing them in a nursery if they did not meet the quota in the previous period.[3] For example, a 30% quota (slightly above the 28% of disadvantaged children in the sample) increases welfare by 1900 EUR for disadvantaged families, while decreasing it by 1400 EUR for advantaged families. Compared to the same mechanism without affirmative action, it 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% capacity increase could make them indifferent to the baseline scenario without affirmative action, while further improving welfare and matches for disadvantaged families, at a government cost of 1200 EUR per parent applying for a spot. A budget-neutral policy of replacing income-based prices with a common price does not generate enough gain to compensate advantaged families.

Fourth, affirmative action policies increase segregation. Additional simulations show that this is primarily driven by the demand side, where residential segregation, combined with travel costs to day care, leads to segregation in day care placements. Furthermore, socio-economic groups also differ in their preferences for nursery locations that are unrelated to their home location, which further contributes to this segregation. In contrast, spillover effects and the existence of institutions that do not

[3]Building on Ehlers et al. (2014), who find that soft bounds -- flexible limits that regulate school priorities dynamically-- Pareto dominates all other fair assignments while eliciting true preferences.

charge income-based prices do not explain this pattern.

Fifth, alternative policies to improve attendance among disadvantaged children are either costly or ineffective. Increasing the number of income-based seats boosts enrollment for all groups. Although costly, this policy is justified by the value parents derive from it. In contrast, making day care free results in large welfare transfers from the government to advantaged families, with minimal impact on participation for either group. While progressive pricing supports disadvantaged families, further increasing progressivity is ineffective, as prices for disadvantaged families are already low.

Related literature As explained at the start of this paper, it contributes to the day care literature by providing solutions to the reverse selection on gains though alternative matching mechanisms. More generally, this paper contributes to the large literature on equity and efficiency concerns in the allocation of children to educational institutions, as well as the methodological literature on the estimation of preferences in matching markets.

Our first contribution is that we play an empirical counterpart to the mostly theoretical literature which characterizes the resulting allocation from an allocation mechanism with affirmative action policies. While under strong conditions (e.g. aligned preference and indifferent schools), the prioritized group will necessarily be better off under affirmative action (Combe, 2018), this does not hold more generally (Kojima, 2012; Hafalir et al., 2013). Given this theoretical ambiguity, there is surprisingly little empirical evidence on how affirmative action policies in a centralized matching market affect welfare and segregation. The empirical literature has studied the welfare impact of different matching algorithms in schools. Abdulkadiroğlu et al. (2017) show that replacing a decentralized mechanism with a deferred-acceptance algorithm resulted in welfare gains in New York; Calsamiglia et al. (2020) showed that changing the Boston mechanism to the deferred-acceptance mechanism decreased welfare while changes to the top-trading cycle increased welfare in Barcelona. Closest to our work, Oosterbeek et al. (2021) empirically explores the impact of affirmative action policies in the centralized matching market of secondary schools in Amsterdam. They find that affirmative action policies reduce segregation by a modest amount, while it reduces welfare.

Our paper empirically investigates the impact of using a centralized algorithm in the day care matching market and how adding affirmative action policies to it affects welfare and segregation. To do this in the context of day care, we apply a new framework for matching with multidimensional knapsack constraints. This framework was originally developed by Delacrétaz et al. (2023) to optimize refugee matching with differently-sized families and was applied using simulated preferences. Our knapsack constraints come from the fact that parents can choose the days of the week they need care. In the context of day care, we illustrate how affirmative action policies that favor the socio-economically disadvantaged can provide solutions for the reverse selection on gains that are found in the literature (Cornelissen et al., 2018). More generally, we highlight a previously unexplored trade-off of affirmative action policies, as we find that it increases segregation due to the strong heterogeneity in location and preferences of different socio-economic groups. This could be problematic as exposure to diverse socio-economic groups in school can reduce prejudice and discrimination (Rao, 2019).

Our second contribution is to the literature on estimating preferences in centralized markets. The availability of a rank-ordered list (ROL) provides rich information on the preferences of parents or students. However, inference from these ROLs can quickly become computationally burdensome when taking into account that agents are strategic, especially as the market becomes larger. Agarwal and Somaini (2019) provide an overview of methodologies using such rich data from a school choice mechanism to estimate student preferences. Abdulkadiroğlu et al. (2017) and Agarwal and Somaini (2018) use Gibb's sampling in estimating parameters and Calsamiglia et al.

(2020) obtain computational gains in a Boston mechanism by solving the ROL using backward induction. Complementing this literature, we model the rank choices in a general, potentially decentralized, matching context, as a dynamic problem. Our estimator is fast and simple to implement by making use of computational gains from the dynamic discrete choice CCP literature (Hotz and Miller, 1993; Arcidiacono and Miller, 2011). We illustrate this in a particularly complicated context as nurseries can accept or reject a child ad hoc (no algorithm is used), many options are available to parents, and parents have an incentive to behave strategically.

Finally, we contribute to the empirical literature on identifying the sources of segregation in education. Prior research has identified several facets of segregation, including factors such as 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). We contribute by exploiting the unique context of day care. First, this allows us to look at the differences by socio-economic status of preferences of parents of very young children, who were not yet exposed to any form of formal education. Second, as day care slots are more flexible and not subject to an academic calendar, we can exploit high-frequency variation in the share of different groups in a given institution to separately identify different channels that explain segregation: residential segregation, group-specific spillovers, and disagreement on the (unobserved) quality of a nursery.

The rest of the paper is structured as follows. In section 2 we discuss the institutional context and a priority policy in the data which provides descriptive evidence of its effect. Section 3 provides a summary of the data to be used for estimation and how socio-economic groups differ. Section 4 discusses the specification of household preferences we will use throughout the paper and section 5 how to identify its unknown parameters, which are summarized in section 6. Section 7 contains the counterfactual simulations of centralized mechanisms and section 8 concludes.

2 Institutional context

Belgium is a federal country with regions and communities that have their own responsibilities. Day care policies are decentralized and under the full authority of three communities: the (Dutch-speaking) Flemish Community, the French Community and the German-speaking Community. We discuss the institutional context in the Flemish community, covering about 60% of the Belgian population and further referred to as "Flanders". We first discuss its institutional context and compare it to other countries and then proceed to the discussion of the matching system and data in the city of Leuven.[4]

2.1 Formal day care in Flanders and the current priority policy

Children can go to free preschool from the age of two and a half. From the first months after pregnancy until entry into preschool, they have access to day care. All formal day care is regulated by the government agency 'Kind & Gezin'. It can be organized by a childminder, who invites children into their home, but the majority of seats are in large-scale nurseries. Childminders and nurseries can freely enter the market and set their own price (="fixed-price institutions"). Most however charge a price based on the income of parents, set by the government ("income-based institutions"). Both receive subsidies, but income-based institutions receive a higher amount, subtracted by the price paid by parents. The local number of income-based institutions that can be allowed is subject to the government's budget decisions. The income-based prices range between 5.24 EUR and 29.09 EUR per day. In exceptional cases, lower

[4]This summarizes overviews that were found in Gaer et al. (2013); Teppers et al. (2019); Van Lancker and Vandenbroeck (2019).

prices are also possible.[5] On average, households pay 14.16 EUR.[6] Prices of other institutions are around 25 to 35 EUR. Furthermore, all parents receive a tax credit of 45% up to a daily price of 11.20 EUR.[7]

If there are capacity constraints, institutions that charge the income-based prices have to follow priority rules set by the Flemish government. Before the decree of 2012, priority had to be given to single parents, parents with a low level of education, a low level of income, or other important social or pedagogical reasons. The decree of 2012 changed the priority rules and made them more explicit (Gezin, 2019). There is an absolute priority for a day care need due to the working situation of parents. Furthermore, priority should be given to single parents, low income parents, foster children or siblings. On a yearly basis, 20% of children should qualify for at least two of the following priority measures: (1) parents work or follow education, (2) single parent, (3) low income, (4) foster child. Furthermore, additional subsidies are attributed if 30% comes from a particularly "vulnerable" priority group. A child is from a vulnerable group if it has at least two of the following characteristics, of which at least one of the latter three: (1) parents work or follow education, (2) single parent, (3) low income, (4) low parental education, (5) problematic health or care condition.[8]

2.2 Flemish day care in an international context

Day care services in Flanders developed rapidly since the 1970s (Van Lancker and Ghysels, 2012). In the 1990s about 20% of children below the age of three attended

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

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

[7]Both types of day care can be organized by public or private actors. Public nurseries charge the income-based price. Childminder can adhere to a public childminder service and then also charge the income-based price. Private nurseries or childminders can choose between a fixed price or the income-based price. While public and private institutions were financed somewhat differently, a new decree of 2012 (in place since 2014), gradually removed these differences, only keeping a difference between institutions that set an income-based price and those that do not.

[8]The income threshold is indexed and was at the level of 28,757 EUR/year in 2019. Parents are considered to have a low level of education if none of them have a high school degree.

day care. This increased to 63% around 2010, far above the Barcelona target of 33% (set by the European Council).

Formal day care in Belgium is cheap, with parents spending about 5% of their income, compared to an OECD average of 12% (OECD, 2011). In terms of participation, the OECD ranks Belgium 7th with an enrollment rate in formal care of 56% among the 0-2 year olds (the OECD average is 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.[9]

2.3 The matching platform of the city of Leuven

Our data come from an online platform in the city of Leuven, the fourth largest city of Flanders, counting a little more than 100,000 inhabitants. It consists of a historical city center, populated by about 30,000 inhabitants, as well as suburbs surrounding the city center. It is a university town, providing higher education to 55,000 students, of which 35,000 live in Leuven during weekdays in the academic year (not included in the population numbers). The availability of academic institutions also results in a more highly educated and richer population than other cities in Flanders. Nevertheless, as in other cities, it is also characterized by large diversity in socio-economic status. In particular, we see 20% of births in poor families, compared to an average of 13% in Flanders.[10]

In 2011, a collaboration between the city of Leuven, local childminders, nurseries and welfare organizations gave rise to a platform called "Loket Kinderopvang". The most visible part of it was a website (kinderopvangleuven.be) through which demand and supply for day care services in Leuven could be matched. The collaboration was considered a success. Despite the lack of a legal ground to force them, all childminders

[9] 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.

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

and nurseries in Leuven decided to organize their services through the platform.

Two principles were maintained in the design of the platform: autonomy for day care providers, and freedom of choice for parents. It is therefore much more flexible than platforms that have been used to organize school choice in many countries. Families interested in a spot in nurseries rank up to five alternatives and specify their requested days and times in the week, as well as the requested start and end date. Families interested in a childminder also register on line and receive access to a list of available spots, 9 months before their requested starting date. If a spot is available, they can contact the childminder directly. The ranked nursery receives a message when a family ranked them. They see all the information on the platform and can respond with a proposal of a spot or reject the child. Some nurseries have specific policies about when they handle requests for different periods, others will respond immediately. The parents can accept or reject the offer. If a child is rejected, parents can choose to remain on a waiting list in case a spot becomes available or replace it with another option.

3 Data and the difference between socio-economic groups

This section serves three purposes. First, it provides an overview of the differences in the data between advantaged and disadvantaged children. Second, it shows motivational evidence on the impact of priority policies on matchings and segregation. Third, it shows the relevance of looking at ranking data on nurseries to learn about formal day care demand, which we need for counterfactual simulations.

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 show statistics for the entire sample of children attending day care between June 2013 and December 2016 (matching sample), as well as for a restricted sample that we will use in the rest of the paper. The latter differs from the former as it only includes families for which we reliably observe ranking data (ranking sample). It excludes in particular families that only considered childminders. A complete overview of the data cleaning can be found in the Appendix.

Table 1 describes the characteristics of all formal day care options in June 2013 (see Appendix Table A13 for December 2016). For all day care options, it shows the

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 A13 for December 2016).
*We drop one nursery in the rest of the analysis because of the small number of children attending it.

capacity, price system and type (childminder or nursery). There is a large variation in the sizes of day care options, and most are income-based. When using capacity weights, one can interpret the numbers at the level of the available spots. 88.5% of the available spots are income-based places and 87.7% of the spots are in nurseries.[11] For nurseries, we have access to more characteristics. First, we observe a dummy for "KUL". KUL denotes the city's largest university (KU Leuven), the university hospital (UZ Leuven) and a large spin-off (IMEC). They collaborate on providing day care in which priority is given to their employees. As these are large employers, they also capture a large number of spots: 22%. We also have location data. This allows us to calculate travel time from home. However, parents might value the nursery location for other reasons (job location, grandparents, neighborhood quality...). Therefore, we calculate distances to five points in the city of Leuven, given by the centroids of (merged) neighborhoods in the city center, and the four cardinal directions and treat them as nursery characteristics. Finally, we observe the operating hours and calculate the average during the week.

Table 2 summarizes the characteristics of families. First, it is important to note

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.

that the matching and ranking samples show largely similar summary statistics, which

[11] 78% of spots with childminders were income-based, for nurseries this was 91%.

is important as our demand estimation will only use the ranking sample. Second, it shows the share of families in the 20% and 30% groups targeted by the government's priority policy. The 20% group captures 28% of the families in Leuven, while the 30% is only slightly smaller (24%). We also take into account more specific family characteristics. 25% is defined as low income, 25% does not speak Dutch (the official language in Flanders) at home, 10% of children grow up in a single-parent home and 9% do not have parents with at least a high school degree. The "work or study" indicator shows that a large majority of parents are either working, studying or looking for a job. Finally, we observe an average daily income-based price of 14 EUR, with substantial variation (SD = 8 EUR). However, we only observe this for a subset of the sample. The data is missing for two reasons. First, nurseries choose if they want to use the platform for their bookkeeping. Only if they do, we observe the prices they charge. Second, we cannot observe it when children never go to an income-based institution. As these reasons are likely non-random, it will be important to take this into account in the estimation procedure.

Since the 20% and 30% are similar in size, we will focus attention on priority policies that target the 20% group and call them "disadvantaged" in the rest of the paper. We now discuss the differences between this group and other ("advantaged") families.

Table 3 shows the same characteristics as Table 2, but only for the matching sample and broken down by socio-economic group. By construction, the two groups are very different. The main defining characteristic turns out to be the low-income category. Only 2.4% of advantaged families are in this category, while it is 86% among disadvantaged families. This is also reflected in the income-based prices they face (17 EUR vs. 7 EUR). Also other characteristics that are reflective of the disadvantaged status show large differences. While the language at home is not used to categorize them, a majority of disadvantaged families does not speak Dutch at home, while it is only 14% for advantaged families. There are virtually no single parent families among

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

the advantaged, and only a few families without a high school degree, while it is the case for respectively 1 out of 3 and 1 out of 5 disadvantaged families. The share of families in an active study or work situation is high in both, but only approaching 100% for the advantaged.

Table 4 shows that the two socio-economic groups have different matching pat-

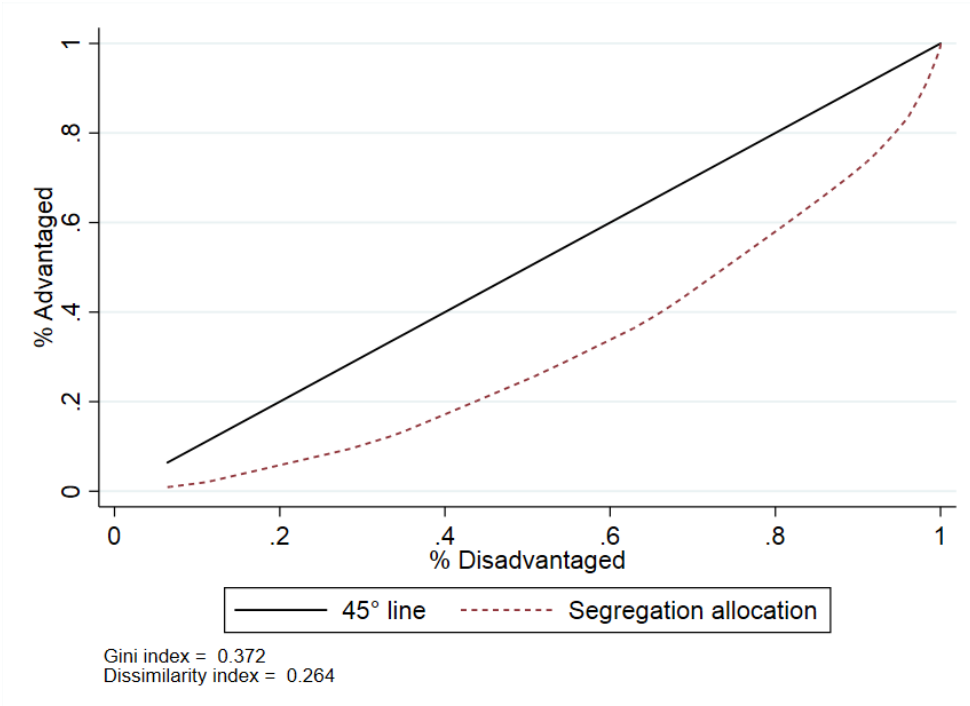
Table 4: Match statistics

	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.

terns. While the number of total matches is similar, disadvantaged children are a

Figure 1: Segregation in day care



bit less likely to be matched with a nursery. When matched to a nursery, they are substantially less likely to match with their highest-ranked alternative. They often end up in different places too. We show this by drawing a segregation curve (for the ranking sample): Figure 1. To draw this, we proceed as follows: we calculate the % of advantaged children in each nursery and order them from low to high. We then calculate the cumulative number of disadvantaged children and put this on the x-axis. Similarly, we calculate the cumulative number of advantaged children and put this on the y-axis. Perfect integration would imply a 45°-line. Instead, we see for example that day cares capturing 40% of disadvantaged children, only have a bit less than 20% of the advantaged children, indicating some level of segregation. This is also confirmed by the larger-than-0 Gini-coefficient (0.37) and dissimilarity index (0.26).[12]

[12]The Gini coefficient calculates the size between the Lorenz-curve and the 45°-line, relative to the total surface under the 45°-line. The dissimilarity index takes the sum of the absolute difference

3.2 Impact of the 2014 policy change

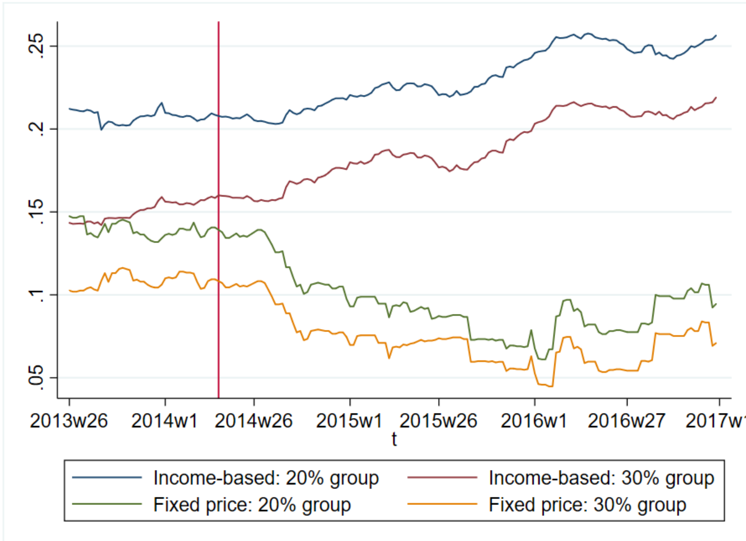
The 2014 policy change (see section 2.1) aimed to increase the participation of disadvantaged children by requiring income-based institutions to prioritize them. Figure 2 shows that the share of disadvantaged groups increases after the policy change for both types of priority groups. Income-based institutions 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.[13] 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.

Figure 3 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.

in the relative share of each group within each day care and divides it by 2.

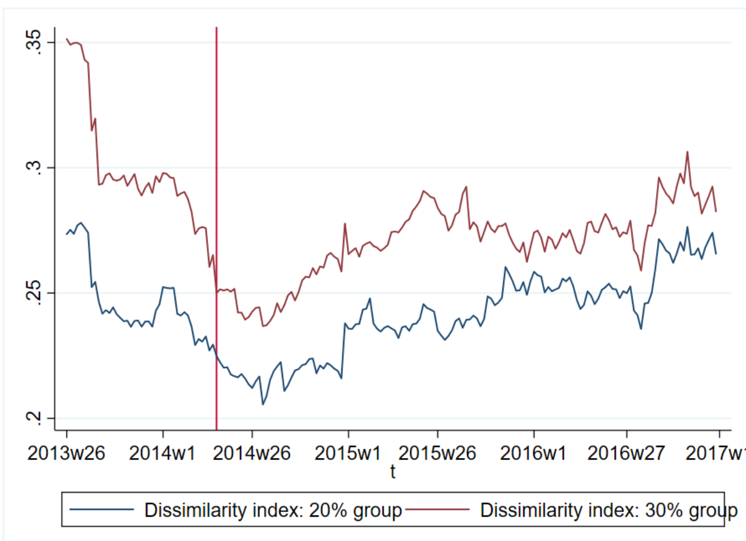
[13] We also noticed that the share of children of different groups finding a match (not shown here) remains constant over time, casting doubt on the policy having an impact on the acceptance behavior of nurseries.

Figure 2: 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 3: 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.

3.3 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. Figure 4 shows the timing of asking for a spot for advantaged and disadvantaged families. Advantaged families ask for a spot earlier before the starting date, while for disadvantaged families we observe a strong peak right before their requested starting date. Similarly, we observe a high peak in demand in the first two months after conception for advantaged families, while demand from disadvantaged families is more gradually spread over time and many more disadvantaged families ask for a spot after birth.

Figure 4: Timing of demand

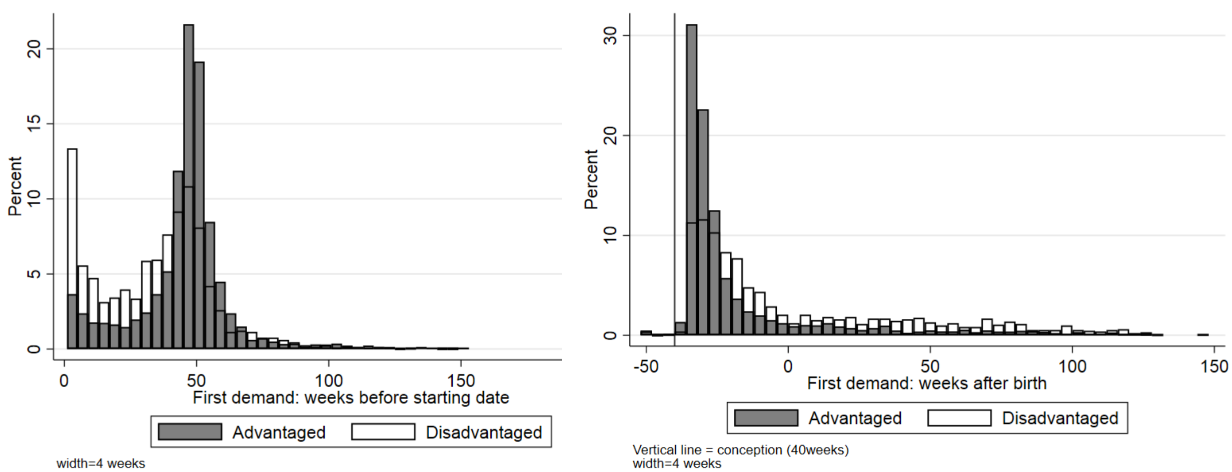
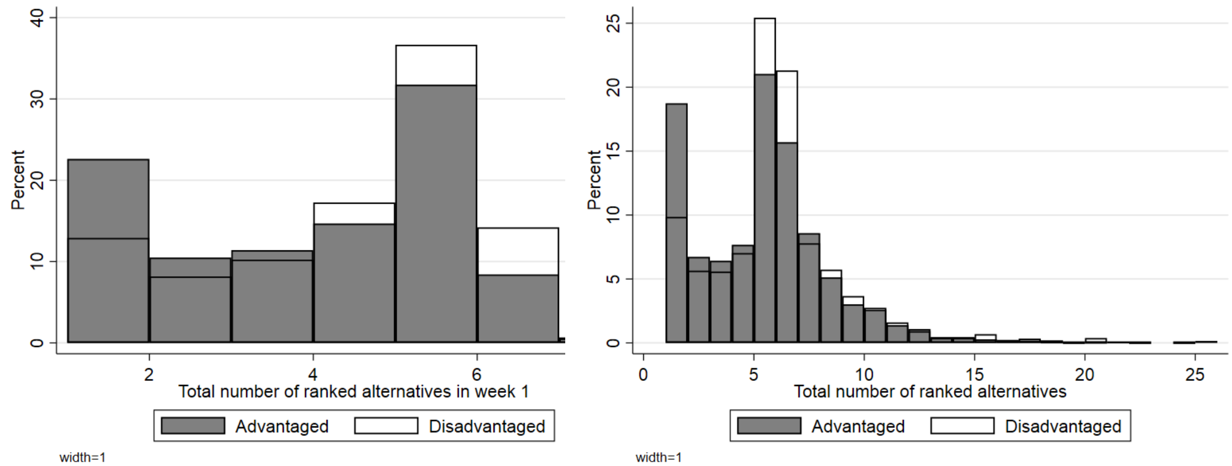


Figure 5 denotes the number of ranked alternatives by both groups. Most families rank the maximum of five but there is also a large group of families who only rank one alternative, especially among advantaged families. Note also that several families rank more than five alternatives. This can be done by removing another alternative from the list, e.g. if there is no spot available to allocate now.

In Table 5 we compare the characteristics of the most preferred alternatives to those of the fifth rank and the average available to them the moment they apply. We include the previously introduced characteristics but also look at characteristics that

Figure 5: Number of nurseries ranked



vary for a given nursery, depending on who applies and when. Note that differences can be explained by preferences, but also by acceptance probabilities if families are ordering strategically. Both groups are very likely to put an income-based nursery first, but advantaged families are more willing to also consider fixed-price institutions. This is consistent with the smaller price difference they face. Both groups also prefer lower travel time, but disadvantaged families are more sensitive to bus times and the fact that the nursery is located in the city center, suggesting they rely more on public transportation.[14] Finally, disadvantaged families are attracted by nurseries where there is currently a high number of disadvantaged children, while this does not explain the rankings of advantaged families.

To finish this section, we show how well the results of the ranking system correspond with actual attendance behavior. As explained in the previous section, there is no centralization of the matching mechanism so there is no guarantee that the highest-ranked alternative that is accepted will indeed be the one where the child goes. Moreover, when the deadline gets closer, also non-ranked nurseries receive info about families looking for a spot and can offer them one. Nevertheless, we see that

[14]We use the location data of the family and the nursery and calculate travel times using google maps. Furthermore, we add five minutes to account for the time to get ready to leave and to avoid 0s in calculating relative differences.

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.

the large majority of families indeed comply with the system. Table 6 shows that after acceptance of 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 6: 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

4 Household preferences

Our goal is to explore how to improve the matching of children, particularly those of low socio-economic status. To do so, we apply various centralized matching algorithms and evaluate welfare differences. In order to apply and evaluate them, we need household preferences. This section proposes the utility function we will use, the next section will discuss how to estimate its parameters using ROLs in a (potentially) strategic context.

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-nursery-specific fixed effect, capturing heterogeneity in the valuation of observed and unobserved characteristics for 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 paid by i if they would attend j. Choosing to opt 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 will be captured by the observed and unobserved characteristics included in W_{ij}. To rationalize all ranking data using a computationally efficient estimation, we will additionally add rank-specific random shocks. As they are uncorrelated over ranks, it is crucial that u_{ij} is sufficiently rich to capture (persistent) heterogeneity in preferences such that it can generate reliable welfare estimates and substitution patterns. The same argument is used in the dynamic matching context (Agarwal et al., 2021) as they also apply a similar estimation strategy. We improve upon this by adding unobserved heterogeneity inside of u_{ij}, as is also the case in random coefficient models to circumvent the restrictive substitution patterns generated by the shock distribution (Berry et al., 1995). We do this by allowing each agent to belong to one of two unobserved preference types (A or B) and let them flexibly interact with nursery characteristics, along with a vector of observed family characteristics (low income, single parent, low education, no Dutch, work or study).

First, we capture a general preference for a nursery in Leuven compared to families' individual outside options. We allow the utility of the inside good to differ by the vector of observed family characteristics, the unobserved preference type, and the time remaining until the desired starting date. We also control for the income-based price families would pay. This is because the outside option also contains income-based places in other municipalities so it is important to control for the fact that there is heterogeneity in prices they would face elsewhere.

Second, we estimate the importance of travel time. 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. Furthermore, we include a preference for the relative time difference if one opts for public transportation.

Third, we capture preferences over nursery characteristics that differ by household. As parents can apply and enroll at any time of the year, we have several variables that contain within-nursery variation.[15] This allows us to estimate their effect through β_s, while controlling for group-specific perceived (unobserved) quality of nurseries through δ_{js}. In particular, we consider that parents take into account the number of children and the proportion of disadvantaged children when applying. By controlling for group-specific fixed effects, we can abstract from the issue of correlated unobservables in identifying peer effects (Manski, 1993) and interpret the parameter as capturing spillovers, i.e. how much parents value the social composition. Moreover, note that nursery prices vary among households due to the income-based pricing structure of most (and not all) places. This allows us to control for unobserved quality of nurseries through fixed effects while identifying the effect of prices on choices. Since utility is normalized with respect to prices, it identifies the scale of the utility function. We also account for more specific reasons for parents to prefer a nursery: having a sibling in the same facility and being employed by KUL for nurseries that belong to the KUL system.

Finally, we capture preference heterogeneity over invariant observed and unobserved nursery characteristics. To capture preference heterogeneity over unobserved characteristics, we first categorize the 45 nurseries into four clusters using a k-means clustering algorithm.[16] We interact a constant, dummy variables for three of the

[15] Note that these nursery characteristics are time-varying as they depend on the time of asking for a spot, or the requested enrollment time. To keep the notation simple, we treat it here as constant for a given individual i such that we do not need an additional (time) subscript. However, in estimation, we relax this as parents could add alternatives to their list at a later time too.

[16] 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,

four clusters and observable characteristics (distances to five locations and operating hours) with the observed family characteristics and unobserved preference type. We will also conduct a second-stage regression of the nursery fixed effects on the same characteristics to gain more insights in what is valued by families.

5 Estimation

This section describes our empirical strategy to uncover household preferences. Note that our main simulations will not require the explicit modeling of the supply side. Nevertheless, we will require data from the nurseries' acceptance decisions to be able to allow for agents' strategic ranking behavior.

In section 5.1, we first explore the estimation strategy when (naive) agents truthfully report their preferences. Section 5.2 instead considers strategic agents. Section 5.3 explains how to derive a likelihood function of a mixture of both, without having to solve the optimal decision of the strategic agents. Finally, section 5.4 discusses identification.

5.1 Truth-telling agents

If family i is naive, i.e. reporting their preferences truthfully, we assume they rank options according to the choice-specific utility u_{ij} denoted by (1), plus an individual shock ϵ_{ijr} for each choice j at each slot r. We also assume they pay a cost to rank c_s. Note that, in contrast to a standard exploded logit, we consider shocks to differ not just by family i and alternative j, but also by rank r. While it does not matter for estimation in the naive case, it will provide substantial computational benefits when considering that agents are strategic. Moreover, it allows us to capture changes with rank, e.g. because of trembling-hand mistakes or new information following from the fact that agents can add alternatives to their ROL at a later time. Unobserved

after subtracting their mean and dividing by their standard deviation.

heterogeneity that is persistent over ranks is contained in u_{ij}.

To rationalize their choice in each rank, it is useful to define the agents' perceived rank-specific value function:

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

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 of the next slot. ϵ_{ir} is a vector of all ϵ_{ijr}. $\mathcal{J}_i$ is the set of options in the market at the time 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 agents are truth-telling, a likelihood function can be composed from these probabilities and gives rise to a standard exploded logit model.[17]

5.2 Strategic agents

A strategic family considers the probability of acceptance and uses it to form their value functions. We first discuss the acceptance probability and then discuss the rank-specific choice probability.

[17]Note however that if all agents are naive, the ranking cost c_s is not identified as it enters the conditional value function of each choice alternative.

5.2.1 Acceptance probability

Denote by q_{ijr} the probability of acceptance of i by j in rank r. 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:

$$q_{ijr} = 1 - \exp(-\lambda_{ijr}\iota_i^p). \tag{3}$$

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.[18] More details on the specification, as well as the estimation results, are in Appendix A.3. Importantly, we allow for rich heterogeneity by socio-economic group and include variables that are excluded from the utility function such as the available capacity, nursery-specific scores based on applicants' characteristics and the order of the nursery in the ROL.

5.2.2 Rank-specific choice probability

A strategic family is assumed to know the probabilities of acceptance when they choose their ROL. In each ranking slot, they choose a nursery, taking into account

[18]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.

the vector of all acceptance probabilities (q_i), resulting in value functions:

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

with the conditional value function now defined as

$$v_{jr}^{\mathrm{s,Str}}\left(X_{ir}^{\tau}, q_i\right) = q_{ijr} u_j^s(W_{ij}) - c_s + \left(1 - q_{ijr}\right) E_{\epsilon_{r+1}}\left[V_{r+1}^{\mathrm{s,Str}}\left(X_{ir+1}, q_i, \epsilon_{ir+1}\right)\right] \tag{4}$$

leading to choice probabilities

$$p_{jr}^{\mathrm{s,Str}}(X_{ir}^{\tau}, q_i) = \frac{\exp(v_{jr}^{\mathrm{s,Str}}(X_{ir}^{\tau}, q_i))/\sigma_s)}{\sum_{j' \in \mathcal{J}_i \setminus H_{ir}} \exp(v_{j'r}^{\mathrm{s,Str}}(X_{ir}^{\tau}, q_i)/\sigma_s)}. \tag{5}$$

Notice that the problem of a strategic family is similar to a dynamic problem of a forward-looking agent as the optimal strategy for lower ranks enters the decision for the current rank.

5.3 Likelihood function

Valuations are parameterized with group-specific parameters (δ_s and β_s), allowing us to estimate the model separately for advantaged and disadvantaged families. Since we cannot observe the family's preference type τ, we estimate its distribution. The probability that a family of observed type s belongs to an unobserved type τ is denoted π_s^{τ}, with $\sum_{\tau=\{A,B\}} \pi_s^{\tau} = 1$. Similarly, we need to estimate the proportion of strategic and naive types. The proportion of the population who is filling out their ROL strategically is denoted as $\lambda_s(S_i, \tau)$, and the proportion of the population who is filling out the ROL naively is $1 - \lambda_s(S_i, \tau)$, with S_i capturing 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 data, i.e., series of $\{d_i, X_i, q_i\} = \{(d_{i1,}X_{i1}, q_{i1}), \ldots, (d_{iR_i}, X_{iR_i}, q_{iR_i})\}$ where d_i is the portfolio of choice, X_i^τ the state if i is of unobserved type τ and q_i the acceptance probabilities is given by the following equation.[19]

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

where

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

where $p_{jr}^{\text{s,Naive}}(X_{ir}^\tau, q_i)$ and $p_{jr}^{\text{s,Str}}(X_{ir}^\tau)$ are the CCPs given by (2) and (5) and $I(d_{ir} = j)$ indicates that the observed choice for i at slot r is the nursery j.

Interpreting the ROL data Parents have the possibility to add alternatives at a later time and they can change the rank. We also do not always know if a family prefers the outside option over non-ranked alternatives. We interpret the data by assuming that later-ranked alternatives are always ranked below previously ranked alternatives. We also only assume i ranks $j = 0$ before non-ranked alternatives when we observe that i left the platform without an allocation. For others, we do not make any assumption on how non-ranked alternatives are compared to the outside option. Note also that when options are added at a later time, characteristics of j could have changed. With some abuse of notation, we take that into account by letting W_{ij} vary with the time of adding the nursery to the ROL. When forming expectations, we assume that agents expect them to stay constant in their expected future values.

Missing price data In several cases, we do not know the income-based prices parents (would) pay. While many are missing for random reasons, we also do not

[19] We treat the acceptance probabilities here as data as we assume they are exogenous to the decision process.

observe them when the child never attends an income-based nursery. As this is likely non-random, we integrate this out of the likelihood function. This is similar to how Arcidiacono et al. (2024) integrate out missing college majors. In practice this gives us 12 types for each socio-economic group s when prices are missing: two behavioral types (naive or strategic) $\times$ two preference types (A or B) $\times$ 3 price types (5, 16 or 23 EUR/day). To facilitate the separate identification of price types, we identify their population probability outside the model in a first step by estimating a Heckman selection model (Heckman, 1979).[20]

CCP estimation and finite dependence To avoid solving the model for strategic agents by backward induction, we proceed to CCP estimation. With the EV1 assumption, Hotz and Miller (1993) and Arcidiacono and Miller (2011) show that we can write

$$E_{\epsilon_{r+1}}\left[V^{\text{s,Str}}_{r+1}\left(X^{\tau}_{ir+1}, q_i, \epsilon_{ir+1}\right)\right] = v^{\text{s,Str}}_{j^*,r+1}(X^{\tau}_{ir}, q_i) - \sigma_s \ln\left[p^{\text{s,Str}}_{j^*,r+1}\left(X^{\tau}_{i,r+1}, q_i\right)\right] \tag{6}$$

with j^* any option in the choice set.[21] 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 $v^{\text{s,Str}}_{j^*,r+1}(X^{\tau}_{ir}, q_i) = -c_s$, hence removing any further dependence on

[20]We use an ordered probit for the three price categories with a selection equation for missing data that takes into account observed characteristics of families, as well as the probability of ranking an alternative first that does not report price information. This probability is obtained from a conditional logit that includes travel time while controlling for nursery fixed effects and other non-price characteristics. I.e. it exploits the heterogeneity in travel time to nurseries that report prices for identification.

[21]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.

the future. We then obtain the following conditional value functions:

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

Aside from rescaling flow utility by q_{ijr}, the estimator is now a standard dynamic discrete choice model with unobserved types and finite dependence so it can be estimated using the adaptation of an EM loop by Arcidiacono and Miller (2011). To facilitate the estimation, we first estimate a specification with a scale normalized to one, but with an estimated price coefficient. Dividing all parameters by the price coefficient then yields the proposed utility specification in euros. Note that we require a reduced form estimate of the CCP of the outside option in (6). We obtain predicted values in a first stage by estimating a logit that depends in a flexible way on the characteristics of families and nurseries, which is then further updated in the adapted EM procedure.

5.4 Identification

Dynamic discrete choice models are identified after normalizing the utility of one option, setting the discount factor and specifying the distribution of the error terms (Magnac and Thesmar, 2002). Apart from common ranking costs, the utility of the outside option is set to 0. The discount factor is set to 1 for strategic agents and to 0 for naive agents. A mixture is identified by exploiting variables that influence the acceptance probability (entering for strategic agents only), but not the utility function. As shown in Appendix A.3, we include several exclusion restrictions that are empirically relevant, in particular, the available capacity at the time of demand and the rank of the nursery in the families' list. We also include a score for each applicant based on a nursery-set formula for family characteristics. While this does not bind nurseries in any way, it influences the order of applications in the list they select from. Note that our model specification differs slightly from that of a standard dynamic discrete choice model as we are multiplying the flow utility by the acceptance probability (see (4)). Note however that this does not introduce identification problems as state transitions are identified outside the model using the acceptance data.

We also allow for unobserved heterogeneity types by making use of the full ROL. By assuming the error term ϵ_{ijr} is iid over ranks, persistent unobserved differences between individuals i can only be captured by the preference types. For example, an individual who ranks only nearby institutions highly, is more likely to care about proximity.

Apart from unobserved preference types, we also need to integrate out unobserved prices for a large part of the sample. Some income-based institutions choose not to report them and for families allocated to fixed-price institutions, we do not know their income-based price either.[22] Note that this is the opposite of the usual selection problem in wage data (where wages below the reservation wage are not observed (Heckman, 1979)), but therefore the intuition for identification is similar as in a Heckman selection model as we can exploit preference shifters, in particular the travel time to nurseries that do (not) report income-based price information.

We also identify the cost of choosing an alternative in each rank. This is identified by the variation in the size of the ROL. However, since only differences in utility are identified, this cost drops out for naive families. For strategic families, it enters the conditional value functions through its expected impact on the lower ranks. Ranking an option with a low acceptance probability changes the need for, and therefore the expected value of, adding lower ranked alternatives to the list. Intuitively, if ranking costs are high, strategic agents will opt for nurseries in which they have a higher acceptance probability so they do not need to add lower-ranked alternatives. Estimating this cost is also useful to counterbalance any (artificial) incentive for

[22]Note that the price paid in fixed-price institutions is not a concern as they are absorbed by fixed effects.

families to be rejected as being rejected also allows them to choose between a new draw of the shocks.

6 Estimation results

We first discuss the main estimates we need for simulations: household preferences. Then, we shortly summarize our results of acceptance probabilities, which are an input needed in the estimation of preferences of strategic parents.

6.1 Household preferences

Table 7 and appendix Tables A14 and A15 show the results of the demand estimation. We first estimate the model with a price coefficient and normalized scale, but then use it to divide all parameters so we get the money-metric utility specification we proposed in (1). This way, estimates can be interpreted as a (daily) willingness to pay. Standard errors are obtained using a bootstrap procedure.[23]

Travel time is crucial for families' preferences. The least sensitive unobserved types experience a baseline effect of 2.6 EUR/day decrease in utility for the advantaged family and a 1.5 decrease for the disadvantaged. As parents would have to do this trip twice a day, it translates into a reasonable estimate of the opportunity cost of time of respectively 1.3 and 0.8 EUR/minute. Note, however, that other unobserved types have a substantially larger opportunity cost of time, while the impact of observable characteristics is relatively small.[24] It is likely that this type captures people who value travel time for other reasons than the opportunity cost of time. For example, parents might value that their children are close to home, or get to know their neighbors. Moreover, the average travel times in the data are low (a one-way

[23] We use 50 bootstrap samples and re-estimate the entire model (supply, demand and auxiliary regressions). Note that we do not run the EM algorithm again. Instead, we keep the weights fixed and thereby assume no uncertainty in the estimation of types.

[24] We omit the interactions with single parent dummy for the advantaged families as there are only twelve in that part of the sample.

Table 7: 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 A14 and Table A15. 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.

trip is around 15 minutes for the average nursery), and do not take into account traffic jams or alternative ways of transportation (e.g. walk or bike). As these travel times are expected to be strongly correlated, it does not entail a problem for the purpose of this paper but does imply that they are overestimating the opportunity cost of time. We do account for the difference in travel time by public transportation. Indeed, for parent-nursery pairs that are not well connected by it, there is a substantial decline in utility. If it takes twice as long by public transportation, instead of the same, it decreases utility by 5 to 6 EUR per day.

We also account for two endogenously evolving characteristics, but effect sizes are small. An increase in the number of children at the nursery by 10 would increase utility by 0.1 to 0.3 EUR per day. Surprisingly, a higher share of disadvantaged children is valued by both groups, but also these effects are small. A (substantial) increase of 10pp in the share, would increase the utility of advantaged families by 1.6 EUR/day and of disadvantaged by 1.9 EUR/day.

Next, we find a strong effect of having a sibling in the facility, increasing utility by 34 EUR for the advantaged and 19 EUR for the disadvantaged. These numbers are reasonable when compared to the travel cost estimates, which is what we would expect as having a sibling in the place, reduces the number of trips by half. A similar magnitude is given to the KUL priority. Note that both variables are also strong predictors of being accepted (see later), but since we model strategic behavior, the effect sizes here are unrelated to that.

Finally, we capture heterogeneity in outside options by interacting the constant with the income-based price of families and the time left until the start. Parents who pay high income-based prices (which also applies outside of Leuven), value their outside options less as each daily increase by 1 euro, makes outside alternatives 2 (advantaged) to 1.5 (disadvantaged) EUR less attractive. As this is more than 1, we expect the estimate also simply captures a higher taste for day care by people with high incomes. Families who apply early are also less likely to go for their outside option, suggesting that people with a high need make sure to apply on time. People who are 10 weeks earlier, value day care more at a rate of around 2 to 3 EUR per day.

Table 7 also reports the estimates of parameters that do not enter u_{ij}. We report the scale of the shocks. Note that a standard deviation of the shocks is given by $\sigma\frac{\pi}{\sqrt{6}}$, which leads to 16EUR for the advantaged, and 10 EUR for the disadvantaged. These parameters test the richness of our model. They are relatively small, given that they are used to rationalize the entire data, after accounting for observables and unobserved types. It shows that we capture the relevant choice margins through u_{ij}. Finally, we find a high cost for adding alternatives to the list, showing that families are unwilling to exploit all their options before resorting to their outside option.

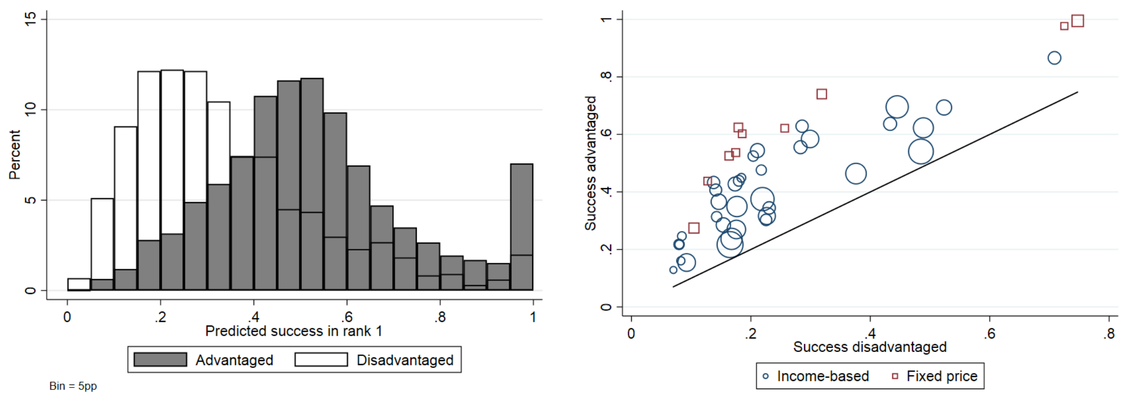
Finally, we report the share of unobserved preference types and of strategic families. The unobserved types are almost equally divided, while we find very few strategic types in both groups (7% in the advantaged group and 10% in the disadvantaged group). Appendix Table A16 reports the average marginal effects for strategic types. For advantaged children, only the weeks left until the starting date are statistically significant, decreasing the probability of being strategic. This also holds for disadvantaged families, while also showing a negative impact of speaking Dutch at home and a positive effect of being a single parent and belonging to type B.

Other tables show a large number of estimates to capture heterogeneous preferences over fixed nursery characteristics. Appendix Table A14 shows that operating hours are valued by all, bet less so for parents who do not speak Dutch. For location preferences, unobserved types are particularly important, but also some observables matter. Non-Dutch-speaking parents prefer places in the city center, while low-income households stay away from the north and south. In Appendix Table A15, we find large differences between how unobserved types value different clusters, but the effects of observables are usually smaller. We do find heterogeneity over clusters with No Dutch parents preferring clusters 2 and 3, and advantaged yet low-income parents preferring cluster 3.

6.2 Acceptance probabilities

The estimates of acceptance probabilities and a discussion can be found in Appendix section A.3 and Tables A19 and A20. To understand them better, we do three things. First, we plot acceptance probabilities of a first-ranked alternative (q_{ij1}) for each group in the left panel of Figure 6. The two distributions differ substantially. Disadvantaged

Figure 6: Acceptance probabilities and socio-economic groups



Notes: plots of the acceptance probabilities in rank 1: q_{ij1} (see equation (3)). 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.

children have a lot of mass on probabilities around 20% and very few on probabilities higher than 60%. Low probabilities are uncommon for the advantaged who have a mode close to 50% and a large number of advantaged families face success rates close to 100%.

Second, we plot the heterogeneous impact of the nursery in the right panel of Figure 6. This shows the probability of acceptance in each nursery for two representative (average) families who only differ because of their disadvantaged status. We also take the time-varying nursery characteristics (available capacity and share of disadvantaged children) to be the same. Differences are driven by the interaction effects with observed and unobserved nursery characteristics, shown in Appendix Table A20. We see a large variation in probabilities among different nurseries, but all have higher acceptance probabilities for advantaged families. It is more pronounced for the fixed-price institutions, but also income-based nurseries (who are expected to prioritize disadvantaged children) do not drop below the 45°-line.

Finally, we calculate marginal effects, i.e. how q_{ijr} chances with (small) changes in characteristics: Appendix Table A21. A few variables with big effects stand out. Having a sibling increases the acceptance probability by around 30 percentage points for both. However, it only applies to a small number of applications. The KUL priority increases it by 14 to 15 percentage points and applies to a much larger number of applications, particularly among the advantaged (13%). The available capacity and the rank are crucial variables. An increase in the share of available seats by 10% points increases the acceptance probability by 1.4% points for the advantaged and by 1.0% points for the disadvantaged.[25] Ranking an alternative first, entails a bonus of 13 to 15 % points for the advantaged, and 7 to 10% points for the disadvantaged. As we saw in the descriptive analysis, disadvantaged families are far more likely to apply for a spot in the final months before the spot is needed. This has a negative impact on the probability of being accepted, but effect sizes are relatively small. Note, however, that this is after controlling for available capacity at the time of demand, which is indeed much lower for the disadvantaged as they apply later. We also see many more families who do not speak Dutch at home among the disadvantaged, which has a negative impact on the acceptance probability.

[25] Note that the average available capacity is negative for the disadvantaged. This is possible as the reported capacity is only a proxy for the actual capacity available in nurseries. For example, they can expect children not to come every day due to illness, or not enroll after accepting.

7 Simulations of centralized mechanisms

In this section, we discuss simulations of the model. We first discuss the inputs and outputs for the mechanisms we evaluate. Subsequently, we analyze the centralization of the mechanism using different algorithms. Finally, we apply additional policies within this framework, including affirmative action and price or capacity changes.

7.1 Inputs and outputs

Timing We split each year into 6 enrollment periods that are defined by the different school vacations in each year as this corresponds to the enrollment periods of preschool. We then assign each family to the period that is closest to what we observe for them in the data.[26] We run matching algorithms sequentially over 20 periods between the end of 2013 and the end of 2016 by simulating utilities for parents, and priorities for nurseries. First, we subtract the capacity taken by the children that were omitted from the analysis from the reported capacity. After each round, we adjust capacities, but also endogenous characteristics (enrollment numbers and the share of disadvantaged children) and let families take into account their value of the last period to calculate utilities. Finally, we only report the results for children in the last 12 periods to allow for some burn-in during the first periods.

ROL We form complete ROLs by simulating from the estimated utility distribution. We abstract from strategic ordering and simulate only truth-revealing mechanisms. We form lists based on $u_j^s(W_{ij})$ and an additive shock.[27]

[26] We take the actual starting dates for children attending nurseries and the requested ones for those without a match. The reason to take actual starting dates is to capture that, in practice, parents might consider a different starting date when they did not find a match at their requested date and this is something we abstract from in the analysis. We also assume parents do not switch nurseries later (in the data, this happens for only 10.5%).

[27] To avoid simulation noise to drive the results, we report the average statistics over 25 simulations but results are very similar with only a single draw per family. We do not consider new draws of shocks after rejections, consistent with the idea of ranking all nurseries at a single moment in time, close to the moment of enrollment.

Priorities While we compare different sets of priorities below, our main results use a common and transparent rule, giving points for having a sibling at the institution (4), working or studying (1), and living in the same neighborhood (1).[28] For KUL-owned nurseries, we replace the location preference with a preference for working at KUL. This still leaves many ties, and we use travel time as a tiebreaker.

Outcomes We want mechanisms to leave few people unmatched, maximize the number of children that are assigned to a high-ranked place and avoid having families envy others for taking a spot they feel belongs to them (defined by having a higher priority than a family matched in the same period), and avoid having segregation. We further quantify the desirability of the system by calculating welfare effects. 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.[29]

7.2 Centralizing the matching process

Explaining data and the role of priorities In a first set of simulations, we simulate deferred acceptance algorithms (DA) under different priorities. A DA is often considered desirable because it avoids envy and truth-telling is a dominant strategy. We account for the differences in days demanded by applying the "Knapsack" DA (KDA) of Delacrétaz et al. (2023), which rejects families who cannot be accom-

[28]These numbers were inspired by the results of Table A18 which regresses self-reported priorities of nurseries on characteristics, as explained in Appendix section A.3.

[29]The daily cost is calculated based on a file 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 cut) income-based price paid for income-based spots and 3.4EUR/day + tax cut 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.

modated by the nursery.[30] The properties of KDA are similar to DA, but will be discussed more in the next subsection.

Table 8 summarizes the main results of this simulation. Before showing the simulation results of centralized mechanisms, the first columns show the predictions from the decentralized mechanism.[31] The second column attempts to replicate these results, but through a centralized (KDA) mechanism that uses acceptance probabilities q_{ijr} as priorities. Note these are not expected to be policy-invariant and will not be used as such, but we find them useful to understand the patterns we see in the data. Results for the number of matches and segregation are indeed comparable between the first two columns, making it an interesting baseline to compare other simulations to. However, centralization does seem to improve the match to the top ranked alternative.[32]

In the second column, we strip the acceptance probabilities from frictions created by the time of application and the bonus given for a high rank. While the number of matches stays quite constant for both groups, we do see a substantial decline in the number of parents receiving their highest-ranked alternative, which is also translated into substantial welfare losses of around 2600 EUR on average for advantaged families and 700 for disadvantaged. On the other hand, segregation is reduced. We will also see in further simulations that there is often a trade-off between increasing welfare and decreasing segregation. Note that the priority for highly ranked offers helped a lot to maximize the number of parents getting such a highly ranked offer. These results suggest that allowing priorities to depend on ranks is beneficial, at least if we value welfare over segregation. However, this inevitably creates incentives to be

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

[31] 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 some of the parents who matched with a nursery to an outside option instead, based on the numbers of Table 6.

[32] A welfare comparison is difficult due to a different role played by the shocks in both scenarios. However, we can compare utility, net of shocks. We find that they are very close, differing only 180 EUR on average, with the decentralized outcome being 130 EUR less favorable for the advantaged and 270 EUR less favorable for the disadvantaged.

Table 8: Simulations: explaining data and alternative priorities

Priorities	Decentralized	KDA Data	KDA No frictions	KDA S(imple)	KDA S + time left	KDA S + travel time
Match (share of families)						
Unmatched	0.32	0.36	0.36	0.35	0.35	0.34
Advantaged	0.27	0.33	0.32	0.34	0.32	0.33
Disadvantaged	0.42	0.43	0.44	0.38	0.42	0.37
Matched with first-ranked	0.57	0.70	0.41	0.43	0.49	0.49
Advantaged	0.61	0.74	0.43	0.44	0.51	0.50
Disadvantaged	0.45	0.59	0.36	0.39	0.42	0.45
With envy		0.12	0.22	0.00	0.02	0.01
Advantaged		0.10	0.24	0.00	0.02	0.01
Disadvantaged		0.17	0.18	0.00	0.02	0.02
Welfare (Δ/capita in 1000 EUR)						
Parental benefit			-2.04	-1.58	-0.88	-0.72
Advantaged			-2.58	-2.20	-1.16	-1.21
Disadvantaged			-0.68	-0.03	-0.17	0.53
Government cost			-0.09	-0.02	-0.07	0.08
Segregation (0 to 1)						
Dissimilarity	0.25	0.27	0.22	0.28	0.26	0.32
Gini	0.35	0.39	0.31	0.39	0.36	0.43

Decentralized denotes simulations from estimated choices, acceptance probabilities and compliance statistics. Other simulations use KDA mechanisms with different priorities. Data-driven priorities are given by the acceptance probabilities. "No frictions" remove the effect of the time of application and the bonus given for a high rank. S(imple) rule 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. S+time left uses the time between the application and requested starting date as a tie-breaker. S+travel time uses travel time between the nursery and the home location instead. 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. These incentives are ignored in this exercise. Moreover, it is not a desirable feature of a good mechanism. However, we can use these insights when developing alternative priority rules.

In the last three columns, we use the priority rule we mentioned before (using sibling, work status, and either location or KUL priority), but using different tiebreakers. With a random tiebreaker, the advantaged would still be worse off but to a smaller extent. The disadvantaged now face similar welfare as in the baseline case. Segregation goes up substantially. Using the time left between the application and the starting date as a tiebreaker leads to substantially better welfare, but only for the advantaged. Using travel time as a tiebreaker, we obtain similar results for the advantaged, but a substantial improvement for the disadvantaged who are now 500 EUR better off than in the baseline case. Segregation increases again.

We conclude that we can develop a transparent matching mechanism to obtain similar results as in the data. Increased welfare is established by making sure that priorities are in line with preferences. The distaste for travel time is particularly useful for that. On the downside: we see that increases in welfare go hand in hand with increases in segregation.

Alternative matching mechanisms Previous simulations used a KDA matching mechanism. We now seek to further improve outcomes by comparing three mechanisms proposed in Delacrétaz et al. (2023): KTTC, KDA and TKDA. They are variants of the well-known Top-Trading Cycles (TTC) and DA mechanisms but adjust for the fact that parents ask for specific days of the week.

KTTC, similar to TTC, is desirable as it is Pareto efficient. However, it is known to suffer from justified envy. DA avoids this but is not Pareto efficient. Similarly, KDA and TKDA satisfy the weaker concept of "weak envy-free" matchings so the number of families that envy others is expected to be small. The reason to also consider TKDA is because, in contrast to a standard DA and to KTTC, KDA does not guarantee truth-telling, which we are assuming in these simulations. While the estimates suggest this might be a minor concern in our sample, we also report the results of this more sophisticated algorithm that restores truthfulness at the cost of efficiency and simplicity.

Table 9 reports the results for different mechanisms, using the proposed priority rule with travel time as a tiebreaker. We obtain the following main findings. First, KDA and TKDA yield almost identical results. Second, KTTC increases the number of families matched with their first-ranked alternative and provides an improvement in welfare for both types of families (440 EUR for the advantaged and 260 EUR for the disadvantaged). Moreover, we escape the trade-off between welfare and reduced segregation as segregation is now substantially less and similar to the data again. However, implementing a KTTC does come at a cost as many families would envy families who have a lower priority, while envy was virtually 0 in the (T)KDA.

7.3 Affirmative action and segregation

Dynamic soft-bound policies In Table 10 we consider five different affirmative action policies, still under the KTTC algorithm with our priority rule and travel time as a tiebreaker. We implement a dynamic variation of a soft-bounds policy. A soft-bounds policy prioritizes the disadvantaged up until a quota is reached. However, it allows other children to take the remaining available seats even if the quota is not satisfied (Ehlers et al., 2014). We slightly modify it to adapt it to our sequential procedure. We run our KTTC algorithm, but we change the priorities of income-based institutions to always be higher for disadvantaged children in nurseries that did not yet reach the quota last period. Note that 28% of children are disadvantaged. We consider a mild quota of 20% and a strong quota of 30%.

First, a mild quota has small overall welfare effects, but favors the disadvantaged (almost 1000 EUR on average), and hurts the larger group of advantaged (700 EUR). A strong quota amplifies this further. As can be expected, envy (calculated at the

Table 9: Simulations: mechanisms

	KDA	TKDA	KTTC
Match (share of families)			
Unmatched	0.34	0.35	0.32
Advantaged	0.33	0.34	0.31
Disadvantaged	0.37	0.37	0.34
Matched with first-ranked	0.49	0.48	0.55
Advantaged	0.50	0.49	0.56
Disadvantaged	0.45	0.44	0.52
With envy	0.01	0.00	0.42
Advantaged	0.01	0.00	0.42
Disadvantaged	0.02	0.00	0.44
Welfare (Δ/capita in 1000 EUR)			
Parental benefit		-0.09	0.39
Advantaged		-0.11	0.44
Disadvantaged		-0.06	0.26
Government cost		-0.08	0.19
Segregation (0 to 1)			
Dissimilarity	0.32	0.32	0.28
Gini	0.43	0.43	0.39

Simulations with different matching mechanisms, using the same 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. 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. We report averages over 25 simulations for the last 12 matching periods.

Table 10: Simulations: affirmative action

	Base	Priority			
		20 %	30 %	30% + Cap 9%	30% + Price flat
Match (share of families)					
Unmatched	0.32	0.32	0.32	0.26	0.32
Advantaged	0.31	0.34	0.36	0.30	0.35
Disadvantaged	0.34	0.28	0.22	0.17	0.26
Matched with first-ranked	0.55	0.53	0.52	0.54	0.52
Advantaged	0.56	0.53	0.50	0.52	0.50
Disadvantaged	0.52	0.54	0.55	0.58	0.58
With envy	0.42	0.44	0.46	0.42	0.45
Advantaged	0.42	0.47	0.51	0.48	0.51
Disadvantaged	0.44	0.38	0.33	0.29	0.30
Welfare (Δ/capita in 1000 EUR)					
Parental benefit		-0.23	-0.47	0.76	-0.31
Advantaged		-0.71	-1.40	0.06	-0.59
Disadvantaged		0.98	1.87	2.52	0.40
Government cost		-0.02	0.00	1.24	0.00
Segregation (0 to 1)					
Dissimilarity	0.28	0.28	0.30	0.32	0.32
Gini	0.39	0.38	0.41	0.43	0.43

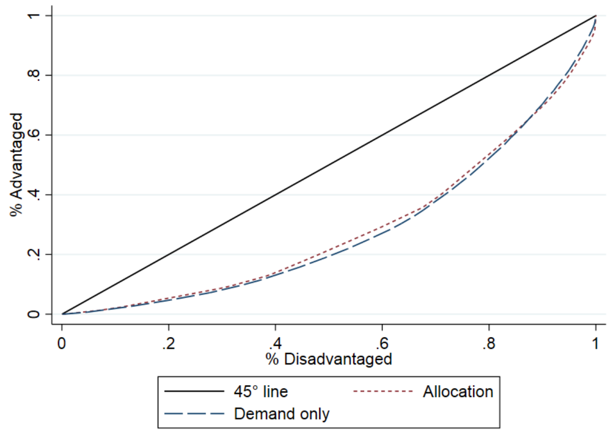
Simulations of KTTC mechanisms with prority rule: points for having a sibling at the institution (4), working or studying (1), and living in the same neighborhood (1). For KUL-owned nurseries, we replace the location preference by a preference for working at KUL. Travel time between the nursery and the home location is used as a tie-breaker. Pricing, capacity and affirmative action policies considered here only affect income-based nurseries. Effects of price and capacity policies without affirmative action can be found in Table 11. Flat prices are set the after tax cut average of 10.5 EUR per day. 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 and no affirmative action). We report averages over 25 simulations for the last 12 matching periods.

original priorities) increases among the advantaged, but decreases a lot among the disadvantaged. To compensate the disadvantaged for their loss, we consider two compensation policies (for their isolated effect, see Table 11). A simultaneous capacity increase of 9% manages to bring the utility impact for the advantaged close to 0 but comes at a public cost of 1200 EUR per person on the platform. Setting a flat price is another way to compensate the advantaged as they usually pay higher prices. While a 10.5 after-tax price manages to keep the budget constant, it does not suffice to compensate the advantaged for their loss. Interestingly, the disadvantaged still prefer this scenario over the baseline case, suggesting that the capacity available to them, rather than prices, is stopping them from entering day care.

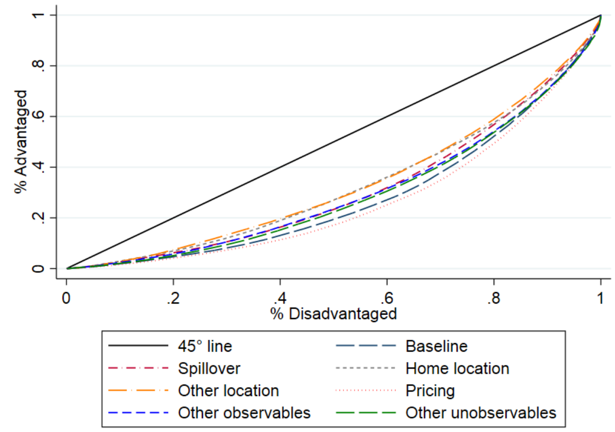
Finally, we again see the trade-off in different policy goals. While a strong affirmative action policy favors the disadvantaged a lot in terms of utility, it also increases segregation.

Decomposing segregation To better understand why segregation increases in affirmative action policies, 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 7a 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 7b 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

Figure 7: 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 A17 for the corresponding statistics.

or grandparents.[33]

7.4 Adjusting prices and capacities

Table 11 simulates alternative policies for income-based institutions. In all simulations, we use the KTTC algorithm with the simple rule and a tie-breaker by travel time. Free day car would change very little to the matches but would entail a large increase in the public cost of day care (2500 EUR per person applying for day care) that would particularly benefit the advantaged families as they are paying higher prices. If everyone would pay the average after-tax-cut daily price of 10.5 EUR, participation of disadvantaged children would substantially go down. However, making the system more progressive by setting the price to 0 for those whose price is below 10 (before taxes), paid for by the group with prices above 20, would only lead to small gains for disadvantaged families, despite the fact that it would affect more than

[33] Appendix Table A17 summarizes all results of the simulations.

75% of them. Capacity increases hold more promise. For example, the 9% increase in capacity (as above, but now without affirmative action) at the start of the simulations would increase participation of the advantaged families by 5 percentage points and of disadvantaged families by 6 percentage points.[34] It would cost 1200 EUR per child applying for a spot, but parental welfare increases by the same amount, meaning that any positive externality of day care would lead to a positive cost-benefit analysis of increased capacity.

8 Conclusion

Disadvantaged children often do not enter day care, despite the documented gains for them. We estimate preferences using a novel method on unique data of parents' application decisions and day care acceptance decisions and use the estimates to illustrate the impact of centralized mechanisms that would improve welfare and increase attendance by disadvantaged children.

Our results have clear policy implications: (1) a centralized matching algorithm can be implemented and is beneficial for disadvantaged families, especially when priorities take into account important preference shifters such as travel time. (2) Affirmative action policies are straightforward to implement in a centralized mechanism and are effective at increasing the participation of disadvantaged children. However, they also increase segregation and are disliked by advantaged families. (3) As prices are already low, there is little to be gained from alternative subsidization policies to encourage participation, but subsidized capacity increases have large effects on participation. Their public cost is justified when compared to the benefits parents derive from it.

Further research could look into how to handle cases in which parents are flexible in the timing of their demand, both within a week, as well as over different months of

[34] Remember that we are always reporting the last 12 out of 20 periods.

Table 11: Simulations: alternative prices and capacity

	Base	Price free	Price flat	Price progressive	Extra capacity
Match (share of families)					
Unmatched	0.32	0.32	0.33	0.32	0.26
Advantaged	0.31	0.31	0.30	0.32	0.26
Disadvantaged	0.34	0.35	0.39	0.33	0.28
Matched with first-ranked	0.55	0.53	0.56	0.55	0.57
Advantaged	0.56	0.54	0.56	0.56	0.58
Disadvantaged	0.52	0.51	0.54	0.51	0.55
With envy	0.42	0.47	0.41	0.43	0.38
Advantaged	0.42	0.47	0.41	0.42	0.38
Disadvantaged	0.44	0.48	0.40	0.45	0.40
Welfare (Δ/capita in 1000 EUR)					
Parental benefit		2.54	0.33	-0.03	1.20
Advantaged		3.19	0.93	-0.22	1.41
Disadvantaged		0.93	-1.20	0.47	0.67
Government cost		2.89	0.10	0.05	1.22
Segregation (0 to 1)					
Dissimilarity	0.28	0.28	0.29	0.28	0.28
Gini	0.39	0.38	0.40	0.39	0.39

Simulations of KTTC mechanisms with prority 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. Flat prices are set the after tax cut average of 10.5 EUR per day. Progressive pricing 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. 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.

enrollment. It would also be interesting to investigate the impact of better matching policies on the labor market outcomes of parents and educational outcomes of children to micro-found the estimated parental welfare gains, and to quantify additional gains through 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.

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. (2019). Revealed Preference Analysis of School Choice Models. Technical Report w26568, National Bureau of Economic Research, Cambridge, MA.

Arcidiacono, P., Aucejo, E., Maurel, A., and Ransom, T. (2024). College attrition and the dynamics of information revelation. *Journal of Political Economy*, (forthcoming).

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.

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).

Combe, J. (2018). Are Affirmative Action Policies so Impossible? *Revue d'économie politique*, 128(6):1123.

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*.

Gezin, K. . (2019). Toegankelijkheid en diversiteit: Voorrangsgroepen en opnamebeleid. Technical report.

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.

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.

Kojima, F. (2012). School choice: Impossibilities for affirmative action. *Games and Economic Behavior*, 75(2):685–693.

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.

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 A12 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 A12: 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 Appendix Tables and Figures

Table A13: 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.

Table A14: 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 7 and A15. 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: 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 7 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 A16: 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 A17: 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.

A.3 Acceptance probabilities

We observe the decisions made by nurseries when they receive an application and use this to estimate acceptance probabilities q_{ijr}, given by equation (3), 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. 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:

A.3.1 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 A18 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.

Table A18: 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.

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.

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.3.2 Estimates

Table A19 and A20 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 A21. 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.[35]

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 spec-

[35]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 (3)).

Table A19: 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 (3). Family and nursery controls can be found in Table A20. Standard errors in parentheses.

Table A20: 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 (3). Other estimates can be found in Table A19. Standard errors in parentheses.

ification, 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 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 A20) 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 A21: 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 (3)). Standard errors calculated using a bootstrap procedure. Mean is the mean over all $i-j$ combinations that are ranked.