Efficiency and fair access in kindergarten allocation policy design

André Veski   Péter Biró   Kaire Põder   Triin Lauri

March 28, 2016

Abstract

We study kindergarten allocations practices in two municipalities of Estonia. We describe their current allocation practices and provide recommendations for improvement. The recommendations are based on child-optimal stable matching, with priorities based on siblings and distance. We evaluate seven policy designs to understand efficiency and fairness trade-offs, based on 2015 admission data from one of the municipalities. Regrettably the data is limited due to being collected under then applicable decentralised allocation procedure. Based on the available data we estimate a ranking function and use that for a counter-factual policy comparison and sensitivity analysis. We find that different ways of considering the same priorities can have significant aggregate effect on the allocation. Additionally we survey a dozen special features that can create significant challenges (both theoretical and practical) in the redesign the allocation mechanism in Estonian kindergartens, and probably also elsewhere in Europe.

1 Introduction

Families have become a much-debated issue in all developed countries and they are focal points in debates of "new risks" and the much needed "new policies" for Western welfare state. The questions of who should care for children, how much and for how long, are at the centre of conflicts around values that shape not only policies and struggles around policies, but also individual and family choices (Saraceno, 2011).

In Eastern-Europe the Soviet legacy has paved the way for publicly provided care to dominate, but in many countries, including our cases, there is a shortage of places for care in early childhood for children aged 18 months to three years. This shortage of places has forced municipalities, who are the main providers, to set priorities for the allocation of these places. One of the most commonly used priorities in Estonian municipalities is the date of application, and in some cases catchment areas are used to ensure proximity. However, as there are municipalities where the shortage of places is acute, even using these priorities does not guarantee a reasonable allocation in terms of equity or efficiency, as families will get places in kindergartens which they do not prefer, if they get them at all. There are several reasons for this, among them: 1) the lack of information on the preferences families;

2) improperly managed priorities; 3) the complicated nature of manually finding a "good" allocation.

We take the approach of a matching mechanism design to propose a "good" way to allocate children to kindergartens. This has been successful in many similar allocation situations such as school choice (Abdulkadiroğlu and Sönmez, 2003; Abdulkadiroğlu et al., 2005a) and job-assignments (Roth, 2008). While mechanism design provides methods for allocation when given welfare criteria like maximising family preferences and stability of the match with some selection priorities, it does not say how these priorities should be applied. The general policy goals for allocation in school choice are siblings in the same school and proximity to a school, and in our kindergarten allocation case studies we see the same principles. Proximity is applied in some cases of school choice as a walk-zone (Shi, 2015) and in other cases as a continuous priority (West et al., 2004), leading to the question of why one or the other method should rather be preferred. Our objective is to shed some light on the trade-offs of efficiency and fairness involved in selecting an design method for the policy goal, by simulating allocations under several alternatives. We view efficiency as the allocation of children to more preferred kindergartens, siblings in the same kindergarten and the minimisation of the travelling distance to kindergartens. Fairness we view as a probability of a child being assigned to the family's most preferred kindergarten.

In section 2 we review the practices of two Estonian municipalities. In section 3 we review common policy designs and propose seven policies for evaluation. In section 4 we review the initial results of the policies using data from one of the municipalities. However, current practices limit the data collection to three preferences, which would not be the case with new policies. From the current data we estimate the full preferences of families and in section 5 we evaluate the alternative policy designs in various counter-factual scenarios. This section is partially inspired by the methodology used in (Shi, 2015), however we also evaluate policy implementations' sensitivity to counter-factual preference profiles. In the appendix we describe the preference estimation procedure.

We summarise in section 6. In section 7 we describe additional aspects of mechanism design that we identified from the two municipalities, but did not explore in the paper.

2 Descriptions of current practices

The chosen municipalities represent two typical cases in Estonia, where there has been a lot of new housing development and more families need to find ways to get a place in a kindergarten. The two municipalities have tackled the problem in slightly different ways. Children in Rae are mostly assigned based on a catchment areas, where most areas have one kindergarten, so parents do not have much choice. This has started to change as additional kindergartens have been built and efforts have been made to give more choice to parents over kindergartens. This choice has been at the heart of kindergarten allocation in Harku municipality, where parents have been asked for their preferences for many years. The similarity in both of these municipalities is that the actual allocation has been up to the heads of the kindergartens, who aim to find an optimal allocation through negotiation. Although both municipalities have policies for how children should be allocated, it is up to the heads how these policies are actually implemented, and they try to make a good judgement with some additional information from parents.

2.1 Harku municipality

Harku has seven kindergartens. In addition there is a childcare that also belongs to the municipality, but charges an additional fee and may not follow the educational curriculum that is compulsory in kindergartens. The demand is for about 200-250 places a year. About 200 kindergarten places are vacated each year and there are about 40 additional places in the childcare, so there is not a significant shortage of places, given that all the available places are acceptable to families.

In the application procedure families can submit up-to three ordered preferences. The application date and the home address are also collected. The application date is important in the allocation as families with an earlier application date have higher priority in kindergartens, so families usually make the application as early as possible, usually a few weeks after child-birth. The application data are typically unchanged until the actual allocation happens. The address can be important, as some heads of kindergartens consider it when assigning places. A qualifying condition is that parents have to be registered as living in Harku, as this determines where local taxes are collected.

The number of vacant places is settled by January and the allocation can start. Offers of a place are made to families by the heads of kindergartens if their kindergarten is the first choice of the family. If there are more families than places then priority is given to applications with earlier dates, although proximity or siblings might occasionally also be a factor. If a place is accepted, the child is assigned to the kindergarten, otherwise the place is offered to the subsequent family on the waiting list.

The head of the kindergarten communicates with other kindergartens to find a place for children who are still unassigned. The communication is usually with the kindergarten listed as second or third choice, though these might already be full. If so, the result may be that families with an earlier application date are rejected from their second choice because the children who were assigned had listed that kindergarten as their first preference, no matter their application date. This is a well-known property of the Immediate-Acceptance mechanism (e.g. Abdulkadiroğlu and Sönmez, 2003) and the procedure used in Harku is very similar to this.

There is often another round of allocations in September. There are usually 2-4 places kept spare to cater to children with special needs, who are considered to take the place of multiple children. This information is not known during the initial allocation, but is observed in the first weeks by the teachers. If there still are no children with special needs, then these extra places are also allocated.

2.2 Rae municipality

Parents in Rae can only apply to kindergartens in their catchment area. Currently there are nine kindergartens in four catchment areas. The demand for kindergarten places is for about 300 places each year. There are several new kindergartens and there is no aggregate shortage of places, although the free places are not evenly distributed. Places are mostly available in one area, but there is a shortage in a second area. The experience of the officials is that families in the second area are not interested in placing their children in a kindergarten in the first area.

In the application procedure, families by default submit an application to all the kindergartens in their catchment area. The application data are public. Like in Harku the application date is important in the allocation.

Allocation begins on April 1 and is centralised. In each catchment area children are allocated in order of their application date, with some rare exceptions being made for municipality officials, civil servants, large families and similar cases. Families with an earlier date are sent an offer for a place and they have two weeks to reply. If they accept the offer, the child is assigned to the kindergarten. If they reject, the place is offered to the subsequent family on the waiting list. Where there are multiple kindergartens in the same catchment area proximity can be an additional priority, but this is not considered consistently. This procedure is similar to the serial dictatorship mechanism ordered by application date.

3 Matching mechanism design

In mechanism design the here relevant allocation is usually modelled as a two-sided matching market, in this case between: 1) families and 2) kindergartens. Both have some ordering over the other side. Families have a preference ordering over kindergartens and they seek to get allocated to their most preferred kindergarten. Kindergartens can be seen as having a priority ordering over children. Priorities become important if there are fewer places available in a particular kindergarten than there are families who would like to have their children attend that kindergarten. In such a case kindergartens then accept children who are higher on the priority list, which in practice might mean children who live closer or who have a sibling in the kindergarten. Kindergartens do not seek to have higher priority children, which is different from some applications of two-sided markets. In college admissions for example (Gale and Shapley, 1962), both students and colleges seek to get more preferred matches, so they might be strategic about it.

There are two prominent "good" mechanisms for solving matching problems, and these are the Deferred-Acceptance (DA) and the Top-Trading Cycles (TTC) mechanisms. The DA mechanism guarantees that no preferences and priorities, or policies, are violated, and there is no child who could get a place in a more preferred kindergarten by priority, so there are no blocking pairs. A matching with no blocking pairs is stable. A blocking pair can also be seen as having justified envy if there is a family that would prefer a kindergarten that either has free places or has accepted a child with a lower priority.

The TTC is another mechanism where the families have an incentive to reveal their true preferences. The difference is that it allows priority violations, and although more children may be allocated to a more preferred kindergarten, some children might end up with a worse match (Abdulkadiroğlu and Sönmez, 2003). Because of this we do not consider TTC viable in this situation and most real-world applications are based on DA (Pathak and Sönmez, 2013).

While there are potentially many stable allocations (Knuth, 1997), the child-proposing DA mechanism that is usually executed results in the best possible preference for the families. The child-proposing DA algorithm works as follows:

1. All children are tentatively assigned to their first preference. If kindergartens have more children assigned than places, they reject children with lower priority.

2. All rejected children are tentatively assigned to their second preference. Again if kindergartens have more children assigned than places, they reject children with lower priority. Note that the kindergarten may reject children tentatively assigned in the previous round, if they have a lower priority.

...

k. In general rejected children are tentatively assigned to their next preference. If a kindergarten has more children assigned than places then children with lower priority are rejected. The process continues until all children are assigned a place, or all preferences have been explored.

To be safe for families to reveal their true preferences with Deferred-Acceptance algorithms, the number of preferences revealed cannot be limited. Currently the parents can list at most three kindergartens. This makes the parents to select their three choices strategically, if they would find more places acceptable. The theoretical properties and disadvantages of such restrictions were studied in Haeringer and Klijn (2009). These findings were backed by laboratory experiments (Calsamiglia et al., 2010) and also with evidences from practical application across the world (Pathak and Sönmez, 2013), where the tendency is to increase these bounds or abolish them. Therefore we do not recommend limiting the number of applications.

Gale and Shapley (1962) showed that the child-proposing DA guarantees the best possible preference for all children, while not violating priorities. The result is child-optimal stable matching. Additionally families have an incentive always to state their true preferences, which might not be the case for kindergartens, if they were strategic agents that were seeking higher priority children (e.g. Roth, 1982).

3.1 Main issues and mechanism design improvements

From these results we can highlight improvements to the current allocation procedure in Harku and Rae. These recommendations usually form the backbone of any matching practice.

Data collection. The biggest problem in Harku starts with data collection. As the parents can get a higher priority if they apply sooner, they tend to apply soon after the birth of the child. During the following three years the family's preferences might change, but that is not usually reflected in the application data. So after the allocation is made, kindergartens are faced with a significant number of cancellations from families and need to reconsider some children. This creates justified envy in the allocation and confusion during the procedure. If a child was initially rejected from a kindergarten due to shortage of places and was allocated somewhere else, they should now be reconsidered. Often this might not happen and the offer is made to the subsequent unallocated child, creating justified envy for the first family, whose child received a place in a kindergarten that was lower on their preference list. Second, if a previously rejected child is offered a place and they accept, another empty place is created in the system and this leads to another child needing to be reconsidered.

The solution for this is fortunately straightforward. Data should be collected close to the actual allocation date. This would mean that the application date cannot be a priority criteria any more, but this is not problematic since in many cases it was the birth date anyway. An alternative approach would be to use siblings and distance as objective criteria for prioritising children as is popular in many places around the world (e.g. Abdulkadiroğlu et al., 2006).

Setting the priorities. In many cases the priorities followed by the kindergartens were not clear. Children with siblings were usually considered with a higher priority, but not always. The regulation requires siblings to be considered, but only after the application date. Since most children have a distinct application date, this would not give a significant advantage to children with siblings. On some other occasions distance was considered an important criterion, but not in every case.

The solution is to create clearly defined priority metrics and have a centralised allocation system that makes sure that the criteria are always followed.

Matching mechanism. The current mechanism followed in Harku could create justified envy. Using the Immediate-Acceptance mechanism creates blocking pairs where some children may prefer to attend a kindergarten other than their current assignment and may also have a higher priority in that kindergarten, if they live closer than some children who were accepted. In this case the family can be considered to be unfairly treated and these situations should be avoided

The solution is to use a stable allocation mechanism. In the literature these are the Deferred-Acceptance mechanisms Abdulkadiroğlu and Sönmez (2003). In this paper we limit our attention to the child-proposing (student-proposing in school choice) Deferred-Acceptance mechanism, because this is the only mechanism that finds a stable allocation that is the best possible for the children and is strategy-proof for the families.

Limited preferences. In Harku the number of preferences is limited to three. This can create some problems for parents trying to decide which kindergartens to list. If there are three popular kindergartens that a family also considers to be their most preferred, it might not be in their best interest to list those three. As they are the most popular, the family has a low probability of receiving a place in them and might thus remain unassigned. This means they should be strategic when revealing their preferences.

In Rae some parents might like to in attend kindergartens in other catchment areas. There is currently no predefined procedure for this, but it is managed on a case-by-case basis. If parents from several catchment areas had preferences elsewhere, then there is a risk of justified envy in the allocation. It might also be an issue because not all families would be aware that it is possible to attend a kindergarten outside their catchment area.

The solution would be to allow families to list all the kindergartens they are willing to attend. Together with the child-proposing Deferred-Acceptance mechanism, this would eliminate the need for families to be strategic about preferences.

3.2 Policy design

For the matching mechanism we chose to use the child-proposing Deferred-Acceptance algorithm, for its incentive compatibility and proposing-side utility-maximising property without priority violations. In the literature on distributive justice, discussion of fair equality of opportunity (fair access in our case) is often accompanied by discussion on the principles of positive discrimination, like the rawlsian difference principle Rawls (1971). In our paper we define fair access as the chance for the family to access their preferred kindergarten. Moreover, we include in our design some positive discrimination, or controlled choice, through policies such as prioritising siblings. However, we do not have any metrics that measure the utility in a similar way to the difference principle, which aims to maximise the utility of the worst-off. In addition, fair access is essentially different to the efficiency metrics for the priorities of local municipalities and the preferences of families, which indicate average or mean outcomes, i.e. mean preference utilisation. We measure fair access as the proportion of families placed in their most preferred kindergarten with at least 10 % of lotteries. The chosen percentage of lotteries is somewhat arbitrary, a greater percentage would be a stronger condition of fairness. Since not all policy designs use lotteries, some will be inherently unfair.

The weaker side in two sided markets is generally the proposing side, in our case the families or children and their preference utilisation, or utility maximisation, is the design goal. In addition, the mechanism allows some policy goals, we consider:

- having siblings in the same kindergarten
- placing children in a kindergarten near their home

Additional goals could be prioritisation of disadvantaged families or children with special needs, but data limitations cause these not to be our focus.

A short list of policy goals does not indicate that policy designs are limited to two alternatives, as the design of policy for siblings and proximity can vary. Children with siblings might always have a priority over others, or might only be prioritised over families living further away. Proximity could be considered in multiple different ways, such as a walk-zone or catchment area or geographical distance.

A simple way to consider geographic aspects is to define catchment areas for each kindergarten, and prioritise the children living in the catchment area where the kindergarten is located. The drawback of this method is that these priorities may not reflect the personalised distances, as a kindergarten might be relatively far from an address in the same area, whilst another kindergarten in a different area can actually be nearby. Therefore it may be more appropriate to use personalised distances. We can use continuous (real) distance or somehow discretise, such as giving priority to a kindergarten within 10 minute walking distance, or giving priority to the closest, or several closest, kindergartens. Another option is to give high priority to some number of nearby kindergartens, as is being evaluated in Boston (Shi, 2015).

When we consider children in walk-zones to be a higher priority and then children with siblings we obtain the following priority groups: 1. siblings in walk-zones, 2. children in walk-zones, 3. siblings, 4. all the rest. We might also consider siblings to be a higher priority, which would give us the priority groups: 1. siblings in walk-zones, 2. siblings, 3.

children in walk-zones, 4. all the rest. This simple classification is used in many US cities, such as New York (Abdulkadiroğlu et al., 2005a) and Boston (Abdulkadiroğlu et al., 2005b), together with a randomised lottery for breaking ties. The lottery can be also conducted in two ways, either as a single lottery which is used in all kindergartens, or as multiple lotteries, with one at each kindergarten. The typical choice, used in most US school choice programmes and also in Irish higher education admissions (Chen, 2012), is the single lottery. We will also use this in our simulations. This question is discussed further in Ashlagi and Nikzad (2015) and Pathak and Sethuraman (2011).

If it is considered undesirable that a high proportion of children get admitted by sibling priority, then one option is to set a quota for siblings, of 50% of the places for example. In this case there is high priority for siblings for only some proportion of the places available, and the remaining places are prioritised by distance only. In such a setting it is crucial how the allocation is implemented. It can be done by allocating the places for siblings first and then the remaining seats or in reverse. In Dur et al. (2013) it is shown that the reverse approach can benefit children with siblings, and in Hafalir et al. (2013) it is shown that reserving places for some minority results in a better allocation for the minority than limiting the quota for the majority does. Under the latter policy both groups, minority and majority, could be worse off. We evaluate the design of policy by reservation of places for siblings or for families living nearby.

The Deferred-Acceptance algorithm is slightly modified to accommodate reserves and quotas. The priority quotas are considered as separate kindergartens. The child is first placed in a quota group higher in the precedence order and if rejected then in one lower and so on. So each child will be placed in the highest precedence quota group.

Table 1: Summary of policies (priority order in parenthesis)

Policy	**Distance** (D)	**Siblings** (S)	**Lottery**	**Quotas (Precedence)**
DA1	absolute (2)	(1)	no	no
DA2	walk-zone (2)	(1)	(3)	no
DA3	walk-zone (1)	(2)	(3)	no
DA4	3 closest (2)	(1)	(3)	no
DA5	absolute (2)	(1)	no	[80%, 20%] ([D, S+D])
DA6	absolute (2)	(1)	no	[20%, 80%] ([S+D, D])
DA7	relative (2)	(1)	(3)	no

From this discussion we settle on seven priority policies (summarised in Table 1) for evaluation:

DA1. Children with siblings always have the highest priority and children living closer have higher priority. Priority classes would be considered in the order: 1) siblings; 2) walking distance.

DA2. Children with siblings always have the highest priority, then children in the walk-zone have higher priority. The walk-zone is defined as 10 minutes walking distance from home. Additional ties are ordered by a single random lottery for all kindergartens. The order of priority classes is: 1) siblings+walk-zone; 2) siblings; 3) walk-zone; 4) remaining.

DA3. Children in the walk-zone always have the highest priority, then children with siblings have higher priority. Additional ties are ordered by a single random lottery for all kindergartens. The order of priority classes is: 1) siblings+walk-zone; 2) walk-zone; 3) siblings; 4) remaining.

DA4. Children with siblings always have highest priority, and children have higher priority for the three closest kindergartens. Additional ties are ordered by a single random lottery for all kindergartens. Priority precedence order: 1) siblings+one-of-three-closest; 2) siblings; 3) one-of-three-closest; 4) remaining

DA5. Children with siblings have highest priority for the reserved 20% of places, otherwise priority is by distance. Precedence order: 1) by distance up to 80%; 2) children with siblings+distance up to 20%; 3) remaining places, if any, by distance.

DA6. Children with siblings have highest priority for the reserved 20% of places, otherwise priority is by distance. Precedence order: 1) children with siblings+distance up to 20%; 2) remaining places, if any, by distance.

DA7. Children with siblings always have highest priority, and children have higher priority in the closest kindergarten, second highest in the second-closest etc. Additional ties are ordered by a single random lottery for all kindergartens. Priority precedence order: 1) siblings; 2) closest-number.

4 Data and initial policy design comparison

We use preference data from Harku kindergarten matching in spring 2015. We have preference data with geographic location for 143 children. The preferences for kindergartens were mostly collected 2-3 years ago when the children were born, as that is the time when parent made their applications to kindergartens. Parents could submit up-to three preferences, and the actual number of preferences submitted is aggregated in Table 2. Most families submit three preferences.

Table 2: Number of preferences submitted

Submitted preferences	**Number of families**
1	13
2	6
3	124

We also have the spring 2015 allocation data from Harku. The allocation was based on the procedure described in section 2.1. In total 123 children from our sample were allocated.

In reality slightly more children were allocated, but some were excluded from our sample because no location data was available for them, because the address was missing or the address lookup failed, or preference data were missing as we had an earlier snapshot of preference data and data are not available for some late applicants.

Table 3: Harku allocation

Kindergarten	**Number of places**
A	26
B	4
C	11
D	22
E	18
F	39
G	3
Total	123

The number of children allocated to each kindergarten in Harku is shown in Table 3. Taking the same number of places available in each kindergarten we look at counter-factual allocations based on the policies in section 3.2 in Table 4. For policies that use lotteries, we ran ten random realisations. The lotteries are also different across policies.

Policy DA0 uses the child-proposing Deferred-Acceptance with the original priorities based on application date and siblings. We see that this policy has a hard time satisfying the locality goal, as on average families live about 4 km, which is quite far, from kindergartens.

Even with the actual Harku matching we see that with the application date priority children are placed on average 4.12 km from kindergartens, which is more than with other policies. Also we see that some children were assigned in kindergartens not in their preference lists, if we exclude them the average distance would become 3.75, so these children are further away than average.

Blocking pairs arise when a family has a higher priority in a kindergarten than the child assigned with the lowest priority and the same family prefers that kindergarten to their current assignment. If the child is assigned in a kindergarten not in their revealed three preferences we consider them assigned to their fourth or unassigned, the result in the number of blocking pairs will be the same. We see that Harku matching has 32 families with blocking pairs, indicating it is hard to guarantee manually that the matching would be stable. With Immediate-Acceptance (IA) matching we see that the number of blocking pairs could be lower. An explanation for this can be that there are cases where the preferences might have changed and the heads of kindergartens have used updated information when making decisions, but this is not reflected in the central data store. An indication that the preferences might have changed is that nine children were assigned in kindergartens not initially listed.

Looking at the siblings data in Table 5 we see that $\frac{23}{32} \approx 72\%$ families listed the kindergarten where a sibling is as their first choice. This also explains why with Immediate-

Table 4: Comparison of policies using reported preferences

Policy	**Mean preference**	**First**	**Unassigned**	**Mean distance (km)**	**With siblings**	**Blocking agents**
Harku[a]	1.25	100	9 (18)	4.12 (3.75)	97 % (97 %)	32
IA	1.13	111.00	18.00	3.40	78 %	13
DA 0	1.28	99	18	3.97	84 %	
DA 1	1.25	102	18	2.77	84 %	
DA 2	1.33	94.80	18.00	3.47	82 %	
DA 3	1.37	90.20	18.00	3.62	87 %	
DA 4	1.33	91.80	18.00	3.34	91 %	
DA 5	1.25	102	18	2.77	84 %	
DA 6	1.25	102	18	2.77	84 %	
DA 7	1.23	103.00	18.40	2.85	80 %	

[a]In parenthesis we also count as unallocated children in kindergartens not in their revealed preferences. They are also excluded when computing mean preference, because their allocated preference is unknown

Acceptance some 22 % of siblings are not in the same kindergarten as their sibling, since that kindergarten was not high in the submitted preferences, because at the time of the application it was not yet known which kindergarten the siblings would attend.

Not surprisingly we obtain the lowest average distance with policy DA1, where prioritisation is by siblings and distance. This policy does not lose in preferences on aggregate when compared to the actual Harku matching, as in both cases the mean preference in the matching is 1.25.

5 Policy sensitivity to preferences

There are multiple points that we are interested in when comparing policies. First we want to evaluate how the policies would work if we collected full preferences from the families, and second we want to evaluate the robustness of the policies to changes in families' preference profiles (see more on preference generation and estimation in the appendix). The main dimensions of the evaluation are the preference rank achieved in an allocation and the effect on societal welfare measures of the average distance from kindergartens and the share of siblings in the same kindergarten.

We compare policies with different generated parametrised preferences profiles. We characterise preference profiles by conditional probabilities (more in appendix) of families:

- ranking a closer kindergarten higher ($\Pr(r_i \succ r_j \mid d_i < d_j)$, $i \neq j$)

Table 5: Preferences with siblings

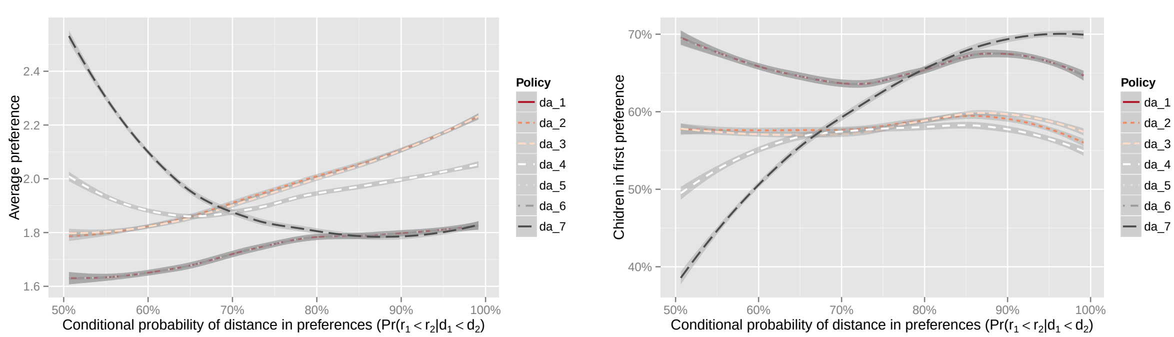
Preference no	**Number of families**
1	23
2	7
3	2
Total	32

- ranking a kindergarten with a sibling higher ($\Pr(r_i \succ r_j \mid s_i > s_j)$, $i \neq j$)

Where r_i is rank of kindergarten i, d_i is distance to kindergarten i and s_i is sibling status in kindergarten i.

For statistical comparison of each of the parameter values we generate ten preference profiles. For each policy that has a lottery we additionally run ten different randomised lotteries for each. In Figures 1, 2, 3 and 4 we show the average results of the ten allocations over policies with a 95% confidence bound for the estimate.

In Figures 1a and 1b we show average preference got and the proportion of families getting their first preference for all policies. Policy DA7 is the most sensitive to changes in families' preferences. When preferences are strictly based on distance with conditional probability of ≈ 1 then it produces the highest average rank score, one which is similar to other policies as DA1, DA5 and DA6. Surprisingly, when families preferences are close to random, with conditional probability of ≈ 0.5, then DA7 is the policy that has by far the lowest average rank and the lowest number of families with a first preference. At face value DA7 seems the most egalitarian as every family has the highest priority in at least one of the kindergartens. However it seems to be that families that do not prefer to be in the closest kindergarten, tend to be rejected from preferred kindergartens further away and where they have a lower priority. Since the preferences and priorities are not aligned, the probability of the family being rejected in some round is higher. With other policies the probability of being rejected in some round seems to be smaller.



(a) Average preference

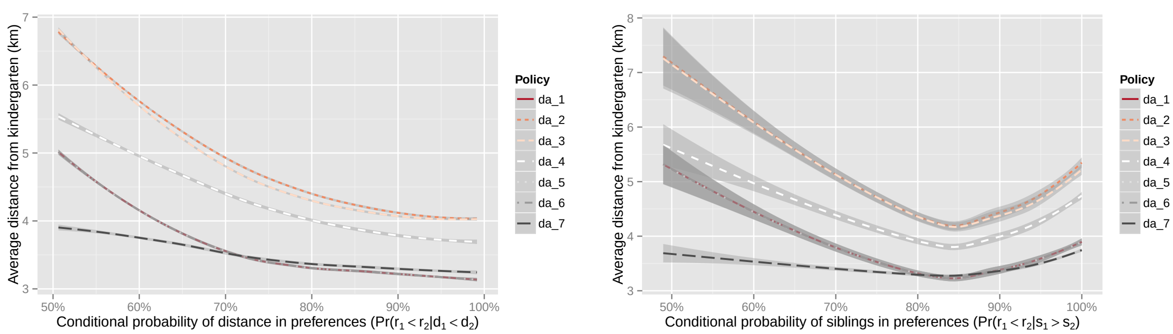
(b) Proportion with first preference

Figure 1: Conditional probability of distance

In terms of average preference policies DA2 and DA3 are indistinguishable from each other, which most likely because there are very few siblings in this dataset. It is also clear when discretised distance is used, as in policies DA2, DA3 and DA4, families do worse on average than when continuous distance is used, as in DA1, DA5 and DA6. The policies DA1 and DA6 always produce exactly the same matching, DA5 is on some occasions, for about 2-6 children, slightly different, but they still have similar aggregate results.

Using reserves for siblings or for children without a sibling we see that it does not make a difference as DA1, DA5 and DA6 are perfectly aligned. Similarly, this is most likely because the selected reserve of 20 % is close to the percent of siblings in the dataset.

Interestingly most policies, with the exception of DA7, are quite robust toward changes in preferences. There is almost always roughly the same proportion of families receiving their first preference, about 50% to 60% with DA2, DA3 and DA4 and 60% to 70% with DA1, DA5 and DA6. In the average preference we see a slight increase when preferences become determined more by distance.



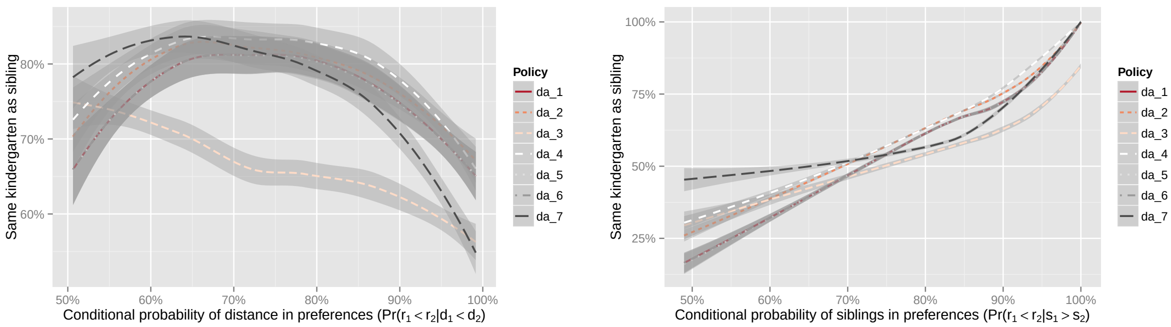
(a) (b)

Figure 2: Average distance and conditional probability of distance in preferences

In Figures 2a and 2b we show the average distance between families and kindergartens. The average distance is smaller with all policies when families' preferences are determined more by distance. Interestingly the smallest average distance is always with DA7, as it minimises distance even at the cost of having worse average preference for families. The average distance is largest with DA2 and DA3, when policies are based on walk-zones, indicating the randomness in priorities.

We get a small improvement in average distance by not using discretisation by walk-zone, but by having a higher priority for a fixed number of kindergartens as in DA4. With random preferences this is a trade-off between achieved preference and distance in DA7 and DA2 and DA3 and on average DA4 is right in the middle of these policies. When preferences are determined more by distance, then DA7 is always better by both, average preference and distance. The policy DA1 has slightly greater average distance with random preferences, but better average achieved preference than DA7, so we can say DA1 has more consideration for families' preferences.

In Figure 2b we show the average distance taking the conditional probability of siblings. Like in Figure 2a we have a decline for all policies in average distance when families have a higher preference for a kindergarten with siblings. This is most probably because the kindergarten for the sibling was already selected by distance, so distance and siblings are correlated. This is not always the case though, as there is an increase in average distance when families always prefer kindergartens with siblings, because not all families have a sibling in a nearby kindergarten.

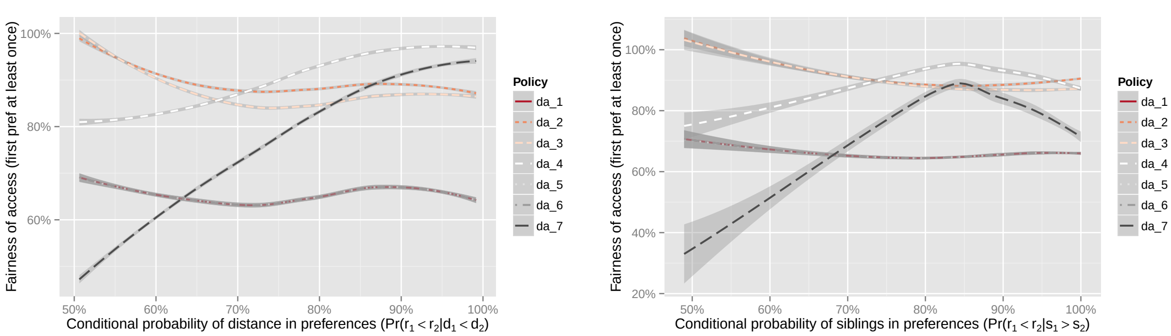


(a) (b)

Figure 3: Probability of being in the same kindergarten with siblings

In Figures 3a and 3b we show the probability of children being in the same kindergarten as their siblings. When families' preferences are random, most policies place about 20% to 30% of siblings in a kindergarten with their siblings, with the exception of DA7, where the percentage is closer to 45%. This higher percentage is most likely due to siblings already being in a nearby kindergarten and, as we have seen, with DA7 families get lower preferences and children are usually placed into nearby kindergartens.

Only DA3 provides the notable exception of having fewer children in kindergartens with siblings, which is because it gives higher priority to walk-zone families than siblings. Were families to have a 100% preference rate for having their children in the same kindergarten, most policies achieve this, and only DA3 has a lower percentage, around 80%. Despite the difference in the precedence order between polices DA5 and DA6, about the same proportion of siblings are places in the same kindergarten in both cases, though this might be a consequence of there being a low number of siblings in the dataset.



(a) (b)

Figure 4: Fairness of access (probability of having first preference with at least one lottery)

In Figures 4a and 4b we measure the probability of a child being matched to the family's first preference in at least one lottery. This is a kind of a measure of fairness of access to kindergartens, similar to the measure of access in Shi (2015). We have plotted policies DA1, DA5 and DA6 the fairness of access even though it does not have any sensible interpretation, since there are no lotteries. These policies are still useful for comparison however.

With the lotteries in DA2, DA3 and DA4, about 80% to 95% of families are placed at least once in a kindergarten that is their first preference. There is a slight fluctuation in some of the polices, but whenever there is a change in the preferences they still stay within these bounds. The policy DA7 seems to be the most unfair, when preferences are random or close to it, as only in 50% of cases are families placed in their most preferred kindergarten.

6 Conclusion and discussion

We have reviewed the kindergarten matching practices in two municipalities in Estonia. While the procedures for allocation were different, the unifying theme was the poor handling of family preferences. In one case the data collected are out-of-date by the time of the allocation and the resulting allocation might not be free of justified envy. Furthermore if the listing of preferences is limited families have to think strategically about how to reveal their preferences.

In the other case the choice is limited to catchment areas. If families have preferences outside the area exchanges might be possible. However, this requires additional effort, so these exchanges between catchment areas rarely materialise. Additionally families can only apply for a kindergarten in their area, but where there are of multiple kindergartens they cannot express their preferences in the application. As a result it is unclear whether the place allocated place is the most preferred for the family or not.

Practices that are well known from matching mechanism design present solutions to some of the problems while also providing a policy tool for the local municipalities. The practices consist of:

- getting full rather than limited preferences from families
- using child-proposing stable matching for allocating places
- defining clear policies for the local municipality to solve the issue of demand in a kindergarten

We realise that while the policy goals might be clear, the choice of exactly which way to implemented can have vastly different results. In most cases the goals of the local municipalities are to have siblings in the same kindergarten and to provide a place in a kindergarten close to home. Of course the main consideration is still to provide a place in a preferred kindergarten. We evaluated seven different policies for implementing the policy goals, using estimated preferences based on data from applications in one of the municipalities.

The simplest policy is to give higher priority to children with siblings and to families living nearby, which is policy DA1. This also demonstrated to be one of the most effective. The resulting allocation had, on average, matched a lot of families with their most preferred kindergarten, while also having one of the smallest average distances. This remained true when the preferences of families were agnostic about distance.

Policy DA1 might occasionally seem unfair, as small differences in distance might affect whether families are placed in their first preference or some lower one. Policies DA2, DA3 and DA4 group kindergartens by distance within equal priority classes, DA2 and DA3 by defining a walk-zone and DA4 by having high priority in the three-closest kindergartens. Families in the walk-zone are treated equally and priorities are defined by lottery. We saw that while this might create a more egalitarian access to kindergartens, it is not without its cost. The average number of children who reach their most preferred kindergarten is usually significantly lower and the average distance is greater. But if having more egalitarian access is important, policy DA4 would do the best of the three. The level of fair access is the same, satisfaction of average preferences is the best, and distance is the lowest.

Another source of seemingly unfair treatment might be perceived from always having higher priority for siblings. If a family already has a child in a kindergarten they are almost guaranteed to get a place in the same kindergarten for a sibling, even when there is another family living even closer. We considered two policies, DA5 and DA6, which limit the number of places in a kindergarten that consider having a sibling to be a priority at up to 20%. Even though the number of places for siblings was low, most families still received a place in that kindergarten if they preferred to. There is almost no difference from policy DA1 on any measure, nor between DA5 and DA6, although theoretically DA6 should provide more opportunity to nearby families, and DA5 to children with siblings.

Clearly an oddity is policy DA7, which was initially designed to deliver more equal access to kindergartens for families who live far from all kindergartens. While policy DA1 would give such families low priority everywhere, DA7 would still give them the highest priority in their closest kindergarten. When most families have a high preference for nearby kindergartens and those where siblings are, the result of DA7 is on par with the best policy designs, as it gives many families their first preference, has the shortest average distance, and has even one of the best results for equality of access. However, the result is radically different when family preferences are mostly idiosyncratic and are almost independent from distance and siblings. In this case DA7 is the worst policy of all for families. Only half the number of children get matched to their first preference, but the average distance is the lowest of all policies and still significantly lower than the second best policy, DA1.

The learning from policy DA7 seems to be that the policy designer looking to have efficiency and fair access to kindergartens needs to be able to predict fairly accurately the preferences of the society. When preferences and priorities are aligned, both of the goals can be met. On the downside, this design is vulnerable to misalignment, as then the price paid in terms of efficiency and fairness is significant. If a local municipality wants to enforce a policy, say by minimising the distances between homes and kindergartens, then this is an option.

Moreover we are left with several unsolved issues that cannot be solved by clever design. For example it is best to coordinate the allocation between neighbouring municipalities, but cooperation is sometimes hard to achieve. Similarly it would be best to know beforehand about children with special needs, but this is often infeasible.

A few interesting aspects of designing a more flexible mechanism might improve the allocation for families. Making decisions on the size and the age composition of the groups in kindergartens, determining this in an optimal way from the application data, could give an additional boost to the number of families receiving a place in their most preferred kindergarten. Some of this research has been done in terms of lower quotas for opening groups (Biró et al., 2010).

A potential way to manage the shortage of kindergarten places is to provide monetary incentives for parents to stay at home with their children or to seek a place in private childcare. The question of how to set this monetary compensation in an optimal way is an interesting research question. Optimal here could mean minimising the cost of providing childcare services in the municipality.

7 Further issues

We identify a dozen further special features that should be considered in the redesign of the mechanism. However, many of these features may pose significant challenges and need additional research. We try to address these issues, and give recommendations for possible adjustments of the basic mechanism.

Children with special needs. In larger cities there are schools for children with special needs, but in smaller municipalities these pupils are mixed in with others. The standard practice is that kindergartens reserve places for children with special needs who require more attention and so are considered to take up the space of three children. Usually it is not known beforehand if there will be any such cases and the special needs may only become evident later. In most cases though, the extra places remain free and can be allocated later to other children. Obviously this has some effect on the fairness of the allocation.

A possible solution would be to have this data available before the allocation is made and to take it into account in the allocation process. However, evaluating in advance all of the children applying could be very costly compared to the extra efforts needed for the reallocation process and the potential issues arising from the extended solution. It would be helpful if the parents of children that are likely to need special treatment were to register for evaluation, but then it should be guaranteed that their chances of admission to their preferred kindergartens would not be worsened, perhaps by giving priority for some number of places in each kindergarten to such children.

Allocation in multiple rounds. Harku currently allocates students in multiple rounds, since two extra places can be taken in each kindergarten where no student with special needs is admitted. The proportion of disadvantaged families is about 10 % (Ministry of Social Affairs, 2015) and children with special needs make up about 3 % (Paat et al., 2011). This question is similar to the question of the design of two-stage allocation mechanisms (Dur and Kesten, 2014) and also to the design of appeal processes (Dur and Kesten, 2015). The first option is to allocate the extra places among the unmatched children only. This is a simple method with no reallocation of children, but it can be seen as unfair for families who are allocated seats in the main round and would prefer an extra place in a kindergarten where they have higher priority than the unallocated children who get those extra places in the second round. Besides the justified envy in the final solution, the parents might also act strategically in the main round, perhaps by not accepting an offer from the kindergarten listed second, especially if they have information that they are first in the waiting list and the creation of extra places is very likely. Therefore is seems reasonable to let everyone apply for the extra places, as is currently done in Harku. However, if the process is not centralised then those who were assigned a place in the first round but now get a better match would then create new available places. Even if this decentralised process could be continued until a stable solution was reached, then this proposal-rejection chain would result in a stable matching that is the worst possible stable matching for the reallocated children, as proved in Blum et al. (1997). Therefore this process would not be strategy-proof for the parents either, and so the only possible solution that is strategy-proof for the parents and avoids justified envy is a centralised second round, where parents can re-apply to all kindergartens with the option of keeping their assignment if they wish to do so (technically this is achieved by putting the children already assigned to the kindergarten at the top of its ranking). Yet, this solution may affect a significant number of children, and in theory possibly all of them, which could result is high reallocation costs. This reallocation cost would be accepted by the parents, since they would always have the option of not changing their assignment, but could be seen as undesired by the local council and the kindergartens.

Children with existing places . The parents of some children may request a transfer for their children. This is especially relevant for those attending a class for 2-3-year olds who then want to go to a different kindergarten for 3-6-year olds since the classes for children aged 2-3 may not be available in their favourite places. It is therefore a question of whether the reallocation of these children should be conducted as part of the yearly matching round. If so, then these children should be guaranteed to get at least as good a seat in the reallocation, meaning they should have the highest priority in their current kindergarten. This question has been studied in the context of Danish daycare allocation Kennes et al. (2014), and also for the reallocation of French teachers Combe et al. (2015).

Overlapping admission processes. Some parents may be registered in more than one municipality, so they are able to apply for a place for their child in two systems, for example in Harku and in Tallinn. This can lead to inefficiencies due to cancellations. Similar problems arise in some US cities where state schools and charter schools hold their admissions separately. Furthermore, the same phenomenon has also appeared in European college admission programmes, where an increasing number of students are applying for programmes in several countries, and this disturbs the national matching schemes.

Outside options with subsidies. Somewhat related to this issue is that private kindergartens operate in Estonia, and some parents also consider the option of home schooling. However, if a municipality cannot provide enough kindergarten places for its resident population, in some cases it may subsidise parents who choose an alternative option. In Rae, the local council pays support to parents who do not receive a place in any kindergarten in their catchment area, but they may withdraw the support if the parents do not accept a place that is offered at a local kindergarten. This conditional support can lead to strategic considerations, since some parents may find an alternative home or private option preferable to a local school if and only if they receive the financial support, but this cannot be stated in the application. This special case can be modelled with the matching with contracts framework. A similar special feature is found in the Hungarian higher education matching scheme, where students can study on the same course under two different contracts, either for free or with a tuition fee. Furthermore, US cadets Sönmez and Switzer (2013) also face such a situation when they decide whether or not they are willing to take some extra years of service in order to increase their chances of admission. The recommended solution is to let the parents list the option of not having a place in the kindergarten but getting the support instead, when they give their applications to kindergartens, meaning that all the listed options are considered preferable to the outside option with no financial support. In such a case, it is then crucial that the parents-optimal stable solution is implemented so as to make the parents reveal their true preferences for these outside options.

Lower quotas, opening of new groups. Sometimes the kindergartens are able to cancel groups or open new ones to fit with the applications. In particular, there are regulations over the minimum number of children they need to start a new group. This feature is similar to the lower quotas used in the Hungarian higher education matching scheme Biró et al. (2010), where programmes can be cancelled if there is a lack of students. This is a natural requirement that makes the education service economical, but the theoretical model for college admissions with lower quotas is not always solvable, meaning a fair solution does not always exist and the problem of finding a fair solution is NP-hard. The problem becomes even more complicated if new groups can be created, since both the closures and the openings in a kindergarten affect the number of students admitted elsewhere. However, clever heuristics and robust optimisation techniques, such as integer programming Biró et al. (2014) can be used to tackle these generalised problems.

Same age groups and mixed groups. In Estonia there are both same age groups and mixed groups. Having only same age groups can vary the number of groups opened in a kindergarten, as a kindergarten with five groups could open only one group in every three years, which would be unsatisfactory for the local children in that year. When mixed groups are created, the number of children admitted can be relatively stable if they can always fill the available places. However, if there are some free places left in a year, then the age distribution of the children can be distorted.

Sharing places. In some kindergartens it is possible that some children attend part of the week only and the rest of the time is used by some other children. This possibility again makes the underlying problem challenging to solve. Specifically, when we have a large number of half-time students then we might face the same problem as allocating doctors and couples to hospitals, which is an NP-hard problem McDermid and Manlove (2010).

Historic dependence of preferences. In Harku, the applications of registered parents are listed in a public website. In Tallinn the number of applications already submitted to the kindergartens is also published. If the registration date is the primary criterion for priority and the parents can see the applications, or the number of applications, made before their turn, then this can affect their true and submitted preferences. It may be that if there are more applications than places then they will find it risky to apply. This can depend on the birth date of the child, because if a child was born soon after 1 October, then the parents could have a good chance for a place everywhere, and so be more truthful. We did not find much evidence of significant changes in the preferences over time in the Harku data, but in a similar study for Tallinn, or other places where the registration date is important, attention should be paid to the potentially biased preferences caused by the published information on past applications.

Smooth transition to a new system. When designing the new mechanism, it should maybe be considered whether there will be a smooth transition between the two systems. This can be especially challenging if the current priorities are based on registration date, since those parents who registered early may see it as unfair for this priority that they earned in the past to be suddenly neglected. Therefore perhaps the transition to a new system should be longer, keeping the priority of those who have already registered in the old regime. Or, at least, the registration date could be used as a tie-breaker in some new policies.

The role of the heads of the kindergartens. The heads of the kindergartens are now actively involved in the allocation system, and the discussions between them and the personal communication with the parents are crucial in the current system. In the centrally coordinated system they may fear losing their chance to influence the allocations, and the same could apply for employees of the local municipality. It should be considered whether they could still have some power to adjust the priorities, or to make other decisions about their kindergartens, such as whether to open a new group.

How fair is it to use proximity as a priority. Whether the use of proximity is fair may depend on the ease/cost of registering: a) whether it is almost costless (as in Hungary); b) whether there are some significant costs such as renting or having a flat locally; or c) whether the family really has to live there, as for example in Barcelona, where somebody who is proven not to live at the stated address can lose their place. When it is easy to register at an address then the parents may play a strategic game in which the first stage is to choose an address. When ownership and actual residency are required, and the priorities are important for the parents, this can affect the housing choices of the families, and influence house prices and also the socio-economic distribution of the population.

Restricting the choice of the parents. A simple way is to allow families to apply to nearby kindergartens only (such as those within the catchment area, as is currently done in Rae. A more sophisticated way is to provide personalised menus of choices, such as the system being proposed in the Boston school choice mechanism Shi (2015). This would potentially provide parents with a choice from schools close to them and also where they have siblings already attending, with a limited number of further options. The advantage of this method over restricting the number of applications is that the mechanism remains strategy-proof, and the parents have a simpler task in ranking the available options. However, the disadvantage is that it is hard to estimate the preferences of the parents, and therefore there is a risk that some highly preferred kindergartens could be missed from some menus. In general, this type of restrictive policy can improve the overall quality of the allocation from the point of view of the municipality, perhaps by reducing the total travel distance, which was the main motivation in the Boston school choice redesign, as the bus costs had to be limited, but it can badly affect the overall welfare of the children. We do not recommend that this policy be used in Harku and Rae, due to their small size, but we would suggest it could be considered in larger cities like Tallinn.

Acknowledgement

This work was partially supported by Norway Grants V653. André Veski acknowledges support from European Cooperation in Science and Technology (COST) action IC1205 and Estonian IT Academy scholarship for PhD students. Péter Biró acknowledges support from the Hungarian Academy of Sciences under its Momentum Programme (LD-004/2010), the Hungarian Scientific Research Fund, OTKA, Grant No. K108673, and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

Abdulkadiroğlu, A., Che, Y.-K., and Yasuda, Y. (2011). Resolving Conflicting Preferences in School Choice: The "Boston Mechanism" Reconsidered. *American Economic Review*, 101(1):399–410.

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2005a). The New York City High School Match. *American Economic Review*, 95(2):364–367.

Abdulkadiroğlu, A., Pathak, P. A., Roth, A. E., and Sönmez, T. (2005b). The Boston Public School Match. *American Economic Review*, 95(2):368–371.

Abdulkadiroğlu, A., Pathak, P. A., Roth, A. E., and Sönmez, T. (2006). Changing the Boston School Choice Mechanism. *National Bureau of Economic Research, Working Paper Series*, 11965.

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

Ashlagi, I. and Nikzad, A. (2015). What matters in tie-breaking rules? How competition guides design.

Biró, P., Fleiner, T., Irving, R. W., and Manlove, D. F. (2010). The College Admissions problem with lower and common quotas. *Theoretical Computer Science*, 411(34-36):3136–3153.

Biró, P., Manlove, D. F., and McBride, I. (2014). The Hospitals / Residents Problem with Couples: Complexity and Integer Programming Models. In Gudmundsson, J. and Katajainen, J., editors, *Experimental Algorithms*, pages 10–21. Springer International Publishing.

Blum, Y., Roth, A. E., and Rothblum, U. G. (1997). Vacancy Chains and Equilibration in Senior-Level Labor Markets. *Journal of Economic Theory*, 76(2):362–411.

Calsamiglia, C., Haeringer, G., and Klijn, F. (2010). Constrained School Choice: An Experimental Study. *American Economic Review*, 100(4):1860–1874.

Chen, L. (2012). University admission practices – Ireland. `http://www.matching-in-practice.eu/higher-education-in-ireland/`. accessed 2016-01-18.

Cohen, W. W., Schapire, R. E., and Singer, Y. (2011). Learning to Order Things. *Journal of Artificial Intelligence Research*, 10:243–270.

Combe, J., Tercieux, O., and Terrier, C. (2015). The Design of Teacher Assignment: Theory and Evidence.

Dur, U. M. and Kesten, O. (2014). Sequential versus Simultaneous Assignment Systems and Two Applications.

Dur, U. M. and Kesten, O. (2015). The Appeals Process in NYC High School Match.

Dur, U. M., Kominers, S. D., Pathak, P. A., and Sönmez, T. (2013). The Demise of Walk Zones in Boston: Priorities vs. Precedence in School Choice.

Gale, D. and Shapley, L. S. (1962). College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 69(1):9–15.

Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company.

Google (2015). The Google Maps Distance Matrix API. `https://maps.googleapis.com/maps/api/distancematrix/`. accessed 2015-12-31.

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic Theory*, 144(5):1921–1947.

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

Kennes, J., Monte, D., and Tumennasan, N. (2014). The Day Care Assignment: A Dynamic Matching Problem. *American Economic Journal: Microeconomics*, 6(4):362–406.

Knuth, D. E. (1997). *Stable Marriage and its Relation to Other Combinatorial Problems*. American Mathematical Society.

McDermid, E. J. and Manlove, D. F. (2010). Keeping partners together: algorithmic results for the hospitals/residents problem with couples. *Journal of Combinatorial Optimization*, 19(3):279–303.

Meinshausen, N. and Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473.

Ministry of Social Affairs (2015). Peretoetuste, teenuste ja vanemapuhkuste roheline raamat. Technical report, Ministry of Social Affairs.

Paat, G., Kaarna, R., and Aaviksoo, A. (2011). Meditsiiniliste erivajadustega laste üldhariduse korralduse analüüs. Technical report, Praxis Centre for Policy Studies, Tallinn.

Pathak, P. A. and Sethuraman, J. (2011). Lotteries in student assignment: An equivalence result. *Theoretical Economics*, 6(1):1–17.

Pathak, P. A. and Sönmez, T. (2013). School Admissions Reform in Chicago and England: Comparing Mechanisms by their Vulnerability to Manipulation. *American Economic Review*, 103(1):80–106.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press, Cambridge, MA.

Roth, A. E. (1982). The Economics of Matching: Stability and Incentives. *Mathematics of Operations Research*, 7(4):617–628.

Roth, A. E. (2008). Deferred acceptance algorithms: history, theory, practice, and open questions. *International Journal of Game Theory*, 36(3-4):537–569.

Saraceno, C. (2011). Family policies. Concepts , goals and instruments.

Shi, P. (2015). Guiding School-Choice Reform through Novel Applications of Operations Research. *Interfaces*, 45(2):117–132.

Sönmez, T. and Switzer, T. B. (2013). Matching With (Branch-of-Choice) Contracts at the United States Military Academy. *Econometrica*, 81(2):451–488.

West, A., Hind, A., and Pennell, H. (2004). School admissions and 'selection' in comprehensive schools: policy and practice. *Oxford Review of Education*, 30(3):347–369.

A Estimating and generating full preference profiles

To evaluate policy behaviour and robustness in a situation when we could collect more preferences from families we need to reconstruct full preference profile from the limited data. This is also useful evaluating situation when families' preferences functions change or are different in future allocations.

In the current dataset we are limited to three preferences which may not be the families' actual three highest preferences, but manipulated. So we cannot use multinomial logit for ranked-order data as in Shi (2015). Instead we transform the data to pairwise comparisons and learn a simple logistic regression model of when one kindergarten is ranked higher compared to another, this is similar to Cohen et al. (2011). An instance of a family with three preferences would provide six data-points for learning: three distinct pairs of kindergartens and two comparisons per pair. This results is 756 pairwise comparisons in our data. With three preferences per child the maximum number of comparisons would be $846 = 141 * 6$, but not all families have reported three preferences and the actual number of comparisons is 756.

By doing pairwise comparisons we assume that the presented three preferences are still ordered correctly, although they might not be the first three preferences. This shown in Haeringer and Klijn (2009) to be the case when using the Deferred-Acceptance algorithm with limited preferences, but is not necessarily the case with Immediate-Acceptance (Abdulkadiroğlu et al., 2011). Still we assume that this is the case as the families do not know the mechanism and how exactly to manipulate it.

For each family and kindergarten we know the geographical location from address lookup from google maps[1] and Estonian Land Board (Maa-amet[2]) and distance calculations taken from Google maps distance[3]. We have a rich dataset for distance, as for each family-kindergarten pair we know the driving and walking distances in kilometres and minutes. We also have the direct distance between the two points calculated with the haversine formula.

The entire feature set in Table 6. Based on this we construct a dataset of pairwise comparisons of kindergartens with feature set as in Table 7. In the table features with prefix e1 are kindergarten 1 features and e2 kindergarten 2 features. For all kindergartens we have the distance between families home and the kindergarten. The dependent variable y is one, when kindergarten e1 is ranked higher than e2 and zero otherwise.

Table 6: Family's kindergarten features

Feature	Description
sibling	1 if kindergarten has a sibling already attending, 0 otherwise
driving_distance_m	driving distance between family's home and kindergarten, based on Google (2015)
driving_distance_sec	driving time between family's home and kindergarten, based on Google (2015)
walking_distance_m	walking distance between family's home and kindergarten, based on Google (2015)
walking_distance_sec	walking time between family's home and kindergarten, based on Google (2015)
haversine_distance_m	direct distance between family's home and kindergarten

[1]https://developers.google.com/maps/documentation/geocoding/intro
[2]http://inaadress.maaamet.ee/geocoder/bulk
[3]https://developers.google.com/maps/documentation/distance-matrix/intro

Table 7: Kindergarten pairwise comparison features for selection

Feature nr	Feature
0	e1_driving_distance_m
1	e1_driving_distance_sec
2	e1_haversine_distance_m
3	e1_sibling
4	e1_walking_distance_m
5	e1_walking_distance_sec
6	e2_driving_distance_m
7	e2_driving_distance_m/e1_driving_distance_m
8	e2_driving_distance_m/e1_driving_distance_sec
9	e2_driving_distance_m/e1_haversine_distance_m
10	e2_driving_distance_m/e1_walking_distance_m
11	e2_driving_distance_m/e1_walking_distance_sec
12	e2_driving_distance_sec
13	e2_driving_distance_sec/e1_driving_distance_m
14	e2_driving_distance_sec/e1_driving_distance_sec
15	e2_driving_distance_sec/e1_haversine_distance_m
16	e2_driving_distance_sec/e1_walking_distance_m
17	e2_driving_distance_sec/e1_walking_distance_sec
18	e2_haversine_distance_m
19	e2_haversine_distance_m/e1_driving_distance_m
20	e2_haversine_distance_m/e1_driving_distance_sec
21	e2_haversine_distance_m/e1_haversine_distance_m
22	e2_haversine_distance_m/e1_walking_distance_m
23	e2_haversine_distance_m/e1_walking_distance_sec
24	e2_sibling
25	e2_walking_distance_m
26	e2_walking_distance_m/e1_driving_distance_m
27	e2_walking_distance_m/e1_driving_distance_sec
28	e2_walking_distance_m/e1_haversine_distance_m
29	e2_walking_distance_m/e1_walking_distance_m
30	e2_walking_distance_m/e1_walking_distance_sec
31	e2_walking_distance_sec
32	e2_walking_distance_sec/e1_driving_distance_m
33	e2_walking_distance_sec/e1_driving_distance_sec
34	e2_walking_distance_sec/e1_haversine_distance_m
35	e2_walking_distance_sec/e1_walking_distance_m
36	e2_walking_distance_sec/e1_walking_distance_sec
37	log_e2_driving_distance_m/e1_driving_distance_m
38	log_e2_driving_distance_m/e1_driving_distance_sec
39	log_e2_driving_distance_m/e1_haversine_distance_m
40	log_e2_driving_distance_m/e1_walking_distance_m

Feature nr	Feature
41	log_e2_driving_distance_m/e1_walking_distance_sec
42	log_e2_driving_distance_sec/e1_driving_distance_m
43	log_e2_driving_distance_sec/e1_driving_distance_sec
44	log_e2_driving_distance_sec/e1_haversine_distance_m
45	log_e2_driving_distance_sec/e1_walking_distance_m
46	log_e2_driving_distance_sec/e1_walking_distance_sec
47	log_e2_haversine_distance_m/e1_driving_distance_m
48	log_e2_haversine_distance_m/e1_driving_distance_sec
49	log_e2_haversine_distance_m/e1_haversine_distance_m
50	log_e2_haversine_distance_m/e1_walking_distance_m
51	log_e2_haversine_distance_m/e1_walking_distance_sec
52	log_e2_walking_distance_m/e1_driving_distance_m
53	log_e2_walking_distance_m/e1_driving_distance_sec
54	log_e2_walking_distance_m/e1_haversine_distance_m
55	log_e2_walking_distance_m/e1_walking_distance_m
56	log_e2_walking_distance_m/e1_walking_distance_sec
57	log_e2_walking_distance_sec/e1_driving_distance_m
58	log_e2_walking_distance_sec/e1_driving_distance_sec
59	log_e2_walking_distance_sec/e1_haversine_distance_m
60	log_e2_walking_distance_sec/e1_walking_distance_m
61	log_e2_walking_distance_sec/e1_walking_distance_sec
62	sibling_diff
63	sqrt_e2_driving_distance_m/e1_driving_distance_m
64	sqrt_e2_driving_distance_m/e1_driving_distance_sec
65	sqrt_e2_driving_distance_m/e1_haversine_distance_m
66	sqrt_e2_driving_distance_m/e1_walking_distance_m
67	sqrt_e2_driving_distance_m/e1_walking_distance_sec
68	sqrt_e2_driving_distance_sec/e1_driving_distance_m
69	sqrt_e2_driving_distance_sec/e1_driving_distance_sec
70	sqrt_e2_driving_distance_sec/e1_haversine_distance_m
71	sqrt_e2_driving_distance_sec/e1_walking_distance_m
72	sqrt_e2_driving_distance_sec/e1_walking_distance_sec
73	sqrt_e2_haversine_distance_m/e1_driving_distance_m
74	sqrt_e2_haversine_distance_m/e1_driving_distance_sec
75	sqrt_e2_haversine_distance_m/e1_haversine_distance_m
76	sqrt_e2_haversine_distance_m/e1_walking_distance_m
77	sqrt_e2_haversine_distance_m/e1_walking_distance_sec
78	sqrt_e2_walking_distance_m/e1_driving_distance_m
79	sqrt_e2_walking_distance_m/e1_driving_distance_sec
80	sqrt_e2_walking_distance_m/e1_haversine_distance_m
81	sqrt_e2_walking_distance_m/e1_walking_distance_m
82	sqrt_e2_walking_distance_m/e1_walking_distance_sec
83	sqrt_e2_walking_distance_sec/e1_driving_distance_m

Feature nr	Feature
84	sqrt_e2_walking_distance_sec/e1_driving_distance_sec
85	sqrt_e2_walking_distance_sec/e1_haversine_distance_m
86	sqrt_e2_walking_distance_sec/e1_walking_distance_m
87	sqrt_e2_walking_distance_sec/e1_walking_distance_sec

We fit a logistic regression model (1), where $F(k_1, k_2)$ is the probability that a kindergarten k_1 is ranked higher than kindergarten k_2.

$$y = F(k_1, k_2) = \frac{1}{1 + e^{-t(k_1,k_2)}} \tag{1}$$

For feature selection we use stability selection Meinshausen and Bühlmann (2010) in learning library Pedregosa et al. (2011). Regularised logistic regression uses penalisation function as in equation (2), where $\frac{1}{\lambda}$ is hyperparameter to be selected.

$$\min_{w,c} \frac{1}{2} w^T w + \frac{1}{\lambda} \sum_{i=1}^{n} \log(\exp(-y_i(X_i^T w + c)) + 1) \tag{2}$$

We show the resulting stability paths with different hyperparameter $\frac{1}{\lambda}$ values in Figure 5. The probabilities are found based on 1000 re-samples of data. We see that the most often selected features are nr 37 and 62, which are $log\_e2\_driving\_distance\_m/e1\_driving\_distance\_m = log(\frac{e2\_driving\_distance\_m}{e1\_driving\_distance\_m})$ and $sibling\_diff = e1\_sibling - e2\_sibling$. We get the functional form for $t(\cdot)$ from Equation (1) as in Equation (3).

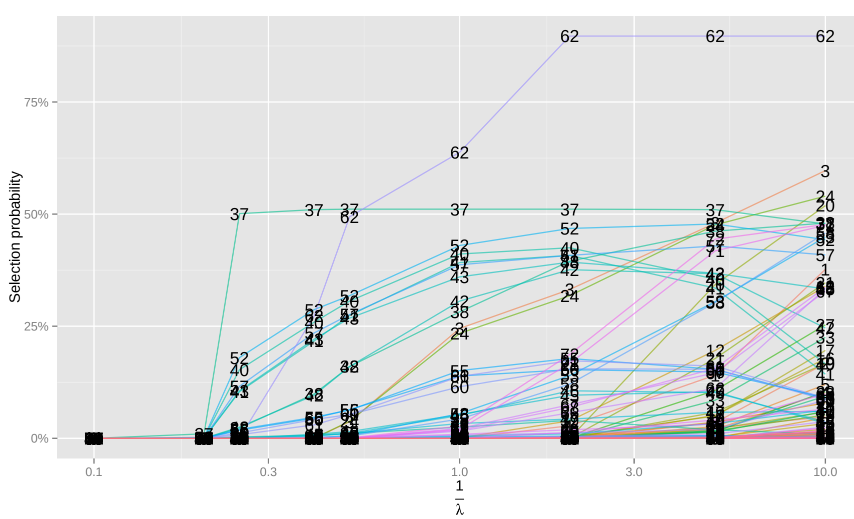


Figure 5: Stability paths

$$t(k_1, k_2) = \beta_0 + \beta_1 \log\left(\frac{d_2}{d_1}\right) + \beta_2 s + \varepsilon \tag{3}$$

We also do a cross-validation with the selected features with L2 regularised logistic regression based on (2) to see the stability of coefficients. We show the results in Figure 6. We see that whatever the regularisation the prediction scores are almost always the same. On one cross-validation fold we obtain a better result with $\frac{1}{\lambda} = 0.1$. As with smaller coefficient the prediction is more stable we prefer higher regularisation.

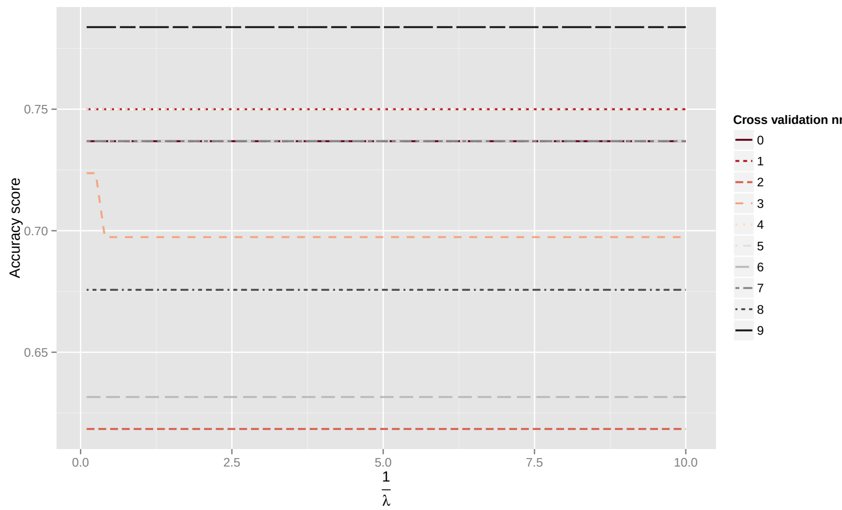


Figure 6: Cross-validation

With $\frac{1}{\lambda} = 0.1$ regularised logistic regression we obtain values for parameters $\beta_0 \approx 0$, $\beta_1 = 0.82$ and $\beta_2 = 1$. In our dataset the number of events when a kindergarten is ranked higher is exactly the same as when it is ranked lower, so we obtain that $\beta_0 \approx 0$. There is no bias in the dataset.

The estimation is an aggregate ranking in Harku not individual utility for obtaining a place in any kindergarten. In function (3) the parameters for two kindergartens, k_1 and k_2, are:

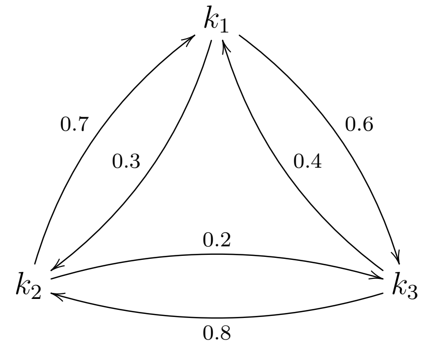
- d_1 distance between the families residence and kindergarten k_1
- d_2 distance between the families residence and kindergarten k_2
- s difference in sibling status of the two kindergartens k_1 and k_2 is
  - 0 when neither of the kindergartens has siblings or both have[4]
  - 1 when only kindergarten k_1 has a sibling
  - -1 when only kindergarten k_2 has a sibling

The ranking function (1) with (3) has at least two useful properties for interpretation, when $\beta_0 = 0$ and not considering the error term ε:

1. $F(k_1, k_2) = 1 - F(k_2, k_1)$. The probability of kindergarten k_1 begin ranked higher is exactly the opposite of probability that k_2 is ranked higher

[4] is this data there was no occasion when a child has siblings in multiple kindergartens

Figure 7: Non-transitive ranking example



2. $F(k_1, k_2) = \frac{1}{2}$, when $d_1 = d_2$ and $s = 0$. When both kindergartens k_1 and k_2 are equal in terms of distance a siblings, the probability than one of them is ranked higher is exactly 50 %.

For policy comparison we estimate the ranking over all kindergartens. We do not model the cut-off levels for outside options, when the family would rather keep the child at home. We assume they would always rather have a place in any of Harku's kindergartens.

To obtain a full ranking of kindergarten we need one more step. The ranking function returns pairwise comparisons, to obtain a full ranked list for a family we need to order all the alternatives. The ordering might be easy by always ranking according to the highest probability when comparing two kindergartens and when there is no non-transitive rankings.

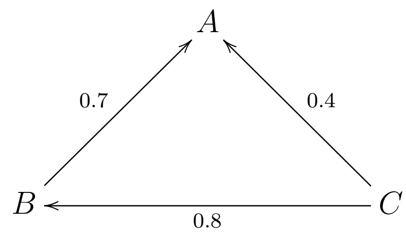
This might not always be the case as we show in Figure 7. Each edge of the graph points in the direction of higher ranked alternative. If we were to keep only the pairwise rankings with higher probability, we would obtain $k_1 \succ k_2$, $k_3 \succ k_1$ and this would indicate that $k_3 \succ k_2$, but in the figure we see that actually $k_2 \succ k_3$ with higher probability. To find a full ordering of alternatives we settle on finding maximum weight of pairwise comparisons, which on the directed graph means finding a maximum acyclic sub-graph. The sub-graph can also be found eliminating edged in minimum feedback arc set. In computer science this is known to be a hard problem (e.g. Garey and Johnson, 1979), but our instance are very small (graph with 7 nodes) so we can solve them exactly. This approach is somewhat similar to Cohen et al. (2011).

By eliminating edges found in minimum feedback arc set, we obtain a maximum acyclic directed graph as we see in Figure 8. This gives us the final ordering of kindergartens: $k_3 \succ k_2 \succ k_1$.

This process would always give us always the same ranking regardless of the differences is distances d_1, d_2 or sibling status s. To obtain a probabilistic outcome we also model an error distribution for ε. We select the distribution for ε so that is would be almost uniformly random when result of the pairwise comparison is $\frac{1}{2}$, but when the probability is really close to 0 or 1 then is would have almost no effect. We use the generalised error distribution with parameters as in equation (4).

$$\varepsilon \sim GN(\mu = 0, \alpha = 0.4, \beta = 8) \tag{4}$$

Figure 8: Maximum acyclic graph for full ordering



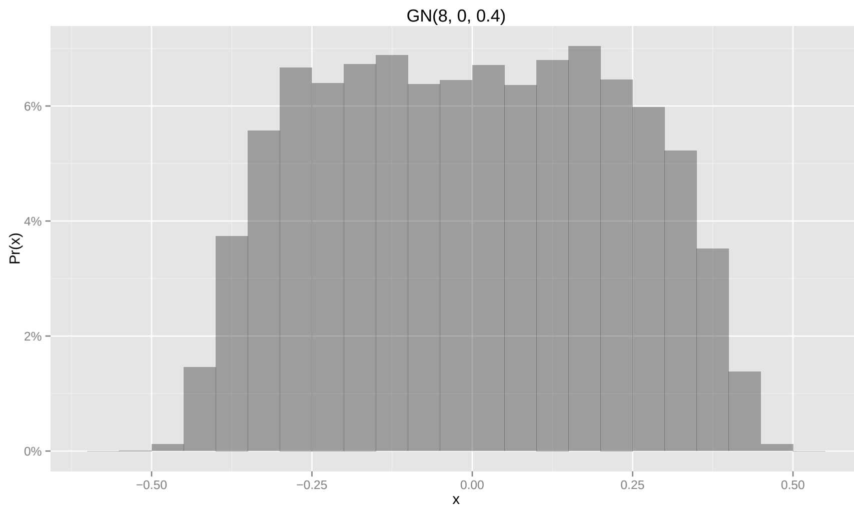


Figure 9: Generalised error distribution

The interpretation of parameters of (4) is intuitive. When $\beta \to \infty$ then it is a uniform distribution $\mathcal{U}(-0.4, 0.4)$, which in the case of two equal kindergartens, where $F(k_1, k_2) = \frac{1}{2}$, is $\mathcal{U}(0.1, 0.9)$. We show an example of a generalised error distribution in Figure 9.

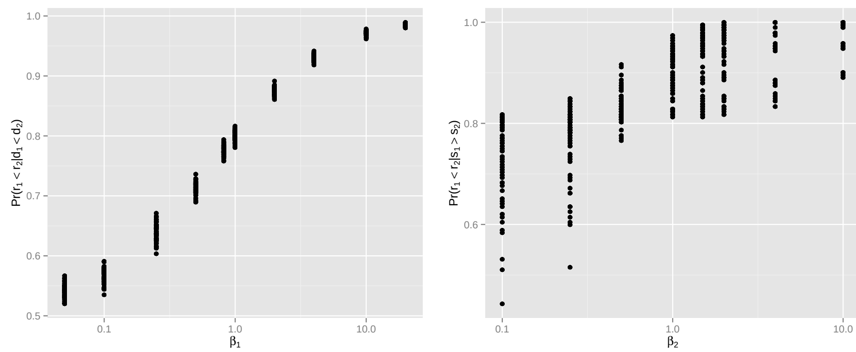
Also for consistency for each family i and every pair of kindergartens $k_{i,a}$ and $k_{i,b}$ we draw a single value from distribution (4), meaning that $\varepsilon(k_{i,a}, k_{i,b}) = \varepsilon(k_{i,b}, k_{i,a})$.

For this we look at several parametrise preferences same as in equation (3), with parameter values as in (5). For each combination of parameters we generate several (7) different preference profiles and evaluate the policies on the average over all the preference profiles.

$$\beta_1 \in \{0.05, 0.1, 0.25, 0.5, 0.82, 1, 2, 4, 10, 20\}, \beta_2 \in \{0.1, 0.25, 0.5, 1, 1.5, 2, 4, 10\} \tag{5}$$

To better interpret the results we look at the results by conditional probabilities of a parameter set. We look at two conditional effects: (a) probability of ranking kindergarten higher given it is closer; and (b)probability of ranking a kindergarten higher given a kindergarten has a sibling. Formally the conditional probabilities are defined in equations (6) and in (7).

$$\Pr(r_1 < r_2 \mid d_1 < d_2) = \frac{\Pr(d_1 < d_2, r_1 < r_2)}{\Pr(d_1 < d_2)} \tag{6}$$



(a) Conditional probability of β_1 (b) PConditional probability of β_2

Figure 10: Coefficients and conditional probabilities

$$\Pr(r_1 < r_2 \mid s_1 > s_2) = \frac{\Pr(s_1 > s_2, r_1 < r_2)}{\Pr(s_1 > s_2)} \tag{7}$$

The mean conditional probabilities with fitted regression parameters, $\beta_1 = 0.82$ and $\beta_1 = 1$, are respectively $\Pr(r_1 < r_2 \mid d_1 < d_2) \approx 0.77$ and $\Pr(r_1 < r_2 \mid s_1 > s_2) = 0.93$, with full preference lists. With limited (up-to 3) kindergartens, based on original submissions not rankings, the conditional probabilities are respectively on average 0.75 and 0.83. The conditional probabilities in the original rankings are respectively $\Pr(r_1 < r_2 \mid d_1 < d_2) \approx 0.7$ and $\Pr(r_1 < r_2 \mid s_1 > s_2) = 0.83$. So the statistical properties in terms of conditional effects is retained in the data. In Figures 10b and 10a we show the relationship between the logistic parameters and the conditional probabilities.

We show the relation between the conditional probabilities and coefficients in Figures 10a and 10b.