Designing the Menu of Licenses for Foster Care *

Altınok, Ahmet † MacDonald, Diana‡

December 2024

Abstract

This paper explores US foster parent licensing, essential for placing foster children. We develop a theoretical matching model to study the optimal menu of licenses designed to screen foster parents, considering heterogeneous agents, adverse selection, and search frictions. Our findings highlight the following: **(i)** optimal allocation calls for a segregation of the market, **(ii)** simple transfer schedules achieve the purpose, **(iii)** complementarities do not ensure Positive Assortative Matching (PAM) in equilibrium. We provide conditions that guarantee PAM. Our results suggest that the current licensing menu partly aligns with optimal solutions but may fall short in screening.

***JEL* Classification:** C78, D47, D82

Keywords : Adverse Selection · Foster Care · Matching · Sorting · Search

*This work benefited from helpful suggestions and discussions with Marek Weretka, and participants at seminars at ACM Conference on Equity and Access in Algorithms, Mechanisms and Optimization, ASSET Annual Meeting, ASU PhD Reunion Conference, Banco de México, CIDE, Conference on Mechanism and Institution Design, EPCS Annual Meeting, ITAM, LACEA Annual Meeting, Nazarbayev University, SING Annual Meeting, SSCW Annual Meeting, Stony Brook International Conference on Game Theory, Universidad de Chile, and Universidad de Guadalajara. The views and conclusions presented in this paper are exclusively of the authors and do not necessarily reflect those of Banco de México. All errors are our own.

†Department of Economics, Nazarbayev University, ahmet.altinok@nu.edu.kz

‡Research Department, Banco de México, diana.macdonald@banxico.org.mx

1 Introduction

Foster care can be viewed as a two-sided matching market with heterogeneous children and parents, where foster parents have preferences over children, and child welfare agencies have preferences over foster parents (on behalf of children).[1] As in many other markets, matches form in the presence of *private information*, since a foster parent's ability to provide care for a child is unknown to the child welfare agency. Aiming at solving for this *adverse selection problem*, a menu of licenses is offered to foster parents. In practice, a license specifies the type of child a parent can foster and the corresponding transfer received by foster parents. Furthermore, as a rule of thumb, children are grouped by the level of care needed, and transfers vary across groups. For example, foster parents in Arizona can choose between two licenses: traditional and therapeutic. In the former, foster parents can only foster children with standard needs, whereas in the latter foster parents can foster children with standard needs and also children with special needs. Parents receive US$20.80 per day for children with standard needs, and US$36.87 for children with special needs. These transfers are based only on the estimated cost of providing care for a child, and do not depend on any other characteristic of the market. This raises the question of whether the current menu of licenses can achieve its screening objective, and more importantly, whether the current mechanism used in the system is optimal.

This paper develops a theoretical matching model to study the optimal menu of licenses designed to screen foster parents in the US foster care system. We construct a two-sided matching model with heterogeneous agents (children differ in the level of care needed and parents differ in their ability to provide care), private information on a parent's attribute, and a designer who coordinates match formation through a menu of contracts.[2] The main innovation of our paper lies in introducing an endogenous search friction that varies with market size, an element not entirely within the designer's control. The analysis focuses on incentive-compatible licenses, which specify an allocation of parents across submarkets of children and the corresponding transfers, and the sorting patterns that might arise in equilibrium.

Our results suggests that the menu of licenses used in practice exhibits some of the properties of the optimal solution. First, we find that it is never optimal to randomly match all types of parents to all types of children, that is, optimal allo-

[1] See Appendix A for a detailed description of foster care in the US.

[2] Our environment can be used to analyze a matching problem between adoptive children and prospective adoptive parents, provided that the designer offers a subsidy to the adoptive parents.

cation calls for a segregation of the market. Second, we show that a simple transfer schedule achieves the purpose, that is, parents holding different licenses and providing care for the same type of child can receive the same monetary transfer. However, the transfers must not only account for the child's attribute as in practice, but also for other features of the market such as the distributions of agents. Lastly, we find that complementarity in child's and parent's attributes is not sufficient to ensure that Positive Assortative Matching (PAM) will arise in equilibrium. Thus, we provide sufficient conditions for the equilibrium sorting to exhibit PAM: either a *stronger* complementarity determined by the distribution of children' attributes, or a lower bound on the share of children with special needs.[3]

The model is as follows. There are two sides of the market populated by a continuum of agents: children and parents. Children are heterogeneous in the level of care needed, low- (x_1) or high-needs (x_2); and parents are heterogeneous in their ability to provide care, low-(y_1) or high-ability (y_2). We start our analysis with this binary type space for parents, which we later extend to a continuum of types. A child's attribute is common knowledge, and a parent's attribute is private information. We assume that children receive greater payoffs when matched than unmatched, and parents incur a cost when a match forms. The designer maximizes expected utility from children minus transfers to parents. We assume that the surplus of each match is nonnegative, thus profitable.[4] As in practice, we construct submarkets for each child's attribute, that is, there is a submarket populated by low-needs children and another submarket populated by high-needs children.

First, the designer announces and commits to a menu of licenses. A license specifies: **(1)** a randomization rule that determines the probability with which a parent is allocated into each submarket, and **(2)** a corresponding transfer when a match forms. After observing the menu, each parent chooses a license. Next, the randomization device is realized and parents are allocated across submarkets[5] determining endogenously the parents-to-children ratio (market tightness) for each submarket. Lastly, within each submarket, meetings take place, matches are formed, and transfers occur. We introduce a search friction by assuming that meetings are not certain, that is, the probability of a child (parent) meeting a parent (child) is represented by a meeting technology which is a function of the mar-

[3] Ideally, we would empirically test our theoretical predictions on optimal sorting patterns, but this is not feasible due to limited data on parents' attributes.

[4] In our framework, surplus of a match is a cost-net benefit function whose argument are parent's ability and child's level of care needed.

[5] We use the language of allocation across submarkets, but it can also be interpreted as the weight the designer assigns to a specific parent holding a particular license to provide care for one type of child or another.

ket tightness. Thus, when parents choose a license that guarantees allocation to a specific submarket with probability one, they are certain about the type of child they will be matched with but remain uncertain about whether the match will occur. However, if a parent selects a license that allocates them to either submarket with a strictly positive probability, they face uncertainty not only about the type of child they will be matched with but also whether a match will take place.

Our search friction assumption is motivated by the fact that child welfare agencies do not act as matchmakers but instead define feasible matches through a menu of licenses and guidelines. In practice, social workers are responsible for contacting parents about a specific child in a decentralized manner. Thus, the randomization mechanism can be interpreted as a guideline: for example, if a parent is deemed 'better' suited to care for low-needs children than high-needs children, the system would aim to allocate that parent to the first submarket with a higher probability. Furthermore, market tightness captures the level of congestion in the market, while the meeting technology accounts for the frictions arising from the decentralized nature of the matching process which strongly depends on the congestion.

It is important to emphasize that the search friction assumption is a crucial element of our model, as it introduces non-trivial effects on the analysis.[6] Specifically, when a mass of type-y parents is reallocated from one submarket to another, three key effects occur: **(i)** *Surplus Effect:* This represents the change in total expected surplus of the market. **(ii)** *Congestion Effect:* The change in market tightness in the submarket where parents are reallocated, leading to a thicker market. **(iii)** *Decongestion Effect:* The change in market tightness in the submarket from which parents are reallocated, resulting in a thinner market. These effects not only add complexity to the analysis but also enrich the predictions of our model.

We begin by examining the case with complete information and derive results for both super- and sub-modular surplus functions. In this section, we focus on the case of super-modularity, while the discussion of sub-modularity is deferred to the main body of the paper.

First, we find that it is never optimal for the designer to allocate type-y parents to both submarkets x_1 and x_2 with strictly positive probability, regardless of

[6]Our paper is closely related to the work of Damiano and Li (2007), which explores how a monopoly matchmaker sorts agents into exclusive meeting places for random pairwise matching. A key assumption in their framework is that all agents have a constant match probability of one, abstracting from size effects that could influence matching probabilities based on market scale. In contrast, our paper incorporates market size effects through the search friction assumption, illustrating how market size influences matching probabilities.

whether the surplus function is supermodular or submodular.[7] This result rationalizes the nested nature of the licenses used in practice, such as the case of the state of Arizona described above.

Second, we show that super-modularity is neither sufficient nor necessary for the optimal sorting to exhibit PAM. In our framework, the randomization device establishes who can match with whom in the market so we use it to define sorting patterns: a sorting exhibits PAM (NAM) if y_2-parents are allocated to submarket x_2 with a greater (smaller) probability than y_1-parents are.[8] For a frictionless environment with a super-modular surplus function, it is well known that matching agents in a positive assortative way maximizes total welfare (Becker, 1973). However, when search frictions are introduced, we find that this result does not hold because the expected total welfare, calculated using the meeting technologies in each submarket, is not necessarily super-modular even if the surplus function is super-modular. By imposing a lower bound on the fraction of type-x_2 children along with super-modularity, we can ensure that PAM arises in equilibrium. Intuitively, type-y_2 parents are more desirable in any submarket, thus the designer would like to allocate them to a more profitable and thicker submarket x_2. Thus, by imposing a lower bound on the share of type-x_2 children we ensure that the market is thick enough.

Third, we find that any transfer scheme that is on the participation constraint for each type of parent is optimal, and it does not affect the equilibrium sorting. Therefore, our framework predicts the same equilibrium sorting regardless of interim or ex-post participation constraints. This is intuitive as, in equilibrium, given a license, parents only care about the expected transfer that equalizes the expected cost. Moreover, the optimal transfers must account for the child's attribute, and other features of the market such as number of children and number of parents.

In this context, one might imagine that the child welfare agency could screen foster parents using observable characteristics such as race, marital status, educational level, employment status, or income. Under this scenario, our complete information analysis would be sufficient. However, the literature suggests that observable characteristics of foster parents do not predict the likelihood of fostering higher-needs children, yet the type of license they hold does.[9] This motivates our

[7]In other words, if the optimal randomization rule is interior for type-y parents, then it is a corner solution for type-y' parents, where y and y' are distinct.

[8]One can equivalently define the sorting pattern through a matching correspondence as standard in the literature, and say that a sorting exhibits PAM if the matching correspondence is a lattice as in Shimer and Smith (2000). Since the randomization device provides more information than the correspondence, our sorting notion is more general: any feasible-unequal allocation of parents in our setting exhibits either PAM or NAM, but not both, unlike Shimer and Smith (2000).

[9]Using a sample of 297 foster mothers and a linear multiple regression analyses, Cox et al.

next analysis relaxing the assumption over the observability of a parent's attribute.

With private information, our results from complete information carry on, except for an additional condition for PAM and the need for the designer to provide informational rents to high-ability parents. Due to the greater expected cost for low-ability parents to provide care, the expected transfer they receive is greater than what high-ability parents receive given the first-best menu of licenses. As a result, high-ability parents have incentives to mimic low-ability parents, thus the designer pays information rent to high-ability parents to eliminate such incentives.

Now, to determine the optimal sorting, one needs to know the cost of a parent-child pairing, as well as the parent distribution, which need not be known under complete information.[10] In this case, a super-modular cost function increases the forces for the equilibrium sorting to be NAM. The intuition is as follows: a super-modular cost function means that the difference between the cost for low-ability parents and high-ability parents of taking care of a child with low-needs is *greater* than the difference of providing care for a child with high-needs. Thus, it would be more expensive to shut down a deviation by high-type parents from high-needs children to low-needs children than a deviation from low-needs to high-needs children. As a result, the designer would pay less information rent if high-ability parents are allocated into the submarket of children with low-needs. Therefore, to ensure that PAM emerges in equilibrium and that parents reveal their type truthfully, we impose sub-modularity on the cost function and establish a lower bound on the proportion of type-x_2 children.

Lastly, we present two significant extensions of our complete information model. First, we retain the classification of children into high- and low-need categories while expanding the parental attribute space to a continuum. We find that the nested hierarchy of licenses observed in the two-type case no longer holds. Instead, parents are allocated into a single submarket. Moreover, the optimal payment now exclusively considers the cost of providing care, aligning more closely with practical scenarios. Second, we conduct comparative statics on the meeting technology to assess the robustness of our results, considering potential variations in the decentralized search process across different States within the United States.

(2011) found no significant association between foster mothers' observable characteristics—such as race, marital status, education level, and income—and the likelihood of fostering children with emotional and behavioral problems.

[10] Knowing the surplus of a match is sufficient to determine the equilibrium licenses under complete information, we do not need to disentangle utility and the cost to determine the optimal sorting. This is not the case in the presence of information friction.

Literature Review. The main contribution of this paper is to develop a theoretical matching model with adverse selection and search frictions to study the optimal menu of licenses in the US foster care system. Previous studies have analyzed foster care as a matching market but with different focuses. For instance, Slaugh et al. (2015) assesses the Pennsylvania Adoption Exchange program, recommending improvements for adoption outcomes, while Robinson-Cortés (2019) uses a confidential dataset to analyze child placements and evaluate policy interventions. Olberg et al. (2021) introduces a dynamic search and matching model with observable attributes to compare search processes used by child welfare agencies, and MacDonald (2022) offers an empirical analysis of match transitions, introducing a model with reversible (foster) and irreversible (adoption) matches. Our paper stands out by incorporating a menu of licenses into the analysis, providing a tailored model that addresses key features of foster care, particularly under conditions of information frictions.

This paper connects to the literature on assortative matching under asymmetric information, specifically within principal-agent frameworks with adverse selection. Previous studies, such as Ghatak (1999), Van Tassel (1999), Ghatak (2000), Guttman (2008) and Altinok (2023), examine sorting patterns in microfinance loan contracts where heterogeneous borrowers are optimally paired. Similar to this paper, lenders in these models can induce PAM or NAM. However, unlike our paper, these studies focus on single-sided markets, without accounting for search frictions or the presence of a match coordinator.

Lastly, our paper relates to the search and matching literature, drawing from frameworks like those of Menzio and Shi (2010a) and Menzio and Shi (2010b) which introduce submarkets, directed search, and market tightness in labor markets. Shi (2001) first demonstrated that super-modularity alone is insufficient for PAM under specific directed search technologies, while Eeckhout and Kircher (2010) identified stronger complementarity conditions necessary for PAM. Our model differs by incorporating private information and transfers from the designer to one side of the market. Additionally, Shimer and Smith (2000) and Smith (2006) explored two-sided matching with random search and complete information, showing that PAM fails unless stronger complementarity conditions, like log-supermodularity, are met. Consistent with this literature, we demonstrate that in the presence of search frictions in foster care, stronger complementarities are needed to achieve PAM. These conditions become even more stringent when information frictions are incorporated into the model.

Organization of the Paper. The rest of the paper is organized as follows. Section 2 introduces the model. Section 3 presents the analysis for the complete information case, and Section 4 widens the analysis to the case of private information. In Section 5, we provide two relevant extensions. Lastly, Section 6 concludes. Appendix A presents an overview of foster care in the US, while Appendix B illustrates key examples. All omitted proofs can be found in Appendices C, D, and E.

2 Model

One side of the market is populated by a continuum of **children** who differ in an observable attribute $x \in X = \{x_1, x_2\}$ where x_1 denotes a low-needs child (without a disability), x_2 denotes a high-needs child (with a disability), and $x_2 > x_1$.The fraction of children with low-needs is $f(x_1) \in [0, 1]$, whereas the fraction with high-needs is $f(x_2) = 1 - f(x_1)$. We refer to the set of children with attribute x as *submarket* x. The other side of the market is constituted by a continuum of **parents** who are heterogeneous in their ability to provide care for a child. In particular, y_1 denotes parents with low-ability, y_2 denotes parents with high-ability, and $y_2 > y_1$.[11] The fraction of parents with low-ability is $g(y_1) \in [0, 1]$, and that with high-ability is $g(y_2) = 1 - g(y_1)$. A parent's ability to provide care is *private information*.

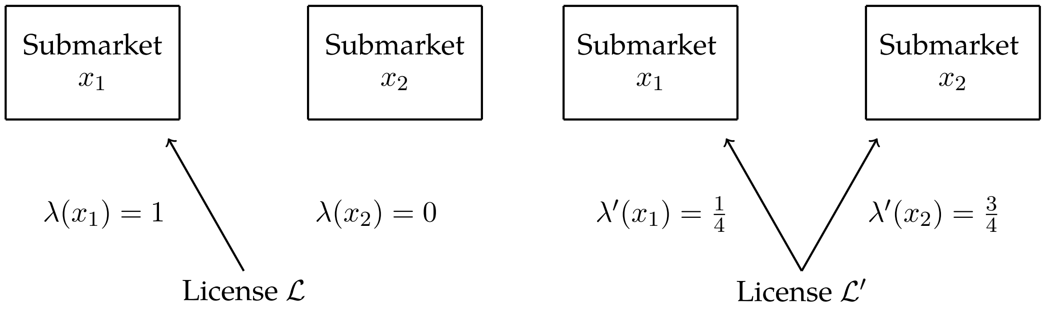
Matches are formed between children and parents on a one-to-one basis.[12] There is a **designer** who facilitates the matching process by offering a menu of licenses to parents. A license $\mathcal{L}$ is represented by a pair (λ, τ) where $\lambda : X \to [0, 1]$ is a randomization device that determines the probability with which a parent is allocated to submarket x, and $\tau : X \to \mathbb{R}$ represents a transfer between the designer and the parent if the parent forms a match with child x.[13] Throughout the paper, we restrict attention to the menu of licenses with the following features: (i) allocations are non-wasteful, that is, $\sum_{x \in X} \lambda(x) = 1$, and (ii) parents have limited liability, that is, $\tau(x) \geq 0$ for any $x \in X$.

Figure 1 represents two examples of randomization devices under separate licenses. Parents holding license $\mathcal{L}$ are allocated to submarket x_1 with probability 1, and to submarket x_2 with probability 0. Analogously, parents holding license $\mathcal{L}'$ are in submarkets x_1 and x_2 with probabilities $1/4$ and $3/4$, respectively.

[11] Section 5 expands the type space of parents to a continuum.

[12] According to Gibbs and Wildfire (2007), the average occupancy rate is 1.5 children per home, indicating that assuming one-to-one matches aligns with the empirical evidence.

[13] Alternatively, $\lambda(x)$ can be interpreted as the probability with which a parent is considered to provide care for a type-x child.



(a) **(b)**

Figure 1: Examples of Randomization Devices

All agents are risk-neutral. The designer maximizes children's welfare net of transfers. Payoffs for unmatched agents are normalized to zero. When a child x and a parent y form a match, the child receives payoffs according to a real-valued function $u(x, y)$, and the parent incurs a cost of providing care according to a real-valued function $c(x, y)$.[14]

Assumption 1. *(a) for all (x, y), $u(x, y) \geq 0$, $c(x, y) \geq 0$ and $u(x, y) - c(x, y) \geq 0$, (b) $u(x, y)$ is increasing in y, and (c) $c(x, y)$ is increasing in x and decreasing in y.*

Assumption 1(a) reflects the following: children are better-off placed with a foster parent than being unmatched, parents incur a cost when providing care for a child, and all matches are profitable. Assumption 1(b) states that children prefer high- to low-ability parents. Finally, Assumption 1(c) implies that parents incur in a smaller cost when matched to low-needs children than to high-needs children, and high-ability parents incur in a smaller cost when providing care than low-ability parents.

Timing is as follows:

1. First, the designer announces and commits to a menu of licenses. By the revelation principle, we restrict attention to direct revelation mechanisms. Thus, without loss of generality, we consider menus with two licenses, one for each type of parent $\{\mathcal{L}^k\}_{k=1}^2 \equiv \left\{ \{ (\lambda^k(x_i), \tau^k(x_i)) \}_{i=1}^2 \right\}_{k=1}^2$.
2. After observing the menu, each parent chooses a license, where $\sigma^y \in \{\mathcal{L}^1, \mathcal{L}^2\}$ denotes this decision. Then, the allocation of parents $\left\{ \{\lambda^k(x_i)\}_{i=1}^2 \right\}_{k=1}^2$ across submarkets is realized.[15]

[14] The parents' cost function can be interpreted as a net cost function, which captures the balance between the benefits of providing care for a child and the associated costs.

[15] Note that, if a license features an interior randomization device, parents who select it have a strictly positive probability of being matched with either type of child before the outcome is realized—that is, at the time the license is chosen.

3. Next, within each submarket, children and parents meet stochastically. The meeting technology can be described in terms of the parents-to-children ratio (*market tightness*). The market tightness of each submarket $x \in X$, denoted by θ_x, is equal to:

$$\theta_x = \frac{\sum_{k=1}^{2} h^k(y_1)\lambda^k(x) + h^k(y_2)\lambda^k(x)}{f(x)}$$

where $h^k(y)$ denotes the endogenous mass of parents $y \in \{y_1, y_2\}$ choosing license k. A child x meets a parent according to a *meeting technology* $\pi^c(\theta_x)$ where $\pi^c : \mathbb{R}_+ \to [0,1]$ is a strictly increasing and strictly concave function such that $\pi^c(0) = 0$. Similarly, a parent meets a child x with probability $\pi^p(\theta_x)$ where $\pi^p : \mathbb{R}_+ \to [0,1]$ is a strictly decreasing and convex function such that $\pi^p(\theta_x) = \frac{\pi^c(\theta_x)}{\theta_x}$ [16] and $\pi^p(0) = 1$.[17]

4. Finally, when a child x and a parent y meet, a match (x, y) is formed and transfers take place according to $\left\{\{\tau^k(x_i)\}_{i=1}^2\right\}_{k=1}^2$.

Designer's Problem: The designer aims to maximize children's welfare while minimizing the transfers. We start by specifying the objective function of the designer. Let $\mathcal{L} \equiv \left\{\left\{\left(\lambda^k(x_i), \tau^k(x_i)\right)\right\}_{i=1}^2\right\}_{k=1}^2$ be an arbitrary menu of licenses. A child x receives utility $u(x, y_j)$ when she matches with a parent y_j. Notice that parent y_j might hold either contract, thus the net utility when a child x matches with parent y_j under contract k is $u(x, y_j) - \tau^k(x)$. Now, conditional on a meeting taking place, the probability that child x has met a parent y_j holding license k is equal to:

$$\frac{\lambda^k(x)h^k(y_j)}{\sum\limits_{k=1}^{2}\left[\lambda^k(x)\sum\limits_{j=1}^{2} h^k(y_j)\right]}$$

Thus, the net expected utility in each submarket x, conditional on a meeting taking place, is:

$$W(x) = \frac{\sum\limits_{k=1}^{2}\left[\sum\limits_{j=1}^{2}\left[u(x, y_j) - \tau^k(x)\right] \cdot \lambda^k(x) \cdot h^k(y_j)\right]}{\sum\limits_{k=1}^{2}\lambda^k(x) \cdot \left[\sum\limits_{j=1}^{2} h^k(y_j)\right]}.$$

[16] This relationship ensures that the probability of a child meeting a parent is consistent with the probability of a parent meeting a child, by equating the expected number of parents' meetings to children's.

[17] The search friction assumption highlights the decentralized matching process in the U.S. foster care system, where market congestion plays a key role. As market tightness increases, parents are less likely to find children, while children are more likely to find parents. These frictions also account for the possibility that congestion may prevent some matches from forming.

Then, the designer's problem is:

$$\max_{\left\{\left\{\left(\lambda^k(x_i),\tau^k(x_i)\right)\right\}_{i=1}^{2}\right\}_{k=1}^{2}} \left\{\sum_{i=1}^{2} \pi^c\big(\theta_{x_i}\big)\ W(x_i)\ f(x_i)\right\} \tag{1}$$

subject to:

$$[\text{FC}] \quad \tau^k(x) \geq 0 \text{ and } \lambda^k(x) \geq 0 \text{ for all } (k,x), \text{ and} \sum_{i=1}^{2} \lambda^k(x_i) = 1 \text{ for all } k = 1, 2.$$

$$[\text{MT}] \quad \theta_x = \frac{1}{f(x)} \cdot \sum_{k=1}^{2} \left[\lambda^k(x) \sum_{j=1}^{2} h^k(y_j)\right], \text{ for all } x.$$

$$[\text{PC}] \quad \sum_{i=1}^{2} \Big[\tau^k(x_i) - c(x_i, y_k)\Big] \lambda^k(x_i) \pi^p(\theta_{x_i}) \geq 0, \text{ for all } k = 1, 2.$$

$$[\text{IC}] \quad \sum_{i=1}^{2} \left[\tau^k(x_i) - c(x_i,y_k)\right] \lambda^k(x_i) \pi^p(\theta_{x_i}) \geq \sum_{i=1}^{2} \left[\tau^{k'}(x_i) - c(x_i,y_k)\right] \lambda^{k'}(x_i) \pi^p(\theta_{x_i}),$$
$$\text{for all } k, k' = 1, 2$$

where [FC] are the feasibility constraints specifying restrictions over each $\lambda^k(x)$ and $\tau^k(x)$. The restrictions [MT] corresponds to the market tightness (parents-to-children ratio) in each submarket. [PC] are the participation constraints guaranteeing that each parent y_j receives a higher expected payoff when holding license $k = j$ than when unmatched. Lastly, [IC] are the incentive compatibility constraints that ensures that our equilibria are truth-telling.

2.1 Definition of Sorting Patterns

Next, we define a matching correspondence and establish sorting patterns using the randomization device of each license $\left\{\lambda^1(x_i), \lambda^2(x_i)\right\}_{i=1}^{2}$.

Definition 1. *A* ***matching correspondence*** *is a map $\mu : Y \mapsto X$ such that $x \in \mu(y_{\mathbf{k}})$ if and only if $\lambda^{\mathbf{k}}(x) > 0$. Moreover, if $\lambda^2(x_2) \geq \lambda^1(x_2)$ then the sorting exhibits* ***Positive Assortative Matching*** *(**PAM**). Analogously, if $\lambda^2(x_2) \leq \lambda^1(x_2)$ then the sorting exhibits* ***Negative Assortative Matching*** *(**NAM**).*

We are interested not only in establishing properties that ensures monotone sorting but also in characterizing the optimal menu of licenses offered by the designer. As a result, our notion of monotone sorting is as follows: We say high-type PAM (NAM) if type-y_2 parents are allocated into both submarket while type-y_1

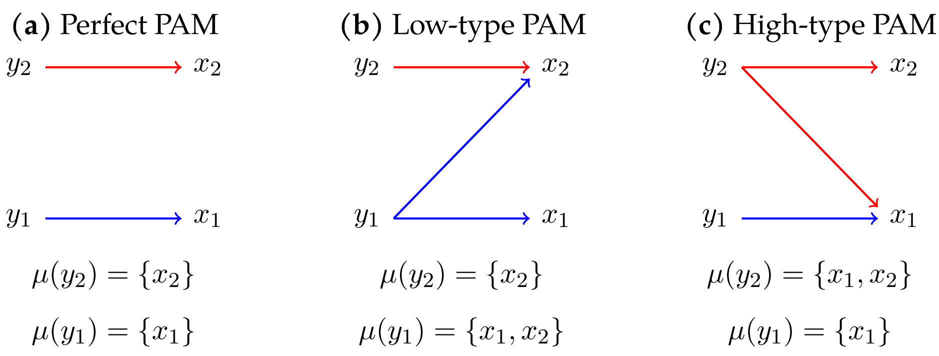


Figure 2: Examples of Positive Assortative Matching (PAM)

parents are allocated only into submarket x_1 (x_2). Analogously, low- type PAM (NAM) follows.

Figure 2 presents examples illustrating our concept of monotone sorting patterns. In Panel 2a, y_2-parents are allocated into submarket x_2 with probability one and y_1-parents are allocated into submarket x_2 with probability zero, thus it follows that $1 = \lambda^2(x_2) \geq \lambda^1(x_2) = 0$. In Panel 2b, y_2-parents are allocated into submarket x_2 with probability one and y_1-parents are allocated into both submarkets with strictly positive probability, thus $1 = \lambda^2(x_2) \geq \lambda^1(x_2) \in (0, 1)$. Lastly, in Panel 2c, y_2-parents are allocated into both submarkets with strictly positive probability and y_1-parents are allocated into submarket x_2 with probability zero, thus $\lambda^2(x_2) \in (0, 1) \geq \lambda^1(x_2) = 0$.

Lastly, note that the randomization device in Panel 2c can represent the menu of licenses used in practice, as outlined in the introduction. In this case, parents holding license 1 are restricted to fostering only low-needs children i.e. $\mu(y_1) = \{x_1\}$, while parents with license 2 are eligible to foster both types of children i.e. $\mu(y_2) = \{x_1, x_2\}$, reflecting the nested structure highlighted previously.

3 Equilibrium Analysis: Complete Information

In this section, we examine the optimal menu of licenses and analyze sorting patterns that might arise in equilibrium under complete information. We focus on symmetric equilibria where same type parents choose the same license. First, note that by incorporating the [PC] constraints into the objective function in Equation

1, reduces the designer's problem to:

$$\max_{\{\lambda^k(x_1),\lambda^k(x_2)\}_{k=1}^2} \left\{ \sum_{i=1}^{2} \pi^p(\theta_{x_i}) \left[\sum_{k=1}^{2} \Big(\underbrace{u(x_i,y_k) - c(x_i,y_k)}_{S(x,y)} \Big) \lambda^k(x_i)\ g(y_k) \right] \right\} \tag{2}$$

subject to [FC] and [MT]. For notational ease, from now on, let θ_1 and θ_2 denote θ_{x_1} and θ_{x_2}, respectively. In addition, let $S(x,y) \equiv u(x,y) - c(x,y)$ denote the **surplus of a match** (x,y) which is increasing in y by Assumption 1.

Lemma 1. *For at least one of the licenses, the optimal randomization rule (allocation) yields a corner solution.*

Proof. See Appendix C.1. □

Lemma 1 states that it is never optimal for the designer to allocate both types of y-parents with strictly positive probability into submarkets x_1 and x_2. To prove Lemma 1, we start by assuming that the designer allocates both types of y-parents into both submarkets with strictly positive probabilities. We use the fact that the market tightness derived from any interior $\big(\lambda^1(x_1), \lambda^2(x_1)\big)$ can be achieved by *any* allocation on a line passing through $\big(\lambda^1(x_1), \lambda^2(x_1)\big)$. Now, since the meeting probabilities (i.e. market tightness) along that line are constant, we show that the designer can always increase the welfare by moving along the line towards the corners. Intuitively, given a submarket, if using one type of parent is more profitable than using the other, then the designer will allocate the entire population of more profitable parents into that submarket. This result speaks to the optimality of the nested hierarchy property exhibited in the licenses used in practice. That is, one license allocates parents into only one submarket, while the other license allocates parents into both submarkets.

Now, to characterize the optimal randomization rule we follow a nonstandard technique due to the presence of corner solutions. We start with an arbitrary interior allocation and examine whether the designer can increase total expected welfare by simply reallocating parents across submarkets. Formally, for each (x,k), let $\lambda^k(x)$ be an arbitrary-feasible interior probability that generates a total welfare equal to:

$$\begin{aligned} W\big(\lambda^1(x_1), \lambda^2(x_1)\big) &= \pi^p(\theta_1) \cdot \Big[g(y_1)\ \lambda^1(x_1)\ S(x_1,y_1) + \big(1-g(y_1)\big)\ \lambda^2(x_1)\ S(x_1,y_2) \Big] \\ &\quad + \pi^p(\theta_2) \cdot \Big[g(y_1)\ \big(1-\lambda^1(x_1)\big)\ S(x_2,y_1) + \big(1-g(y_1)\big)\ \big(1-\lambda^2(x_1)\big)\ S(x_2,y_2) \Big] \end{aligned}$$

After trembling $\lambda^1(x_1)$ by ε_1 and $\lambda^2(x_1)$ by ε_2 such that $\varepsilon_2 \equiv -\frac{\varepsilon_1 g(y_1)}{1-g(y_1)}$, ensuring that the market tightness remains constant, the change in welfare is equal to:

$$\begin{aligned}\Delta_W &= W\big(\lambda^1(x_1)+\varepsilon_1, \lambda^2(x_1)+\varepsilon_2\big) - W\big(\lambda^1(x_1), \lambda^2(x_1)\big) \\ &= \varepsilon_1\, g(y_1)\, \underbrace{\Big(\pi^p(\theta_2)\, \big[S(x_2,y_2)-S(x_2,y_1)\big] - \pi^p(\theta_1)\, \big[S(x_1,y_2)-S(x_1,y_1)\big]\Big)}_{Z^{CI}(\theta_1)}\end{aligned}$$

Note that, $\theta_2 = \frac{1-f(x_1)\theta_1}{1-f(x_1)}$, thus $Z^{CI}(\cdot)$ can be written as a function of only θ_1. From the change in welfare, it is easy to see that the designer can always increase total welfare by changing $\big(\lambda^1(x_1), \lambda^2(x_1)\big)$ such that the market tightness remains constant. The optimal allocation of parents can be characterized by $Z^{CI}(\theta_1)$, which represents the expected difference in gains between children x_2 and x_1 from being matched to a high-ability parent as opposed to a low-ability parent. Moreover, the sign of $Z^{CI}(\theta_1)$ determines the equilibrium sorting. Let $\bar{\theta}_1$ be such that $Z^{CI}(\bar{\theta}_1) = 0$, then the following result holds:

Proposition 1. *Let θ_1^* be the equilibrium market tightness.* ***(i)*** *If $\theta_1^* > \bar{\theta}_1$ then the equilibrium sorting exhibits PAM.* ***(ii)*** *If $\theta_1^* < \bar{\theta}_1$ then the equilibrium sorting exhibits NAM.* ***(iii)*** *$\theta_1^* = \bar{\theta}_1$ is never optimal.*

Proof. See Appendix C.2. □

Proposition 1 states that if the equilibrium market tightness θ_1^* is such that $Z^{CI}(\theta_1^*)$ is positive then PAM arises in equilibrium. To see this, note that $Z^{CI}(\theta_1)$ is increasing in θ_1, that is, the change in welfare increases as θ_1 increases. Thus, for any $\theta_1 > \bar{\theta}_1$ it follows that $Z^{CI}(\theta_1)$ is positive. Therefore, when $Z^{CI}(\theta_1)$ is positive, we can pick $\varepsilon_1 > 0$, increasing the share of y_1-parents allocated in submarket x_1 and decreasing the share of y_2-parents allocated in submarket x_1, until either $\lambda^1(x_1) = 1$ or $\lambda^2(x_1) = 0$. Either way, we move in the direction of PAM. Intuitively, a high θ_1^* translates into a small probability of a parent meeting a child in submarket x_1. Since y_2-parents generate a greater surplus, it would be optimal to minimize the probability with which they remain unmatched. Thus, the designer chooses to use y_2-parents in submarket x_2, leading to PAM. Analogously, NAM follows.

Figure 3 illustrates environments capturing Lemma 1 and Proposition 1. In each box, the x- and y-axis correspond to the probability with which parents holding license 1 and 2 are allocated into submarket x_1, respectively. Thus, every point inside the box $\big(\lambda^1(x_1), \lambda^2(x_1)\big)$ is a feasible allocation of parents. Yet, note that by Lemma 1 only the points at the borders can be an equilibrium. In addition, each black-dashed line corresponds to the values of $\big(\lambda^1(x_1), \lambda^2(x_1)\big)$ such

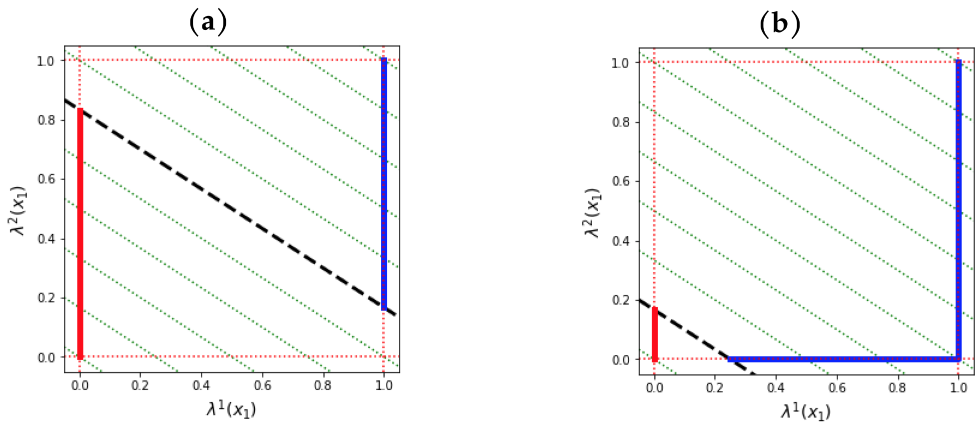


Figure 3: Illustration of PAM and NAM given $Z^{CI}(\overline{\theta}_1)$

that $Z^{CI}(\overline{\theta}_1) = 0$, each blue line shows the feasible allocations that can be an equilibrium when $Z^{CI}(\theta_1) > 0$ (above the black-dashed line), and each red line shows the feasible allocations that can be an equilibrium when $Z^{CI}(\theta_1) < 0$ (below the black-dashed line). In Panel 3a, the equilibrium candidates are along the vertical blue line and vertical red line. In the former, allocations are such that $\lambda^2(x_2) \geq \lambda^1(x_2) = 0$, which corresponds to high-type PAM. In the latter, allocations are such that $1 = \lambda^1(x_2) \geq \lambda^2(x_2)$, which corresponds to high-type NAM. Analogously, in Panel 3b, the equilibrium candidates are along the red and the blue lines.

Now, we are interested in establishing sufficient conditions for PAM and NAM to arise in equilibrium. Corollary 1 follows directly from Proposition 1.[18]

Corollary 1. ***(i)*** *If* $\frac{S(x_2,y_2)-S(x_2,y_1)}{S(x_1,y_2)-S(x_1,y_1)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)}$ *holds, then the equilibrium sorting exhibits PAM.* ***(ii)*** *If* $\frac{S(x_1,y_2)-S(x_1,y_1)}{S(x_2,y_2)-S(x_2,y_1)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_1)}\right)}$ *holds, then the equilibrium sorting exhibits NAM.*

Proof. See Appendix C.3. □

For Corollary 1(i), notice that $Z^{CI}(\theta_1)$ reaches its minimum value at $\theta_1 = 0$, implying that $\pi^p(0) = 1$ and $\theta_2 = \frac{1}{f(x_2)}$. Thus, we ensure PAM by imposing that the minimum value of $Z^{CI}(\theta_1)$ is positive. Observe that (i) requires a super-modular surplus function since the right-hand side is greater than 1. Moreover, the greater the left-hand side of (i) is, the stronger the super-modularity is. Thus, *strong* super-modularity on the surplus function dominates the adversary

[18]Corollary 1 ensures that PAM or NAM will arise in equilibrium, but it does not specify whether we will observe either low-type PAM (NAM), high-type PAM (NAM), or perfect PAM (NAM). See Appendix C.4 for a detailed characterization.

effect of the search friction, and becomes sufficient to induce PAM at the optimum.[19] Alternatively, one can think of the inequality (i) as a lower bound over the share of children with high-needs to ensure PAM in equilibrium. Intuitively, by imposing a lower bound on the share of type-x_2 children we ensure that the market is thick enough for the more desirable type-y_2 parents , that is, the probability of meeting a child in submarket x_2 is bounded below. This is in line with the literature in dynamic search and matching, which imposes stronger complementarity conditions to ensure that more desirable partners have incentives to wait for a more desirable partner from the other side of the market.[20] Similar arguments and intuition follows for (ii).

Figure 4 exhibits environments illustrating Corollary 1. In Panel 4a, the equilibrium sorting can only exhibit PAM, since $Z^{CI}(\bar{\theta}) = 0$ is located in the left-bottom corner. Analogously, in Panel 4b, the equilibrium sorting can only exhibit NAM, since $Z^{CI}(\bar{\theta}) = 0$ is located in the right-top corner.

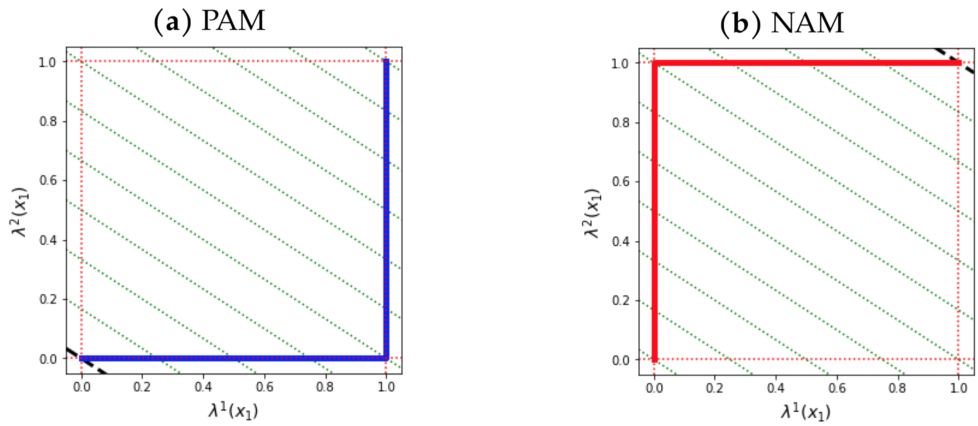


Figure 4: Illustration of Sufficient Conditions for Monotone Sorting

Next, we study the optimal transfer schemes. By fixing the optimal allocations

[19]By *strong* super-modularity on the surplus function we mean: $[S(x_2, y_2) - S(x_2, y_1)] \cdot \pi^p \left(\frac{1}{f(x_2)}\right) \geq S(x_1, y_2) - S(x_1, y_1)$ which introduces a constraint that is sensitive to the underlying distribution and the specific meeting technology. Similarly, the condition of *strong* sub-modularity on the surplus function follows.

[20]Shimer and Smith (2000) and Smith (2006) analyze a dynamic two-sided matching setting with heterogeneous agents, random search and complete information. The former paper assumes that utility is fully transferable and establishes as a sufficient condition not only supermodularity on the value of a match $f(x, y)$ where x and y are the agent's attributes, but also supermodulariy on $log\ f_x$ and $log\ f_{xy}$. The latter paper assumes that utility is strictly non-transferable and establishes as sufficient conditions monotonicity and log-supermodularity in $f(x, y)$. In both papers, these conditions ensure that, in the search process, high-partners do not settle for a low-partner but instead wait for the arrival of a high-partner. This is in the same spirit as our condition: we are also making sure that the payoffs received from matching high-types together compensate for the adversary effect of search frictions.

$\{\lambda^{k*}(x_1), \lambda^{k*}(x_2)\}_{k=1}^2$ from Equation 2, the designer solves the following:

$$\min_{\{\tau^k(x_1), \tau^k(x_2)\}_{k=1}^2} \left\{ \sum_{i=1}^{2} \pi^p\left(\theta_i^*\right) \sum_{k=1}^{2} \tau^k(x_i)\lambda^{k*}(x_i)\, g(y_k) \right\}$$

subject to [FC], [MT], and [PC] from Equation 1. The following proposition states that the transfer scheme is characterized by [PC]:

Proposition 2. *Given an equilibrium allocation of parents* $\{\lambda^{k*}(x_1), \lambda^{k*}(x_2)\}_{k=1}^2$*, any feasible transfer schedule for which the participation constraints hold with equality is an equilibrium.*

Proof. See Appendix C.5. □

Recall that the optimal allocation of at least one type of parent is a corner solution, in which case the transfer can be trivially pinned down. As an example, suppose that the equilibrium sorting exhibits perfect PAM, that is, y_1-parents are allocated into submarket x_1 with probability one, while y_2-parents are allocated into submarket x_2 with probability one. Then, the optimal transfer scheme is $\tau^{1*}(x_1) = c(x_1, y_1)$ and $\tau^{2*}(x_2) = c(x_2, y_2)$. That is, parents receive exactly the cost of providing care as is current practice.

In the case of an interior solution for at least one license, the optimal transfer scheme is not unique. As an example, suppose that the equilibrium sorting exhibits high-type PAM, that is, y_1-parents are allocated into submarket x_1 with probability one, while y_2-parents are allocated into both submarkets with strictly positive probability. Note, this is similar to the example of Arizona discussed in the introduction where low-needs children can be fostered by parents holding any of the two licenses, and high-needs children can only be fostered by parents holding one particular license. Here, the optimal transfer scheme is $\tau^{1*}(x_1) = c(x_1, y_1)$, $\tau^{2*}(x_1) \geq 0$ and $\tau^{2*}(x_2) = c(x_2, y_2) - [\tau^{2*}(x_1) - c(x_1, y_2)]\frac{\pi^p(\theta_1^*)\lambda^{2*}(x_1)}{\pi^p(\theta_2^*)\lambda^{2*}(x_2)}$. Now, as in practice, let's suppose that we include a restriction imposing that parents who provide care in the same market receive the same transfer, that is, $\tau^{1*}(x_1) = \tau^{2*}(x_1) = c(x_1, y_1)$. In this case, the optimal transfer for parent y_2 in submarket x_2 would be the following:

$$\tau^{2*}(x_2) = c(x_2, y_2) - \left[c(x_1, y_1) - c(x_1, y_2)\right]\frac{\pi^p(\theta_1^*)\lambda^{2*}(x_1)}{\pi^p(\theta_2^*)\lambda^{2*}(x_2)}$$

Remark 1. *A* ***simple transfer schedule*** *suffices; in equilibrium, parents holding different licenses and caring for the same type of child can receive the same transfer. This follows directly from the nested nature of the equilibrium allocation, meaning that markets are segregated. In contrast, for a fully interior allocation, such a simple transfer schedule would not simultaneously satisfy* *both [PC] conditions. Since the designer can select any randomization device, it is far from trivial that simple transfers can always form part of an equilibrium. Furthermore, equilibrium transfers depend on additional market features, including the number of children, the number of parents, the allocation itself, and the meeting technology.*

Now, we present an example of an environment where super-modularity does not imply PAM in equilibrium.

Example 1. (***Positive Assortative Matching Fails***) *Consider an environment where* $f(x_1) = 0.8$, $g(y_1) \in (0,1)$, *and the match values are determined by a super-modular function* $S(x,y)$ *with the following values:* $S(x_2,y_2) = 191$, $S(x_1,y_2) = 201$, $S(x_2,y_1) = 40$ *and* $S(x_1,y_1) = 51$. *Additionally, assume the meeting technology is given by* $\pi^p(\theta) = {}^{\pi^c(\theta)}/_{\theta} = {}^1/_{1+\theta}$.[21]

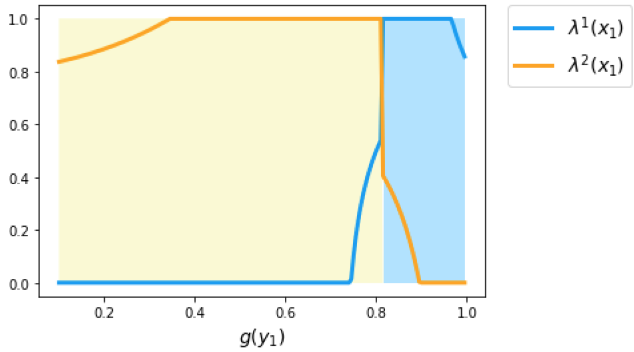


Figure 5: Randomization Device - Complete Information

Figure 5 illustrates the optimal allocation (randomization) of parents into submarket x_1 *for any* $g(y_1) \in (0,1)$. *For instance, when* $g(x_1) = 0.5$, *the optimal allocation is* $\lambda^{1*}(x_1) = 0$ *and* $\lambda^{2*}(x_1) = 1$, *reflecting perfect-NAM. In fact, the optimal menu exhibits NAM for any value of* $g(y_1)$ *below approximately* 0.8, *despite the surplus function* $S(x,y)$ *being super-modular. Refer to Appendix B.1 for a detailed analysis of this sample environment.* □

4 Equilibrium Analysis: Private Information

In this section, we analyze the case where a parent's ability is private information by solving the problem specified in Equation 1. Recall that high-ability parents incur a smaller cost when providing care for any child than low-ability parents do. Thus, high-ability parents receive a smaller expected transfer under the optimal menu specified for the complete information setting. As a result, high-ability

[21]Note the following: (i) the share of low-needs children is similar to what is observed in practice (see Appendix A), and (ii) the condition on primitives mentioned in Corollary 1(i) is not satisfied.

parents have incentives to mimic low-ability parents in order to receive a greater expected transfer in the presence of private information. This is true regardless of the sorting pattern/equilibrium allocation. Therefore, [PC] for low-ability parents and [IC] for high-ability parents must be binding in equilibrium. After incorporating these two constraints into the objective function in Equation 1, the designer's problem reduces to:

$$\max_{\{\lambda^k(x_1),\lambda^k(x_2)\}_{k=1}^2} \Bigg\{ \sum_{i=1}^{2} \pi^p(\theta_i) \left[\sum_{k=1}^{2} \Big(\underbrace{u(x_i,y_k) - c(x_i,y_k)}_{S(x,y)} \Big) \lambda^k(x_i)\, g(y_k) \right]$$
$$- \Big[c(x_1,y_1) - c(x_1,y_2)\Big] \lambda^1(x_1) \pi^p(\theta_1) g(y_2) - \Big[c(x_2,y_1) - c(x_2,y_2)\Big] \lambda^1(x_2) \pi^p(\theta_2) g(y_2) \Bigg\} \tag{3}$$

subject to [FC], [MT], and [IC] for low-ability parents.[22,23]

As one can see, in the objective function, extra terms appear in the second line due to information frictions . Looking closely, it corresponds to the expected gain that a high-ability parent obtains by mimicking low-ability parents. Even so, we start by showing that Lemma 1 and Proposition 1 presented in the previous section hold under private information (see Appendix D.1 and D.2). Following the same arguments as we did previously, the term that characterizes the optimal allocation of parents across submarkets becomes:

$$Z^{PI}(\theta_1) = \pi^p(\theta_2) \Bigg(\left[u(x_2,y_2) - \frac{c(x_2,y_2)}{g(y_1)} \right] - \left[u(x_2,y_1) - \frac{c(x_2,y_1)}{g(y_1)} \right] \Bigg)$$
$$- \pi(\theta_1) \Bigg(\left[u(x_1,y_2) - \frac{c(x_1,y_2)}{g(y_1)} \right] - \left[u(x_1,y_1) - \frac{c(x_1,y_1)}{g(y_1)} \right] \Bigg) \tag{4}$$

Note that $Z^{PI}(\theta_1)$ is analogous to $Z^{CI}(\theta_1)$, adjusted by the cost due to information

[22] Notice, [PC] for low-ability parents and [IC] for high-ability parents imply [PC] for high-ability parents, see proof of Proposition 4 in Appendix D.5.

[23] For the complete information case, the assumption $\sum_{x\in X} \lambda(x) = 1$ does not play a role in our results: if we relax it to $\sum_{x\in X} \lambda(x) \leq 1$, at the optimum this inequality will still be binding. In the private information setting, the optimum could change if we relax this equality: the designer might find it optimal to leave some foster parents out of the market to mitigate the incentives of mimicking. However, we believe that our assumption is reasonable considering that foster care exhibits a shortage of foster parents, who have to pass a rigorous assessment to be accepted to participate in the market. Thus, imposing that the system would like to employ all available parents is in line with the child welfare agencies objectives. In addition, relaxing this assumption would make the problem intractable for the private information case.

friction. Recall,

$$Z^{CI}(\theta_1) = \pi^p(\theta_2)\Big(\big[u(x_2, y_2) - c(x_2, y_2)\big] - \big[u(x_2, y_1) - c(x_2, y_1]\Big)$$
$$- \pi^p(\theta_1)\Big(\big[u(x_1, y_2) - c(x_1, y_2)\big] - \big[u(x_1, y_1) - c(x_1, y_1)\big]\Big) \quad (5)$$

A couple of significant insights are worth highlighting. First, if $g(y_1) = 1$ then Equations 4 and 5 are equivalent. In words, if there is no high-ability parents then there is no screening problem. Second, $\frac{c(x,y)}{g(y_1)}$ is greater than $c(x, y)$ for all (x, y). That is, in the private information case, the cost of providing care is amplified by the information friction. Third, as $g(y_1)$ increases, $\frac{c(x,y)}{g(y_1)}$ decreases and approaches to $c(x, y)$. In words, as the share of low-ability parents increases, the cost of information frictions decreases.

As in Section 3, the sign of $Z^{PI}(\cdot)$ at the equilibrium θ_1 determines the equilibrium sorting (see Appendix D.2). Now, we present sufficient conditions for monotone sorting under private information, analogous to Corollary 1:[24]

Corollary 2. *(i) If* $\frac{S(x_2,y_2)-S(x_2,y_1)+\frac{g(y_2)}{g(y_1)}\cdot[c(x_2,y_1)-c(x_2,y_2)]}{S(x_1,y_2)-S(x_1,y_1)+\frac{g(y_2)}{g(y_1)}\cdot[c(x_1,y_1)-c(x_1,y_2)]} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)}$ *and* $\frac{c(x_2,y_1)-c(x_2,y_2)}{c(x_1,y_1)-c(x_1,y_2)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)}$ *hold, then the equilibrium sorting exhibits PAM.*

(ii) If $\frac{S(x_1,y_2)-S(x_1,y_1)+\frac{g(y_2)}{g(y_1)}\cdot[c(x_1,y_1)-c(x_1,y_2)]}{S(x_2,y_2)-S(x_2,y_1)+\frac{g(y_2)}{g(y_1)}\cdot[c(x_2,y_1)-c(x_2,y_2)]} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_1)}\right)}$ *and* $\frac{c(x_1,y_1)-c(x_1,y_2)}{c(x_2,y_1)-c(x_2,y_2)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_1)}\right)}$ *hold, then the equilibrium sorting exhibits NAM.*

Proof. See Appendix D.3. □

Unlike the complete information case, the surplus function is not sufficient to elicit the equilibrium sorting pattern under the presence of private information. Here, we need to take into account the cost of a match as well as the distribution of parents. The first condition of Corollary 2(i) ensures that the allocation that maximizes the objective function exhibits PAM, and the second condition guarantees that the allocation is implementable—incentive compatible. Note, the second condition requires $c(x, y)$ to be a *strong* sub-modular function.[25] That is, the difference between the cost of providing care for a child x_2 and a child x_1 must be greater for low-ability parents than for high-ability parents. A sub-modular cost function implies that the informational rents paid to high-ability parents are

[24] See Appendix D.4 for a detailed characterization.

[25] By *strong* sub-modularity on the cost function we mean: $[c(x_2, y_1) - c(x_2, y_2)] \cdot \pi^p\left(\frac{1}{f(x_2)}\right) \geq c(x_1, y_1) - c(x_1, y_2)$ which introduces a constraint that is sensitive to the underlying distribution and the specific meeting technology. Similarly, the condition of *strong* super-modularity on the cost function follows.

lower under PAM than NAM. In other words, it is cheaper for the designer to motivate high-ability parents to report truthfully while inducing PAM. Thus, such a cost function pushes forces for the equilibrium sorting towards PAM. Note that, even if the surplus function $S(x, y)$ is sub-modular we may observe PAM, as in the complete information case. Analogous for 2(ii).

Next, motivated by the fact that the child welfare agency may not know the distribution of parents' attributes, we establish conditions that do not depend on this primitive:

Corollary 3. ***(i)*** *If* $\frac{S(x_2,y_2)-S(x_2,y_1)}{S(x_1,y_2)-S(x_1,y_1)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)}$ *and* $\frac{c(x_2,y_1)-c(x_2,y_2)}{c(x_1,y_1)-c(x_1,y_2)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)}$ *hold, then the equilibrium sorting exhibits PAM.*

(ii) *If* $\frac{S(x_1,y_2)-S(x_1,y_1)}{S(x_2,y_2)-S(x_2,y_1)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_1)}\right)}$ *and* $\frac{c(x_1,y_1)-c(x_1,y_2)}{c(x_2,y_1)-c(x_2,y_2)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_1)}\right)}$ *hold, then the equilibrium sorting exhibits NAM.*

Corollary 3 follows directly from Corollary 2. In item (i), we require $S(x, y)$ to be a strong super-modular function as in the complete information case, plus the condition of strong sub-modularity in $c(x, y)$ to ensure incentive-compatibility. Thus, we add an extra condition to the complete information result.

It is now important to examine the problem under more relaxed conditions, specifically when the conditions outlined in Corollaries 2 and 3 are not satisfied.

Remark 2. *The incentive-compatibility conditions are satisfied if and only if:*

$$\left(\lambda^2(x_2) - \lambda^1(x_2)\right) \underbrace{\left(\pi^p(\theta_2)\left[c(x_2, y_1) - c(x_2, y_2)\right] - \pi^p(\theta_1)\left[c(x_1, y_1) - c(x_1, y_2)\right]\right)}_{C(\theta_1)} \geq 0$$

where $C(\cdot)$ represents the expected cost difference between caring for child x_2 and child x_1 when comparing a high-ability parent to a low-ability parent. Note that, $C(\theta_1)$ increases with θ_1. Furthermore, $\lambda^2(x_2) - \lambda^1(x_2)$ is non-negative if the given allocation exhibits PAM.

Next, we provide a partial characterization. Let $C(\theta_1') = 0$ and $Z^{PI}(\theta_1'') = 0$ for some θ_1' and θ_1''. Thus, $C(\theta_1) > 0$ if and only if $\theta_1 > \theta_1'$, and $Z^{PI}(\theta_1) > 0$ if and only if $\theta_1 > \theta_1''$.

Proposition 3. *Let θ_1^{**} denote the equilibrium market tightness derived from the solution $\{\lambda^{k^{**}}(x_1)\}_{k=1}^2$ in Equation 3.* ***(i)*** *If $\theta_1^{**} \geq \max\{\theta_1', \theta_1''\}$ then $\{\lambda^{k^{**}}(x_1)\}_{k=1}^2$ is a solution to Equation 1 which exhibits PAM.* ***(ii)*** *If $\theta_1^{**} \leq \min\{\theta_1', \theta_1''\}$ then $\{\lambda^{k^{**}}(x_1)\}_{k=1}^2$ is a solution to Equation 1 which exhibits NAM.* ***(iii)*** *Otherwise, $\{\lambda^{k^{**}}(x_1)\}_{k=1}^2$ and the corresponding induced θ_1^{**} do not solve Equation 1.*

Proposition 3 is analogous to Proposition 1. It essentially states that the solution to the relaxed problem forms an equilibrium for the more constrained problem if the resulting market tightness in submarket-x_1 is either sufficiently high or low, relative to specific thresholds determined by the meeting technology, surplus, and cost functions: $\max\left\{\theta_1', \theta_1''\right\}$ and $\min\left\{\theta_1', \theta_1''\right\}$.

Next, we analyze the equilibrium transfers under private information. By fixing the optimal allocations $\{\lambda^{k**}(x_1), \lambda^{k**}(x_2)\}_{k=1}^{2}$ from Equation 3, the designer solves the following:

$$\min_{\{\tau^k(x_1),\tau^k(x_2)\}_{k=1}^2} \left\{ \sum_{i=1}^{2} \pi^p\left(\theta_i^{**}\right) \sum_{k=1}^{2} \tau^k(x_i)\lambda^{k**}(x_i)\; g(y_k) \right\}$$

subject to [FC], [PC] and [IC] from Equation 1. Here, the [PC] for low-ability parents, and the [IC] for high-ability parents determine the equilibrium transfer scheme. Formally:

Proposition 4. *Fix an equilibrium allocation of parents* $\{\lambda^{k**}(x_1), \lambda^{k**}(x_2)\}_{k=1}^{2}$. *Any feasible transfer schedule for which the PC for y_1-type parents as well as the IC for y_2-type parents are satisfied by equality is an equilibrium, if one of the following hold:*

(i) $\lambda^{2**}(x_2) \geq \lambda^{1**}(x_2)$ *and $c(x, y)$ is strong sub-modular, or*

(ii) $\lambda^{1**}(x_2) \geq \lambda^{2**}(x_2)$ *and $c(x, y)$ is strong super-modular*

Proof. See Appendix D.5. □

Note that, conditions in Corollaries 2 and 3 are also sufficient for Proposition 4. In particular, Corollary 2(i) or Corollary 3(i) ensure two things: (1) the cost function $c(x, y)$ is strong sub-modular, and (2) the randomization device exhibits PAM, that is, $\lambda^{2**}(x_2) \geq \lambda^{1**}(x_2)$. Thus, conditions in Corollaries 2 or 3 satisfy the conditions in Proposition 4(i). Analogous, for item (ii).

When the equilibrium sorting exhibits perfect PAM, the optimal transfer scheme is as follows: $\tau^{1**}(x_1) = c(x_1, y_1)$ and $\tau^{2**}(x_2) = c(x_2, y_2) + [c(x_1, y_1) - c(x_1, y_2)]\frac{\pi^p(\theta_1^{**})}{\pi^p(\theta_2^{**})}$. That is, y_1-parents receive exactly the cost of providing care, while y_2-parents receive the cost of providing care plus informational rents. Thus, it is no longer optimal to transfer parents just the cost of providing care; screening requires to compensate high-ability parents to disclose their type truthfully.

Similarly, let's suppose that the equilibrium sorting exhibits high-type PAM. In this case, the optimal transfers are as follows: $\tau^{1**}(x_1) = c(x_1, y_1)$, $\tau^{2**}(x_1) \geq 0$ and $\tau^{2**}(x_2) = c(x_2, y_2) - [\tau^{2**}(x_1) - c(x_1, y_2)]\frac{\pi^p(\theta_1^{**})\lambda^{2**}(x_1)}{\pi^p(\theta_2^{**})\lambda^{2**}(x_2)} + [c(x_1, y_1) - c(x_1, y_2)]\frac{\pi^p(\theta_1^{**})}{\pi^p(\theta_2^{**})\lambda^{2**}(x_2)}$.

For instance, suppose that parents who provide care in the same market *must* receive the same transfer, that is, $\tau^{1**}(x_1) = \tau^{2**}(x_1) = c(x_1, y_1)$. The optimal transfer for parent y_2 in submarket x_2 would be the following:

$$\tau^{2**}(x_2) = c(x_2, y_2) - \left[c(x_1, y_1) - c(x_1, y_2)\right] \frac{\pi^p(\theta_1^{**})\lambda^{2**}(x_1)}{\pi^p(\theta_2^{**})\lambda^{2**}(x_2)} + \left[c(x_1, y_1) - c(x_1, y_2)\right] \frac{\pi^p(\theta_1^{**})}{\pi^p(\theta_2^{**})\lambda^{2**}(x_2)}$$

As one can observe, compared to the complete information setting, a positive extra term appears in the transfer for high-ability parents who provide care in submarket x_2. This is to incentivize high-ability parents to reveal their true types.

Lastly, we revisit Example 1 by incorporating private information.

Example 2. (***Positive Assortative Matching Fails***). *Recall the environment in Example 1:* $f(x_1) = 0.8$, $g(y_1) \in (0, 1)$, $S(x_2, y_2) = 191$, $S(x_1, y_2) = 201$, $S(x_2, y_1) = 40$, $S(x_1, y_1) = 51$, *and that* $\pi^p(\theta) = {}^1\!/_{1+\theta}$. *Moreover suppose that the cost function is supermodular with the following values:* $c(x_2, y_2) = 15$, $c(x_1, y_2) = 1$, $c(x_2, y_1) = 20$ *and* $c(x_1, y_1) = 15$.[26]

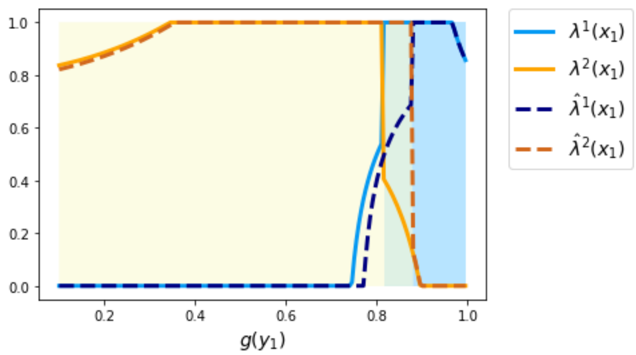


Figure 6: Randomization Device - Private & Complete Information

Figure 6 shows that the optimal randomization devices for both complete and private information scenarios are remarkably similar. Specifically, $\lambda^1(x_1)$ *and* $\lambda^2(x_2)$ *closely resemble* $\hat{\lambda}^1(x_1)$ *and* $\hat{\lambda}^2(x_2)$*, respectively.*[27] *However, when* $g(y_1)$ *is approximately in* $(0.8, 0.9)$*, the equilibrium sorting pattern is PAM under complete information, whereas it is NAM with private information. To see the intuition, consider an equilibrium menu of licenses that implements perfect sorting under the complete information, such that*

[26] Notice, the cost function here guarantees the existence of a separating menu of licenses under NAM, whereas any equilibrium exhibiting PAM does not screen parents.

[27] This similarity arises due to the values of the surplus and cost functions. If the values of the cost function were to increase, it would lead to a notable disparity in the optimal randomization rule between the complete and private information settings.

$\tau^k(x) = c(x, y_k)$. *If the menu implements NAM(PAM), then type-y_2 parents pretend to be type-y_1 to be able to match with type-x_2(type-x_1) instead of type-x_1(type-x_2) children. The misreport under NAM allows parent type-y_2 parent to (ex-post) gain as much as $\tau^1(x_2) - c(x_2, y_2) = 5$ whereas the misreport under PAM does $\tau^1(x_1) - c(x_1, y_2) = 14$. That is, type-$y_2$ parents have stronger incentives to misreport if the equilibrium sorting is PAM than when it is NAM. Thus, it is cheaper for the designer to switch the equilibrium sorting from PAM to NAM for the (roughly) specified region of $g(y_1)$. Notice, this intuition is in line with the counterpart of Corollary 2(i). See Appendix B.2 for an in-depth analysis of the sample environment discussed in this section.*

□

5 Discussions and Extensions

In this section, we extend our complete information model in two key directions and present the corresponding findings. First, we maintain the classification of children into high- and low-need categories while broadening the parental attribute space to a continuum. Second, we analyze the sensitivity of our results by conducting comparative statics on the meeting technology.

5.1 Continuous type of Parents

Motivated by the fact that a parents ability to provide care might be a continuous variable, we now assume that parents differ in $y \in |\underline{y}, \overline{y}| \equiv Y \subset \mathbb{R}^+$, which follows a continuous and differentiable cumulative distribution function (CDF) $G(x)$ with a strictly positive probability density function (PDF) $g(x)$. We uphold our regularity conditions: $S(x, y)$ is increasing in y, while $c(x, y)$ is increasing in x and decreasing in y.

Now, a license for type-y parent is $\mathcal{L}(y) \equiv \left\{\left(\lambda(x_i, y), \tau(x_i, y)\right)\right\}_{i=1}^2$. Thus, conditional on a meeting taking place, the probability that child x has met a parent y is equal to:

$$\frac{\lambda(x, y)g(y)}{\int_{\underline{y}}^{\overline{y}} \lambda(x, y)g(y)\, dy}$$

Thus, the net expected utility in each submarket x, conditional on a meeting taking place, is:

$$W(x) = \frac{\int_{\underline{y}}^{\overline{y}} [u(x,y) - \tau(x,y)] \lambda(x,y) g(y) \, dy}{\int_{\underline{y}}^{\overline{y}} \lambda(x,y) g(y) \, dy}.$$

Then, the designer's problem is:

$$\max_{\left(\left\{\left(\lambda(x_i,y),\tau(x_i,y)\right)\right\}_{i=1}^{2}\right)_{y \in Y}} \left\{ \sum_{i=1}^{2} \pi^c(\theta_i) \, W(x_i) \, f(x_i) \right\} \tag{6}$$

subject to [FC] and [PC] defined as in earlier, and

$$[\mathrm{MT}] \quad \theta_i = \frac{1}{f(x)} \cdot \int_{\underline{y}}^{\overline{y}} \lambda(x,y) g(y) \, dy \, , \text{ for all } i \in \{1,2\}$$

After incorporating the [PC] equations into the objective function, and using the [MT] along with the relationship $\pi^c(\theta)/\theta = \pi^p(\theta)$, the designer's problem reduces to:

$$\max_{\left(\left\{\lambda(x_i,y)\right\}_{i=1}^{2}\right)_{y \in Y}} \left\{ \sum_{i=1}^{2} \pi^p(\theta_i) \cdot \int_{\underline{y}}^{\overline{y}} S(x_i,y) \lambda(x_i,y) g(y) \, dy \right\} \tag{7}$$

It is easy to see that the randomization device $\lambda(x,y)$ is independent of whether interim or ex-post participation constraints are satisfied (see Corollary C.1). Furthermore, the segregation result from the two-type case (see Lemma 1) extends to the current environment, albeit with a caveat:

Lemma 2. *Any **interior** randomization $\lambda(x,y)$ is **not** optimal.*

Proof. See Appendix E.1.1. □

Lemma 2 states that for each parent y, the randomization device will take a value of either zero or one. Notably, unlike in the complete information case, a parent will foster only one type of child here. Therefore, the nested structure of licenses described in the introduction does not appear in this case. The proof closely follows the two-type case. Begin with an interior randomization $\lambda(x_1,y) \in (0,1)$ for any y such that $\lambda(x_2,y) = 1 - \lambda(x_1,y)$ by the [FC]. Next, define a monotone function $\varepsilon : Y \to (0,1)$ such that $\int \varepsilon(y) g(y) \, dy = 0$. Apply a perturbation to the initial pair $\lambda(y) = \left(\lambda(x_1,y), \lambda(x_2,y)\right)$ to obtain a new allocation $\tilde{\lambda}(y) \equiv \left(\lambda(x_1,y) - \varepsilon(y), \lambda(x_2,y) + \varepsilon(y)\right)$ ensuring that market tightness remains unchanged under both $\lambda(y)$ and $\tilde{\lambda}(y)$. The change in welfare resulting from this perturbation is:

$$\Delta_W = \int_{\underline{y}}^{\overline{y}} \underbrace{\left[\pi^p(\theta_2)S(x_2, y) - \pi^p(\theta_1)S(x_1, y)\right]}_{\hat{Z}^{CI}(y|\theta)} \varepsilon(y)g(y)\, dy,$$

where $\theta = (\theta_1, \theta_2)$. Then, we construct the perturbation function $\varepsilon(y)$ such that it is monotone increasing over regions where $\hat{Z}^{CI}(y|\theta)$ is increasing and monotone decreasing over regions where $\hat{Z}^{CI}(y|\theta)$ is decreasing. This yields a strictly positive change in welfare, which finishes the proof.

Next, let's extend the definition of sorting patterns:

Definition 2. *If $\lambda(x_2, y)$ is non-decreasing (non-increasing) in y, then the sorting exhibits PAM (NAM).*

Proposition 5 states that, given an equilibrium market tightness (θ_1^*, θ_2^*), there exists a threshold parent level such that all parents below this threshold are allocated to one submarket, while all parents above it are allocated to the other submarket. This result reflects a complete segregation of the market.

Proposition 5. *Let (θ_1^*, θ_2^*) be an equilibrium market tightness.*

(*i*) *If $\pi^p(\theta^*)S(x, y)$ is super-modular, then the equilibrium sorting exhibits PAM. That is, $\lambda^*(x_2, y) = 0$ for all $y \leq \hat{y}^{PAM}$ and $\lambda^*(x_2, y) = 1$ for all $y > \hat{y}^{PAM}$ where $\theta_1^* = {G(\hat{y}^{PAM})}/{1-f(x_2)}$ and $\theta_2^* = {1-G(\hat{y}^{PAM})}/{f(x_2)}$.*

(*ii*) *If $\pi^p(\theta^*)S(x, y)$ is sub-modular, then the equilibrium sorting exhibits NAM. That is, $\lambda^*(x_2, y) = 1$ for all $y \leq \hat{y}^{NAM}$ and $\lambda^*(x_2, y) = 0$ for all $y > \hat{y}^{NAM}$ where $\theta_1^* = {1-G(\hat{y}^{NAM})}/{1-f(x_2)}$ and $\theta_2^* = {G(\hat{y}^{NAM})}/{f(x_2)}$.*

Proof. See Appendix E.1.2. □

In the following, we establish sufficient conditions for monotone sorting. To this end, let $S(x, y)$ be continuous and differentiable over Y, and let $S_y(x, \cdot)$ denote the partial derivative of $S(x, y)$ with respect to y.

Corollary 4.

(*i*) *If $\frac{S_y(x_2,\grave{y})}{S_y(x_1,\grave{y})} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)}$ where $\grave{y} := \arg\min_{y \in Y} \pi^p\left(\frac{1}{1-f(x_1)}\right) S_y(x_2, y) - S_y(x_1, y)$, then the equilibrium sorting exhibits PAM. That is, $\lambda^*(x_2, y) = 0$ for all $y \leq \hat{y}^{PAM}$ and $\lambda^*(x_2, y) = 1$ for all $y > \hat{y}^{PAM}$ with:*

$$\hat{y}^{PAM} := \arg\max_{\hat{y} \in Y} \pi^p\left(\frac{G(\hat{y})}{f(x_1)}\right) \int_{\underline{y}}^{\hat{y}} S(x_1, y)g(y)dy + \pi^p\left(\frac{1 - G(\hat{y})}{1 - f(x_1)}\right) \int_{\hat{y}}^{\overline{y}} S(x_2, y)g(y)dy.$$

(ii) *If* $\frac{S_y(x_1,\acute{y})}{S_y(x_2,\acute{y})} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_1)}\right)}$ *where* $\acute{y} := \arg\max_{y \in Y} S_y(x_2,y) - \pi^p\left(\frac{1}{f(x_1)}\right) S_y(x_1,y)$*, then the equilibrium sorting exhibits NAM. That is,* $\lambda^*(x_2,y) = 1$ *for all* $y \leq \hat{y}^{NAM}$ *and* $\lambda^*(x_2,y) = 0$ *for all* $y > \hat{y}^{NAM}$ *with:*

$$\hat{y}^{NAM} := \arg\max_{\hat{y} \in Y} \pi^p\left(\frac{1-G(\hat{y})}{f(x_1)}\right) \int_{\hat{y}}^{\overline{y}} S(x_1,y)g(y)dy + \pi^p\left(\frac{G(\hat{y})}{1-f(x_1)}\right) \int_{\underline{y}}^{\hat{y}} S(x_2,y)g(y)dy.$$

Proof. See Appendix E.1.3. □

Corollary 4 (i) implies that $S(x,y)$ is supermodular. This suggests that our condition is slightly stronger than the standard supermodularity of the surplus function required to ensure PAM, as in the two-type case. Analogously, a sufficient condition—Corollary 4 (ii)—for NAM follows. Furthermore, these conditions result in a segregation of the market, with one key distinction to the two-type case: no parent is allocated across both submarkets. Specifically, a threshold type-y parent emerges, dividing the type space into two distinct partitions, each allocated to a separate submarket. This contrasts with environments involving a finite number of parent types. An immediate implication is that if the conditions for monotone sorting are satisfied, the designer's problem simplifies to selecting the threshold type-y parent that partitions the type space. This choice is effectively equivalent to determining the market tightness.

Remark 3. *The model and analysis can be readily extended to accommodate an arbitrary environment with* $X = \{x_1, x_2, \ldots, x_n\}$ *where* $n \geq 2$ *while* $y \sim Y$ *with a nonzero PDF* $g(\cdot)$*. In such a setting, whenever* $\lambda(x,y) \in (0,1)$ *for some* $x \in \{x_i, x_j\}$ *over some non-zero measure* $Y' \subseteq Y$*, a similar perturbation—between submarkets* i *and* j *over the type-space* Y' *without altering MT at each submarket—yields the following change in the welfare:*

$$\Delta_W = \int_{y \in Y'} \left[\pi^p(\theta_{x_i})S(x_i,y) - \pi^p(\theta_{x_j})S(x_j,y)\right]\varepsilon(y)g(y)dy.$$

Thus, the steps outlined above can be easily followed to replicate the analysis and derive characterizations analogous to Proposition 5 as well as Corollary 4.

Lastly, regarding the transfers, they can be easily determined by the [PC] since parents are allocated to exactly one submarket. Specifically, $\tau^*(x_i,y) = c(x_i,y)$ if $\lambda^*(x_i,y) = 1$ for all (x_i,y). Thus, in this case, we observe that paying parents exactly the cost of providing care, as mentioned in the Arizona example, constitutes an equilibrium. However, it is important to note that this outcome is optimal only when licenses are not nested.

5.2 Improvement in the Meeting Technology

In this section, we analyze the role of meeting technology in modeling search frictions within the allocation process. Specifically, we investigate how changes in search technology—whether advancements or setbacks—affect our sorting results. This analysis is motivated by the observed differences in the effectiveness of child welfare agencies in matching children with suitable foster families.

Formally, we define what constitutes an improvement in search technology, concentrating, without loss of generality, on its application to parents. Recall that $\pi^p(\cdot)$ is a strictly decreasing and strictly convex function bounded by $\pi^p(0) = 1$ and $\lim_{\theta\to\infty} \pi^p(\theta) = 0$.[28] Let $\mathbf{\Pi}^p$ be the set of all such bounded, strictly decreasing, and strictly convex functions. For any $\pi^p, \hat{\pi}^p \in \mathbf{\Pi}^p$, we say that $\hat{\pi}^p$ is an **improved technology** compared to π^p if $\partial\hat{\pi}^p(\theta)/\partial\theta > \partial\pi^p(\theta)/\partial\theta$ for any finite θ. Note, this also implies $\hat{\pi}^p(\theta) > \pi^p(\theta)$. If $\hat{\pi}^p$ is an improved technology of π^p, we denote $\pi^p \rhd \hat{\pi}^p$. Now, the following partially characterizes the equilibrium sorting as meeting technology improves:

Proposition 6. *Suppose $S(x, y)$ is super-modular (sub-modular). If the equilibrium sorting is PAM (NAM) with some meeting technology π^p, then it remains PAM (NAM) for any $\hat{\pi}^p$ such that $\pi^p \rhd \hat{\pi}^p$.*

Proof. See Appendix E.2.1. □

6 Concluding Remarks

This paper analyzes the foster care system in the US as a two-sided matching market wherein one side consists of children who are heterogeneous in level of care needed, and the other side consists of parents who differ in their ability to take care of a child. We solve for the optimal menu of licenses which specifies an allocation of parents across submarkets of children as well as the corresponding transfers, under the presence of search and information frictions.

With a discrete type space, the paper establishes two key results that hold regardless of the information frictions: **(i)** it is not optimal to mix multiple types of parents into multiple submarkets of children, and **(ii)** super-modularity and sub-modularity of the surplus of a match are neither sufficient nor necessary conditions for the optimal sorting to exhibit PAM and NAM, respectively. The former

[28]Moreover, recall that $\pi^c(\theta)/\theta \equiv \pi^p(\theta)$. And thus, for any other meeting technology for parents $\tilde{\pi}^p(\theta)$, it has to be the case that $\tilde{\pi}^p(\theta)\cdot\theta = \tilde{\pi}^c(\theta)$ is strictly increasing and concave with the following bounds $\tilde{\pi}^c(0) = 0$ and $\lim_{\theta\to\infty} \tilde{\pi}^c(\theta) = 1$.

rationalizes the nested nature of the menu of licenses offered by various states in the US. The latter has implications on the optimal allocation of parents: even if the surplus shows complementarity (substitutability) in child and parent's attributes, allocating parents into submarkets such that the sorting exhibits PAM (NAM) is not necessarily optimal due to search frictions.

We also make inferences once information friction is introduced: as the share of low-type parents increases, the allocation of parents approaches to the first-best (complete information). Because, high-type parents mimic the low-type ones to receive a greater expected transfer. As a result, the designer pays information rents to high-type parents to overcome such incentives. The smaller the share of high-type parents, the less the designer cares about such mimicking incentives. However, if the proportion of high-type parents is big enough, then not only the allocation diverges from the first-best, but also the optimal sorting may reverse.

Lastly, we analyze the sensitivity of our results by introducing a continuous attribute space for parents and briefly discuss the implications of expanding the discrete attribute space for children.

References

Altinok, A. (2023): "Group Lending, Sorting, and Risk Sharing," *Games and Economic Behavior*, 140, 456–480.

Becker, G. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81, 813–846.

Cox, M., D. Cherry, and J. Orme (2011): "Measuring the willingness to foster children with emotional and behavioral problems," *Children and Youth Services Review*, 33, 59–65.

Damiano, E. and H. Li (2007): "Price Discrimination and Efficient Matching," *Economic Theory*, 30, 243–263.

DeVooght, K. and D. Blazey (2013): "Family Foster Care Reimbursement Rate in the US: A Report from a 2012 National Survey on Family Foster Care Provider Classifications and Rates," *Child Trends Publications.*

Eeckhout, J. and P. Kircher (2010): "Sorting and Decentralized Price Competition," *Econometrica*, 78, 539–574.

Ghatak, M. (1999): "Group Lending, Local Information, and Peer Selection," *Journal of Development Economics*, 60, 27–50.

——— (2000): "Screening by the Company You Keep: Joint Liability Lending and the Peer Selection Effect," *Economic Journal*, 110, 601–631.

Gibbs, D. and J. Wildfire (2007): "Length of service for foster parents: Using administrative data to understand retention," *Children and Youth Services Review*, 29, 588–599.

Guttman, J. (2008): "Assortative Matching, Adverse Selection, and Group Lending," *Journal of Development Economics*, 81, 51–56.

MacDonald, D. (2022): "Foster Care: A Dynamic Matching Approach," *Working Paper*.

Menzio, G. and S. Shi (2010a): "Block recursive equilibria for stochastic models of search on the job," *Journal of Economic Theory*, 145, 1453–1494.

——— (2010b): "Directed Search on the Job, Heterogeneity, and Aggregate Fluctuations," *The American Economic Review*, 100, 327–332.

O'Hare, W. (2008): "Data on Children in Foster Care from the Census Bureau," *Annie E. Casey Foundation*.

Olberg, N., L. Dierks, S. Seuken, V. W. Slaugh, and U. Unver (2021): "Search and Matching for Adoption from Foster Care," *Working Paper*.

Robinson-Cortés, A. (2019): "Who Gets Placed Where and Why? An Empirical Framework for Foster Care Placement," *Working Paper*.

Rosinsky, K., M. Fischer, and M. Haas (2023): "Child Welfare Financing SFY 2020: A Survey of Federal, State, and Local Expenditures," *Child Trend Publications*.

Shi, S. (2001): "Frictional Assignment. I. Efficiency," *Journal of Economic Theory*, 98, 232–260.

Shimer, R. and L. Smith (2000): "Assortative Matching and Search," *Econometrica*, 68, 343–369.

Slaugh, V., M. Akan, O. Kesten, and U. Unver (2015): "The Pennsylvania Adoption Exchange Improves Its Matching Process," *Boston College Working Papers in Economics 858*.

Smith, L. (2006): "The Marriage Model with Search Frictions," *Journal of Political Economy*, 114, 1124–1144.

Van Tassel, E. (1999): "Group Lending Under Asymmetric Information," *Journal of Development Economics*, 60, 3–25.

A Appendix: Foster Care in the US

A.1 Overview

During 2020 Federal Fiscal Year (FFY),[29] child welfare agencies across the United States received more than 3.9 million allegations of suspected child abuse or neglect (equivalent to approximately 7.1 million children). Out of these children, 9 percent were removed from their homes and placed into foster care. According to Rosinsky et al. (2023), the national spending on child welfare in 2020 FFY was approximately US$34.1, out of which US$15.2 billion was federally funded, and the remaining was financed directly by States. Furthermore, 45 percent of the national spending was destined to foster care placement expenditure, including payments to foster parents.

Using the Foster Care Files from AFCARS,[30] we observed that in 2020 FFY there were 631,254 children in foster care. On average, these children were almost 7 years old, 49 percent were females, 69 percent were white, and 24 percent were clinically diagnosed with a disability.[31] Thus, based on the disability variable, we can infer that at least 24 percent of children in the US foster care are special needs.[32] During their stay in foster care, 77 percent of these children were placed with foster parents, 9 percent were placed in institutional care, and the remaining had other arrangements. Foster parents caring for children with and without a disability received an average payment of US$1,423 and US$ 2,704 per month, respectively. In this data set, foster parents are not identifiable; only family structure, race and year of birth are reported. Thus, since we do not know how many times a foster parent might appear, we can not provide reliable statistics.

Most of the information regarding foster parents comes from Census data and surveys. Using Census data from 2000, O'Hare (2008) finds that households with foster children, compared to all other households with children, are: less likely to be married-couples, less likely to have a member who finished college, less likely to work full-time, more likely to be low income families, and more likely to receive public assistance income. Now, after conducting a survey of 297 foster mothers, Cox et al. (2011) finds that the average age is 44.1 years old, 88.2 percent

[29] October 1, 2019 to September 30, 2020.

[30] AFCARS is a federally mandated data collection system. All fifty US states and the District of Columbia are required to collect data on all children in foster care and all children adopted from foster care.

[31] A disability includes conditions such as blindness, glaucoma, arthritis, multiple sclerosis, down syndrome, personality disorder, attention deficit, and anxiety disorder, among others.

[32] In the majority of the cases, once a child enters the foster care system, a mandatory medical evaluation is performed, therefore we assume that the level of care needed is common knowledge.

are European-American, 75.1 percent are married, 28.9 percent have a bachelor's degree, 33 percent works full-time, and 50.1 percent have an annual family income less than USD$50,000.

A.2 Matching Process

Foster care is overseen and managed at the state level by Child Protective Services (CPS). Upon receiving an allegation regarding a child's well-being, CPS assigns a social worker to the case, starting an investigation. If sufficient evidence supporting an accusation is identified, the case is presented to a juvenile or family court. The judge then determines whether the child should be removed from their birth family home and placed in foster care.

In many states, decisions regarding the placement of children are made by social workers. Acting on behalf of the child, the social worker (a) searches for and contacts foster parents, (b) facilitates a meeting between the foster parent and child to assess compatibility, and (c) decides on the placement of the child. In this search process, the social worker can only consider fosters parent who are certified, through a license, to provide care for the child.

Foster parents must obtain a license to provide care for children. The licensing process involves a home study and mandatory training. The home study ensures the foster parent's residence is clean, in good condition, and free from hazards. Initial training, ranging from 15 to 30 hours, covers topics such as agency policies, foster parent roles and responsibilities, and behavior management. The menu of licenses varies across states (for more details see DeVooght and Blazey (2013)). As we mentioned in the introduction, children are grouped by the level of care needed, and transfers vary across groups. These transfers follow the principle that foster parents caring for children with high-needs receive greater transfers.

B Appendix: Omitted Examples

In this section, we provide a more in-depth discussion of Examples 1 and 2. Additionally, we present environments that satisfy Corollaries 1 and 2, along with an explanation of how they fulfill the conditions for PAM.

B.1 Complete Information

Example B.1. (***Detailed explanation of Example 1***). Figure B.1 illustrates an environment where super-modularity in the surplus function $S(x, y)$ is not a sufficient condition for PAM. We assume that the share of low-needs children is $f(x_1) = 0.8$, the functional form of the meeting technology is $\pi^p(\theta) = \frac{1}{1+\theta}$, and $S(x, y)$ is a super-modular function with values $S(x_2, y_2) = 191$, $S(x_1, y_2) = 201$, $S(x_2, y_1) = 40$ and $S(x_1, y_1) = 51$. Here, the condition over primitives presented in Corollary 1(i) is violated: $1 = \frac{S(x_2,y_2)-S(x_2,y_1)}{S(x_1,y_2)-S(x_1,y_1)} \not\geq \frac{1}{\pi^p(1/f(x_2))} = 5.99$

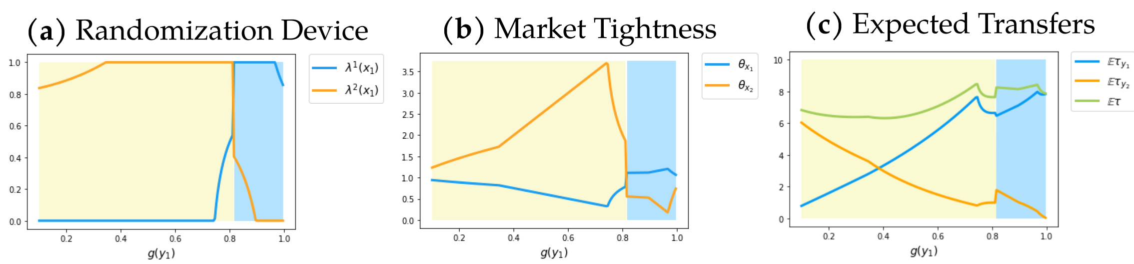


Figure B.1: Monotone Sorting Fails

Panel B.1a presents the optimal probability with which parents holding licenses 1 (blue line) and 2 (orange line) are allocated into submarket x_1. The y-axis corresponds to these probabilities while the x-axis presents possible values for the share of low-ability parents, $g(y_1)$. In Panel B.1b, we plot the optimal market tightness for submarket x_1 (blue line) and x_2 (orange line) as a function of the share of low-ability parents, $g(y_1)$. Here, the y-axis corresponds to possible values for the market tightness. In Panel B.1c, we plot the optimal expected transfers received by all y_1-parents (blue line) and all y_2-parents (orange line) as a function of the share of low-ability parents, $g(y_1)$. In addition, we also include the optimal total expected transfers (green line), or equivalently, the optimal total cost incurred by the child welfare agency to implement the optimal sorting. Lastly, in every graph, the blue- and golden- shaded regions correspond to PAM and NAM, respectively.

As Panel B.1a illustrates, for small enough values of $g(y_1)$, the equilibrium sorting exhibits NAM, even when the surplus function is super-modular. Thus, super-modularity is not a sufficient condition for PAM to hold in equilibrium. For the same interval of $g(y_1)$, Panel B.1b shows that the equilibrium market tightness is greater in submarket x_2 than x_1; thus, parents are more likely to meet a child in submarket x_1. This induces the designer to allocate y_2-parents in submarket x_1, resulting in NAM. As $g(y_1)$ increases, the equilibrium market tightness becomes larger in submarket x_1 than in x_2, and thus the equilibrium sorting reverses to PAM. Lastly, we can see from Panel B.1c that the total expected cost of imposing NAM increases as the share of low-ability parents increases.[33] This is intuitive, since low-ability parents incur a greater cost for providing care than high-ability parents. Therefore, the designer must pay greater transfers to low-ability parents to satisfy the [PC]. □

Example B.2. (***Positive Assortative Matching Holds***). Figure B.2 considers an environment that satisfies the condition presented in Corollary 1(i) to ensure that PAM will arise in equilibrium. In this environment, we assume that the share of low-needs children is equal to 0.8, and $S(x, y)$ is a super-modular function with revised values $S(x_2, y_2) = 100$, $S(x_1, y_2) = 201$, $S(x_2, y_1) = 30$ and $S(x_1, y_1) = 191$. This set of primitives satisfies the following: $7 = \frac{S(x_2,y_2)-S(x_2,y_1)}{S(x_1,y_2)-S(x_1,y_1)} \geq \frac{1}{\pi^p(1/f(x_2))} = 5.99$

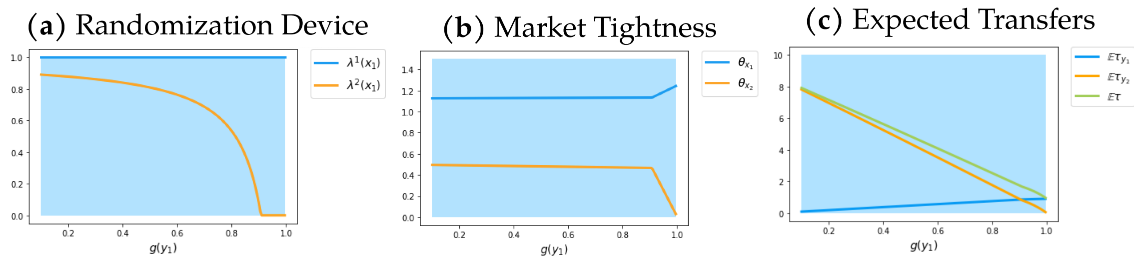


Figure B.2: Monotone Sorting Holds

As Panel B.2a illustrates, the equilibrium sorting exhibits PAM for any value of $g(y_1)$. Moreover, for sufficiently high values of $g(y_1)$, there is a perfect segregation of the market such that all type-y_i parents are allocated into submarket x_i.

Panel B.2b shows that the market tightness in both submarkets remains flat for a fair range of values of $g(y_1)$, even though the share of y_2-parents being allocated into submarket x_1 decreases. This is due to two effects compensating: **(i)** $\theta_1^\star$ increases as $g(y_1)$ increases, and **(ii)** $\theta_1^\star$ decreases as $\lambda^{2\star}(x_1)$ decreases. Similarly, for

[33] Though the cost structure is not necessary for the analysis of equilibrium allocations, it directly determines the equilibrium total transfers. Panel B.1c uses the following cost function: $c(x_1, y_1) = 15, c(x_1, y_2) = 1, c(x_2, y_1) = 20, c(x_2, y_2) = 15$.

$\theta_2^\star$. In addition, note that the market tightness is larger in submarket x_1 than in submarket x_2, resulting in parents being more likely to meet a child in submarket x_2. This is in line with the intuition that the designer would like to allocate more profitable parents into thicker submarkets.

Lastly, in Panel B.2c, the total expected cost of implementing PAM is decreasing in $g(y_1)$, unlike the intuition presented in the previous example.[34] Here, the total expected transfers received by y_1-parents (blue line) are increasing in $g(y_1)$, but not enough to compensate for the decrease of the total expected transfers received by y_2-parents (orange line). □

B.2 Private Information

Example B.3. (***Detailed explanation of Example 2***)**.** Figure B.3 illustrates the environment in Example B.1, where super-modularity in the surplus function is not a sufficient condition for PAM.[35] In all panels, the solid lines represent the equilibrium objects under the complete information, while the dashed lines correspond to the private information. The cost function is super-modular with the following values: $c(x_2, y_2) = 15$, $c(x_1, y_2) = 1$, $c(x_2, y_1) = 20$ and $c(x_1, y_1) = 15$. Notice, it guarantees the existence of a separating menu of licenses under NAM, whereas any equilibrium exhibiting PAM does not screen parents.

(**a**) Randomization Device (**b**) Market Tightness (**c**) Expected Transfers

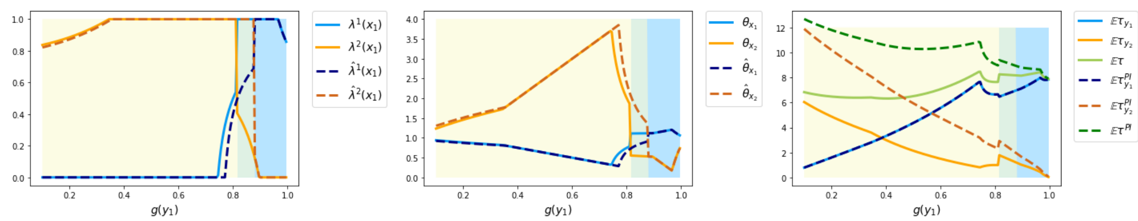


Figure B.3: Monotone Sorting Fails under Private Information

In Panel B.3a, one can observe that the optimal randomization devices, $\lambda^1(x_1)$ and $\lambda^2(x_2)$, are very similar for the complete and private information cases.[36] As a result, Panel B.3b represents that the optimal market tightness coincides at a

[34] The cost structure here is as follows: $c(x_1, y_1) = 2, c(x_1, y_2) = 1, c(x_2, y_1) = 20, c(x_2, y_3) = 15$.

[35] Recall that the primitives used in Example B.1 are as follows: $f(x_1) = 0.8$, $\pi^p(\theta) = \frac{1}{1+\theta}$, $S(x_2, y_2) = 191$, $S(x_1, y_2) = 201$, $S(x_2, y_1) = 40$ and $S(x_1, y_1) = 51$.

[36] This similarity arises due to the values of the surplus and cost functions. If the values of the cost function were to increase, it would lead to a notable disparity in the optimal randomization rule between the complete and private information settings.

fairly large interval for $g(y_1)$. However, when $g(y_1)$ is approximately in $(0.8, 0.9)$, the equilibrium sorting pattern is PAM under complete information, whereas it is NAM with private information. To see the intuition, consider an equilibrium menu of licenses that implements perfect sorting under complete information, such that $\tau^k(x) = c(x, y_k)$. If the menu implements NAM, then type-y_2 parent mimics y_1 and matches with child x_2 instead of x_1, and if it implements PAM, type-y_2 mimics y_1 and matches with x_1 instead of x_2. The former misreport allows parent y_2 to (ex-post) gain as much as $\tau^1(x_2) - c(x_2, y_2) = 5$ whereas the latter does $\tau^1(x_1) - c(x_1, y_2) = 14$. That is, y_2 has *stronger* incentives to misreport if the equilibrium sorting is PAM than when it is NAM. Thus, it is cheaper for the designer to switch the equilibrium sorting from PAM to NAM for the (roughly) specified region of $g(y_1)$. Notice, this intuition is in line with the counterpart of Corollary 2(i).

Lastly, in Panel B.3c we observe that the total expected transfers received by low-ability parents (blue lines) coincides under complete and private information. The reason is that, for low-ability parents, transfers are pint-down by the [PC] regardless of the information friction. For high-ability parents (orange lines), total expected transfers are greater under private than complete information. This is intuitive, since the designer must pay informational rents to incentivize high-ability parents to reveal their type truthfully when informational fictions are introduced. □

Example B.4. (***Positive Assortative Matching Holds***)**.** Figure B.4 illustrates the equilibrium objects of the environment in Example B.2 that satisfies the additional conditions presented in Corollary 3(i).[37] We assume $c(x, y)$ is a strong submodular function with values $c(x_2, y_2) = 13$, $c(x_1, y_2) = 1$, $c(x_2, y_1) = 20$ and $c(x_1, y_1) = 2$.

One can easily verify that the cost function guarantees the existence of a separating menu of licenses under PAM. In this case, the conditions over primitives presented in Corollary 3(i) are satisfied:

$$\frac{S(x_2, y_2) - S(x_2, y_1)}{S(x_1, y_2) - S(x_1, y_1)} = 7 = \frac{c(x_2, y_1) - c(x_2, y_2)}{c(x_1, y_1) - c(x_1, y_2)} \geq \frac{1}{\pi^p\left(1/f(x_2)\right)} = 5.99$$

As Panel B.4a illustrates, the equilibrium sorting exhibits PAM for any value of $g(y_1)$. Thus, PAM is robust to informational frictions, unlike Example B.3 where we observe PAM and NAM. In Panel B.4b, we observe that the market tightness in

[37] Recall that the primitives used in Example B.2 are as follows: $f(x_1) = 0.8$, $\pi^p(\theta) = \frac{1}{1+\theta}$, $S(x_2, y_2) = 100$, $S(x_1, y_2) = 201$, $S(x_2, y_1) = 30$ and $S(x_1, y_1) = 191$.

both submarkets remains flat for a fair range of values of $g(y_1)$. Lastly, in Panel B.4c we observe that the total cost of imposing PAM decreases with $g(y_1)$. Note that, in this example, the equilibrium allocations are almost identical under complete and private information.

(a) Randomization Device (b) Market Tightness (c) Expected Transfers

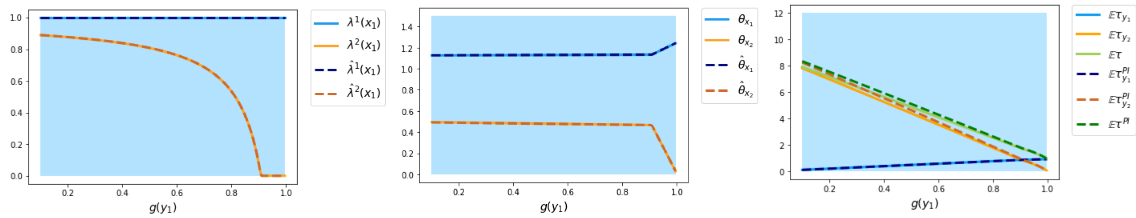


Figure B.4: Monotone Sorting Holds under Private Information

□

C Appendix: Analysis of Complete Information

In this section, we prove the results for the complete information case. For each parent y_k with $k = \{1, 2\}$, the designer offers a licenses (λ^k, τ^k). The designer solves the following problem:

$$\max_{\left\{\left(\lambda^k(x_i), \tau^k(x_i)\right)_{i=1}^2\right\}_{k=1}^2} \left\{ \sum_{i=1}^{2} \pi^c\big(\theta_i\big) \, \frac{\sum_{k=1}^2 \left[u(x_i, y_k) - \tau^k(x_i)\right] \lambda^k(x_i) g(y_k)}{\sum_{k=1}^2 \lambda^k(x_i) \; g(y_k)} \; f(x_i) \right\}$$

subject to:

$$\text{[FC]} \; \tau^k(x) \geq 0 \text{ and } \lambda^k(x) \geq 0 \text{ for all } (k, x), \text{ and} \sum_{i=1}^{2} \lambda^k(x_i) = 1 \text{ for all } k = 1, 2.$$

$$\text{[MT]} \quad \theta_x = \frac{1}{f(x)} \cdot \sum_{k=1}^{2} \left[\lambda^k(x) \sum_{j=1}^{2} h^k(y_j) \right], \text{ for all } x.$$

$$\text{[PC]} \; \sum_{i=1}^{2} \left[\tau^k(x_i) - c(x_i, y_k)\right] \lambda^k(x_i) \pi^p(\theta_i) \geq 0 \text{ , for all } k = 1, 2.$$

Now, recall that $\pi^p(\theta) = \frac{\pi^c(\theta)}{\theta}$. Thus, the objective function can be written as:

$$\max_{\left\{\left(\lambda^k(x_i), \tau^k(x_i)\right)_{i=1}^2\right\}_{k=1}^2} \left\{ \sum_{i=1}^{2} \pi^p\big(\theta_i\big) \; \sum_{k=1}^{2} \left[u(x_i, y_k) - \tau^k(x_i)\right] \lambda^k(x_i) g(y_k) \right\}$$

Next, by rearranging terms from the objective function:

$$\begin{aligned}
&\sum_{i=1}^{2} \pi^p\big(\theta_i\big) \; \sum_{k=1}^{2} u(x_i, y_k) \lambda^k(x_i) g(y_k) - \sum_{i=1}^{2} \pi^p\big(\theta_i\big) \sum_{k=1}^{2} \tau^k(x_i) \lambda^k(x_i) g(y_k) \\
&\Rightarrow \sum_{i=1}^{2} \pi^p\big(\theta_i\big) \; \sum_{k=1}^{2} u(x_i, y_k) \lambda^k(x_i) g(y_k) \\
&- \Big[\sum_{i=1}^{2} \tau^1(x_i) \lambda^1(x_i) \pi^p\big(\theta_i\big) g(y_1) + \sum_{i=1}^{2} \tau^2(x_i) \lambda^2(x_i) \pi^p\big(\theta_i\big) g(y_2) \Big]
\end{aligned}$$

At the optimum, we know that the [PC] hold with equality (see Proof of Proposition 2):

$$\sum_{i=1}^{2} \tau^k(x_i) \lambda^k(x_i) \pi^p(\theta_i) = \sum_{i=1}^{2} c(x_i, y_k) \lambda^k(x_i) \pi^p(\theta_i) \tag{C.1}$$

Thus, by replacing Equation C.1 into the objective function, the optimization problem is:

$$\max_{\left\{\lambda^k(x_1),\lambda^k(x_2)\right\}_{k=1}^2} \left\{ \sum_{i=1}^{2} \pi^p(\theta_i) \sum_{k=1}^{2} \left[u(x_i,y_k) - c(x_i,y_k)\right]\lambda^k(x_i)g(y_k) \right\}$$

subject to feasibility constraints[FC] and market tightness[MT] defined above. The following corollary is immediate:

Corollary C.1. *In the first best, the randomization device* $\left\{\lambda^k(x_1), \lambda^k(x_2)\right\}_{k=1}^2$ *is independent of whether we consider interim or ex-post participation constraints.*

Proof. This follows from the fact that the objective function is independent of the transfers after incorporating the participation constraints. □

C.1 Proof of Lemma 1

For each (x, k), let $\lambda^k(x)$ be an arbitrary-feasible interior probability that generates a total welfare equal to:

$$W\left(\lambda^1(x_1), \lambda^2(x_1)\right) = \pi^p(\theta_1) \cdot \left[g(y_1)\, \lambda^1(x_1)\, S(x_1,y_1) + \left(1-g(y_1)\right)\, \lambda^2(x_1)\, S(x_1,y_2)\right]$$
$$+ \pi^p(\theta_2) \cdot \left[g(y_1)\, \left(1-\lambda^1(x_1)\right)\, S(x_2,y_1) + \left(1-g(y_1)\right)\, \left(1-\lambda^2(x_1)\right)\, S(x_2,y_2)\right]$$

where:

$$\theta_1 = \frac{g(y_1)\, \lambda^1(x_1) + \left(1-g(y_1)\right)\, \lambda^2(x_1)}{f(x_1)} \quad \text{and} \quad \theta_2 = \frac{g(y_1)\, \left(1-\lambda^1(x_1)\right) + \left(1-g(y_1)\right)\left(1-\lambda^2(x_1)\right)}{1-f(x_1)} \tag{C.2}$$

After trembling $\lambda^1(x_1)$ by ε_1 and $\lambda^2(x_1)$ by ε_2 such that $\varepsilon_2 \equiv -\frac{\varepsilon_1 g(y_1)}{1-g(y_1)}$, ensuring that the market tightness in each market remains constant, the new total welfare is:

$$W\left(\lambda^1(x_1)+\varepsilon_1, \lambda^2(x_1)+\varepsilon_2\right) = \pi^p(\theta_1) \cdot \left[g(y_1)\, \lambda^1(x_1)\, S(x_1,y_1) + \left(1-g(y_1)\right)\, \lambda^2(x_1)\, S(x_1,y_2)\right]$$
$$+ \pi^p(\theta_2) \cdot \left[g(y_1)\, \left(1-\lambda^1(x_1)\right)\, S(x_2,y_1) + \left(1-g(y_1)\right)\, \left(1-\lambda^2(x_1)\right)\, S(x_2,y_2)\right]$$
$$+ \varepsilon_1\, g(y_1)\, \left(\pi^p(\theta_2)\, \left[S(x_2,y_2) - S(x_2,y_1)\right] - \pi^p(\theta_1)\, \left[S(x_1,y_2) - S(x_1,y_1)\right]\right)$$

Thus, the change in welfare is equal to:

$$\begin{aligned}\Delta_W &= W\big(\lambda^1(x_1)+\varepsilon_1, \lambda^2(x_1)+\varepsilon_2\big) - W\big(\lambda^1(x_1), \lambda^2(x_1)\big) \\ &= \varepsilon_1\, g(y_1)\, \underbrace{\Big(\pi^p(\theta_2)\,\big[S(x_2,y_2)-S(x_2,y_1)\big] - \pi^p(\theta_1)\,\big[S(x_1,y_2)-S(x_1,y_1)\big]\Big)}_{Z^{CI}(\theta_1)}\end{aligned}$$

where θ_1 and θ_2 are defined as in Equation C.2. Note that, $\theta_2 = \frac{1-f(x_1)\theta_1}{1-f(x_1)}$, thus Z^{CI} can be written as a function of only θ_1. It is easy to see that $Z^{CI}(\theta_1)$ is strictly increasing in θ_1. Therefore, $Z^{CI}(\theta_1^{\max}) \geq Z^{CI}(\theta_1) \geq Z^{CI}(0)$ for any $\theta_1 \in [0, \theta_1^{\max}]$ where $\theta_1^{\max} = \frac{1}{f(x_1)}$. Now, we analyze three cases:

1. Suppose $Z^{CI}(\theta_1) > 0$. Then, pick $\varepsilon_1 > 0$ with $\varepsilon_2 = -\frac{\varepsilon_1 g(y_1)}{1-g(y_1)}$ such that either $\hat{\lambda}^1(x_1) \equiv \lambda^1(x_1) + \varepsilon_1 = 1$ or $\hat{\lambda}^2(x_1) \equiv \lambda^2(x_1) + \varepsilon_2 = 0$. In the former case, $\hat{\lambda}^1(x_2) = 0$ and $\hat{\lambda}^2(x_2) \in (0,1)$; and in the latter case, $\hat{\lambda}^1(x_2) \in (0,1)$ and $\hat{\lambda}^2(x_2) = 1$. In both cases, the definition of PAM is satisfied.

2. Suppose $Z^{CI}(\theta_1) < 0$. Then, pick $\varepsilon_1 < 0$ with $\varepsilon_2 = -\frac{\varepsilon_1 g(y_1)}{1-g(y_1)}$ such that either $\hat{\lambda}^1(x_1) \equiv \lambda^1(x_1) + \varepsilon_1 = 0$ or $\hat{\lambda}^2(x_1) \equiv \lambda^2(x_1) + \varepsilon_2 = 1$. In the former case, $\hat{\lambda}^1(x_2) = 1$ and $\hat{\lambda}^2(x_2) \in (0,1)$; and in the latter case, $\hat{\lambda}^1(x_2) \in (0,1)$ and $\hat{\lambda}^2(x_2) = 0$. In both cases, the definition of NAM is satisfied.

3. Suppose $Z^{CI}(\theta_1) = 0$. We show that an interior randomization device can not be an equilibrium. To see this, first tremble $\lambda^1(x_1)$ by ε_1, and calculate welfare:

$$\begin{aligned}W\big(\lambda^1(x_1)+\varepsilon_1, \lambda^2(x_1)\big) &= \pi^p(\hat{\theta}_1)\cdot\Big[g(y_1)\,\lambda^1(x_1)\,S(x_1,y_1)+\big(1-g(y_1)\big)\,\lambda^2(x_1)\,S(x_1,y_2)\Big] \\ &+ \pi^p(\hat{\theta}_2)\cdot\Big[g(y_1)\,\big(1-\lambda^1(x_1)\big)\,S(x_2,y_1)+\big(1-g(y_1)\big)\,\big(1-\lambda^2(x_1)\big)\,S(x_2,y_2)\Big] \\ &+ \varepsilon_1 g(y_1)\Big[\pi^p(\hat{\theta}_1)\,S(x_1,y_1)-\pi^p(\hat{\theta}_2)\,S(x_2,y_1)\Big]\end{aligned}$$

where $\hat{\theta}_1 = \theta_1 + \frac{\varepsilon_1 g(y_1)}{f(x_1)}$, $\hat{\theta}_2 = \theta_2 - \frac{\varepsilon_1 g(y_1)}{1-f(x_1)}$, and θ_1, θ_2 are defined as in Equation C.2. Now, let's tremble $\lambda^2(x_1)$ by ε_2, and calculate welfare:

$$\begin{aligned}W(\lambda^2(x_1), \lambda^2(x_1)+\varepsilon_2) &= \pi^p(\tilde{\theta}_1)\cdot\Big[g(y_1)\,\lambda^1(x_1)\,S(x_1,y_1)+\big(1-g(y_1)\big)\,\lambda^2(x_1)\,S(x_1,y_2)\Big] \\ &+ \pi^p(\tilde{\theta}_2)\cdot\Big[g(y_1)\,\big(1-\lambda^1(x_1)\big)\,S(x_2,y_1)+\big(1-g(y_1)\big)\,\big(1-\lambda^2(x_1)\big)\,S(x_2,y_2)\Big] \\ &+ \varepsilon_2\big(1-g(y_1)\big)\Big[\pi^p(\tilde{\theta}_1)\,S(x_1,y_2)-\pi^p(\tilde{\theta}_2)\,S(x_2,y_2)\Big]\end{aligned}$$

where $\tilde{\theta}_1 = \theta_1 + \frac{\varepsilon_2(1-g(y_1))}{f(x_1)}$, $\tilde{\theta}_2 = \theta_2 - \frac{\varepsilon_2(1-g(y_1))}{1-f(x_1)}$, and θ_1, θ_2 are defined as in Equation C.2.

For any small ε_1 with $\varepsilon_2 \equiv \frac{\varepsilon_1 g(y_1)}{1-g(y_1)}$, it follows that $\hat{\theta}_1 = \tilde{\theta}_1$ and $\hat{\theta}_2 = \tilde{\theta}_2$. Pick such ε_2. Then, increasing $\lambda^1(x_1)$ is marginally more profitable than increasing $\lambda^2(x_1)$ if and only if

$$\underbrace{\pi^p(\hat{\theta}_2) \cdot \big[S(x_2, y_2) - S(x_2, y_1)\big] - \pi^p(\hat{\theta}_1) \cdot \big[S(x_1, y_2) - S(x_1, y_1)\big]}_{Z^{CI}(\hat{\theta}_1)} \geq 0$$

Since $Z^{CI}(\hat{\theta}_1) > Z^{CI}(\theta_1) = 0$, then the inequality holds. Therefore, at least one of the partial derivatives of W at $\big(\lambda^1(x_1), \lambda^2(x_1)\big)$ is non-zero, meaning that $\big(\lambda^1(x_1), \lambda^2(x_1)\big)$ at $Z^{CI}(\theta_1) = 0$ is not an equilibrium. This finishes the proof.

C.2 Proof of Proposition 1

By assumption $S(x, y)$ is increasing in y, thus $Z^{CI}(\theta_1)$ is increasing in θ_1. Therefore, items (i) to (iii) from the previous proof of Lemma 1 apply here.

C.3 Proof of Corollary 1

Notice that, $Z^{CI}(\theta_1)$ is increasing in θ_1 reaching its minimum value at $\theta_1 = 0$, and when $\theta_1 = 0$ it follows that $\pi^p(0) = 1$ and $\theta_2 = \frac{1}{1-f(x_1)}$. Therefore, from Proposition 1, we can ensure PAM by imposing that the following inequality must hold:

$$\pi^p\Big(\frac{1}{1 - f(x_1)}\Big) \cdot \big[S(x_2, y_2) - S(x_2, y_1)\big] - \big[S(x_1, y_2) - S(x_1, y_1)\big] \geq 0$$

Now, $Z^{CI}(\theta_1)$ reaches its maximum value at $\theta_1 = \frac{1}{f(x_1)}$.Therefore, from Proposition 1, we can ensure NAM by imposing that the following inequality must hold:

$$\big[S(x_2, y_2) - S(x_2, y_1)\big] - \pi^p\Big(\frac{1}{f(x_1)}\Big) \cdot \big[S(x_1, y_2) - S(x_1, y_1)\big] \leq 0$$

C.4 Assortative Matching in Equilibrium

Under the light of the results above, we can characterize the equilibrium sorting patterns. We will start by providing some auxiliary lemmas.

Lemma C.1. *The rate of change in Welfare $W\big(\lambda^1(x_1), \lambda^2(x_1)\big)$ monotonically decreases in $\lambda^k(x_1)$ for each $k = 1, 2$.*

Proof. Recall the total welfare:

$$W\big(\lambda^1(x_1),\lambda^2(x_1)\big) = \pi^p(\theta_1)\cdot \underbrace{\Big[g(y_1)\ \lambda^1(x_1)\ S(x_1,y_1) + \big(1-g(y_1)\big)\ \lambda^2(x_1)\ S(x_1,y_2)\Big]}_{\mathbb{E}U_1}$$
$$+\ \pi^p(\theta_2)\cdot \underbrace{\Big[g(y_1)\ \big(1-\lambda^1(x_1)\big)\ S(x_2,y_1) + \big(1-g(y_1)\big)\ \big(1-\lambda^2(x_1)\big)\ S(x_2,y_2)\Big]}_{\mathbb{E}U_2}$$

where

$$\theta_1 = \frac{g(y_1)\ \lambda^1(x_1) + \big(1-g(y_1)\big)\ \lambda^2(x_1)}{f(x_1)} \ \text{ and } \ \theta_2 = \frac{1-\theta_1 f(x_1)}{1-f(x_1)}$$

Fix $\lambda^{-k}(x_1)$. Increasing $\lambda^k(x_1)$ by a small amount $\varepsilon > 0$, increases $\mathbb{E}U_1$ and θ_1 linearly, and decreases $\mathbb{E}U_2$ and θ_2 linearly. Recall that, $\pi^p(\cdot)$ is a decreasing and convex function, thus the rate of increase through $\pi^p(\theta_1)\cdot\mathbb{E}U_1$ decreases, while the rate of decrease through $\pi^p(\theta_2)\cdot\mathbb{E}U_2$ increases in $\lambda^k(x_1)$, for any $k=1,2$. □

Lemma C.1 is useful since it implies that $\frac{\partial W\left(\lambda^1(x_1),\lambda^2(x_1)\right)}{\partial\lambda^k(x_1)}$ is monotonically decreasing. Thus, if it is zero at some $\lambda^{k\prime}(x_1)$, then it is negative at any $\lambda^k(x_1)$ if and only if $\lambda^k(x_1) > \lambda^{k\prime}(x_1)$ for any $\lambda^{-k}(x_1)$. Note that the same analysis applies to any pair $(\lambda^1(x_1),\lambda^2(x_1))$ that yields the same market tightness. Now, another useful lemma follows:

Lemma C.2. *Fix $(\hat\lambda^1(x_1),\hat\lambda^2(x_1))$. For any $(\tilde\lambda^1(x_1),\tilde\lambda^2(x_1))$ such that $\theta_1(\hat\lambda^1(x_1),\hat\lambda^2(x_1)) = \theta_1(\tilde\lambda^1(x_1),\tilde\lambda^2(x_1))$ and $\hat\lambda^1(x_1)\geq\tilde\lambda^1(x_1)$, the following holds:*

$$\frac{\partial W(\lambda^1(x_1),\lambda^2(x_1))}{\partial\lambda^k(x_1)}\Big|_{(\hat\lambda^1(x_1),\hat\lambda^2(x_1))} \leq \frac{\partial W(\lambda^1(x_1),\lambda^2(x_1))}{\partial\lambda^k(x_1)}\Big|_{(\tilde\lambda^1(x_1),\tilde\lambda^2(x_1))}$$

Proof. Taking partial derivatives on welfare yields the followings:

$$\frac{\partial W(\lambda^1(x_1),\lambda^2(x_1))}{\partial\lambda^1(x_1)} =$$
$$g(y_1)\cdot V(\lambda^1(x_1),\lambda^2(x_1)) + g(y_1)\cdot\Big[\pi^p(\theta_1)S(x_1,y_1) - \pi^p(\theta_2)S(x_2,y_1)\Big]$$

$$\frac{\partial W(\lambda^1(x_1),\lambda^2(x_1))}{\partial\lambda^2(x_1)} =$$
$$\big(1-g(y_1)\big)\cdot V(\lambda^1(x_1),\lambda^2(x_1)) + \big(1-g(y_1)\big)\cdot\Big[\pi^p(\theta_1)S(x_1,y_2) - \pi^p(\theta_2)S(x_2,y_2)\Big]$$

with

$$V(\lambda^1(x_1),\lambda^2(x_1)) = \frac{\pi^{p\prime}(\theta_1)}{f(x_1)} \cdot \mathbb{E}U_1(\lambda^1(x_1),\lambda^2(x_1)) - \frac{\pi^{p\prime}(\theta_2)}{1-f(x_1)} \cdot \mathbb{E}U_2(\lambda^1(x_1),\lambda^2(x_1))$$

where $\mathbb{E}U_1(\lambda^1(x_1),\lambda^2(x_1))$ and $\mathbb{E}U_2(\lambda^1(x_1),\lambda^2(x_1))$ are defined as in Lemma C.1. It is easy to verify that $V(\lambda^1(x_1),\lambda^2(x_1))$ decreases as we move down on the market tightness θ_1, that is, as we increase $\lambda^1(x_1)$ while decreasing $\lambda^2(x_1)$. This implies that the rate of change with respect to $\lambda^1(x_1)$ decreases as one moves down on the same market tightness, which finishes the proof. □

Now, by using Lemmas C.1 and C.2, we can characterize the equilibrium allocation of parents across submarket step by step. We establish the equilibrium allocation of parents when the sufficient conditions of Corollary 1 hold. Then we leave the analysis for the case where the sufficient conditions are violated to the readers.

Proposition C.1 (**Positive Assortative Matching (PAM)**). *Suppose* $\frac{S(x_2,y_2)-S(x_2,y_1)}{S(x_1,y_2)-S(x_1,y_1)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)}$ *holds. The equilibrium sorting exhibits:*

i. low-type PAM with $\lambda^{1\star}(x_1) \in (0,1)$ and $\lambda^{2\star}(x_1) = 0$ if

$$\frac{\partial W\left(\lambda^1(x_1),\lambda^2(x_1)\right)}{\partial \lambda^1(x_1)}|_{\{\lambda^2(x_1)=0\}} = 0 \text{ for some } \lambda^{1\star}(x_1) \in (0,1) \tag{C.3}$$

ii. perfect PAM with $\lambda^{1\star}(x_1) = 1$ and $\lambda^{2\star}(x_1) = 0$ if

$$\frac{\partial W\left(\lambda^1(x_1),\lambda^2(x_1)\right)}{\partial \lambda^1(x_1)}|_{\{\lambda^2(x_1)=1,\lambda^2(x_1)=0\}} \geq 0 \geq \frac{\partial W\left(\lambda^1(x_1),\lambda^2(x_1)\right)}{\partial \lambda^2(x_1)}|_{\{\lambda^1(x_1)=1,\lambda^2(x_1)=0\}} \tag{C.4}$$

iii. high-type PAM with $\lambda^{1\star}(x_1) = 1$ and $\lambda^{2\star}(x_1) \in (0,1)$ if

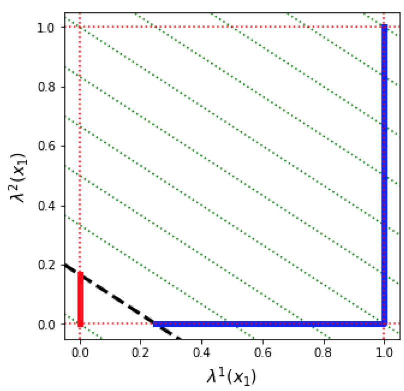
$$\frac{\partial W\left(\lambda^1(x_1),\lambda^2(x_1)\right)}{\partial \lambda^2(x_1)}|_{\{\lambda^1(x_1)=1\}} = 0 \text{ for some } \lambda^{2\star}(x_1) \in (0,1) \tag{C.5}$$

Proof. By assumption, $\frac{S(x_2,y_2)-S(x_2,y_1)}{S(x_1,y_2)-S(x_1,y_1)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)}$, which implies that $Z^{CI}(\theta_1) \geq 0$ for any θ_1. Therefore, starting from an initial allocation $\lambda^1(x_1) = 0$ and $\lambda^2(x_1) = 0$, the designer first allocates y_1-parents into submarket x_1 until either parents are exhausted or it is not profitable anymore. Accordingly, perfect PAM and high-type PAM follows. □

Proposition C.1 characterizes the equilibrium sorting patterns when the conditions specified in Corollary 1(i) hold. One can easily characterize the equilibrium distribution of parents across submarkets for NAM and extend it to the case where the sufficient conditions are violated, with a parallel argument. As is represented by the dashed-lines in Figure C.1, suppose that $Z^{CI}(\bar{\theta}_1) = 0$ for some $\bar{\theta}_1 \in (0, {}^1/_{f(x_1)})$. Then, for NAM (red-lines in Figure C.1), there exists either, **(i)** $\tilde{\lambda}^1(x_1) = 0$ and $\tilde{\lambda}^2(x_1) \leq 1$ or **(ii)** $\tilde{\lambda}^1(x_1) > 0$ and $\tilde{\lambda}^2(x_1) = 1$, with $\bar{\theta}_1 = \frac{g(y_1)\tilde{\lambda}^1(x_1)+(1-g(y_1))\tilde{\lambda}^2(x_1)}{f(x_1)}$. Similarly, for PAM (blue-lines in Figure C.1), there exists either, **(i)** $\hat{\lambda}^1(x_1) \leq 1$ and $\hat{\lambda}^2(x_1) = 0$ or **(ii)** $\hat{\lambda}^1(x_1) = 1$ and $\hat{\lambda}^2(x_1) \geq 0$, with $\bar{\theta}_1 = \frac{g(y_1)\hat{\lambda}^1(x_1)+(1-g(y_1))\hat{\lambda}^2(x_1)}{f(x_1)}$. In what follows, we study each possible case illustrated in Figure C.1.

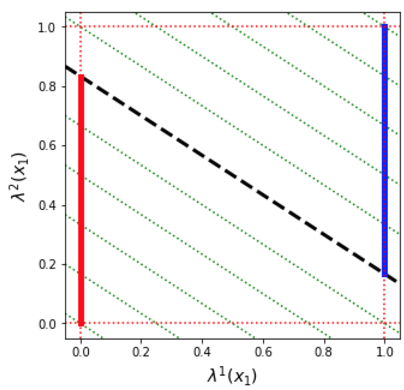
Figure C.1: Possible Cases given $Z^{CI}(\bar{\theta}_1)$

(**a**) Case 1A



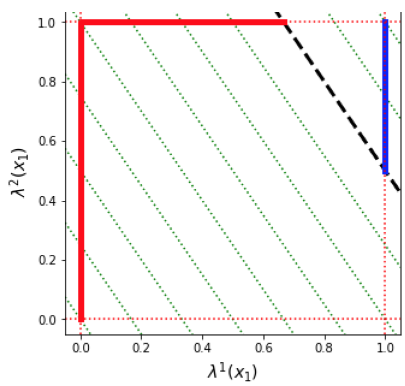
$\tilde{\lambda}^1(x_1) = 0$ and $\tilde{\lambda}^2(x_1) \leq 1$
$\hat{\lambda}^1(x_1) \leq 1$ and $\hat{\lambda}^2(x_1) = 0$

(**b**) Case 1B



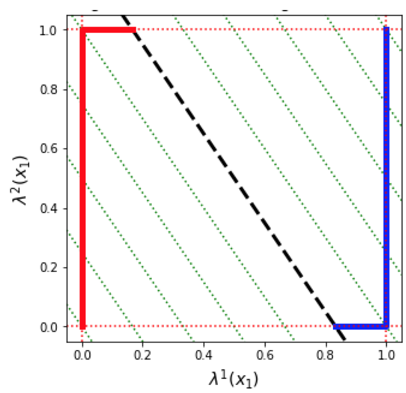
$\tilde{\lambda}^1(x_1) = 0$ and $\tilde{\lambda}^2(x_1) \leq 1$
$\hat{\lambda}^1(x_1) = 1$ and $\hat{\lambda}^2(x_1) \geq 0$

(**c**) Case 2A



$\tilde{\lambda}^1(x_1) \geq 0$ and $\tilde{\lambda}^2(x_1) = 1$
$\hat{\lambda}^1(x_1) = 1$ and $\hat{\lambda}^2(x_1) \geq 0$

(**d**) Case 2B



$\tilde{\lambda}^1(x_1) \geq 0$ and $\tilde{\lambda}^2(x_1) = 1$
$\hat{\lambda}^1(x_1) \leq 1$ and $\hat{\lambda}^2(x_1) = 0$

C.5 Proof of Proposition 2

The designer solves the following problem:

$$\max_{\left\{\left(\lambda^k(x_i),\tau^k(x_i)\right)_{i=1}^2\right\}_{k=1}^2} \left\{\sum_{i=1}^{2} \pi^c(\theta_i)\ \frac{\sum_{k=1}^2 \left[u(x_i,y_k)-\tau^k(x_i)\right]\lambda^k(x_i)g(y_k)}{\sum_{k=1}^2 \lambda^k(x_i)\ g(y_k)}\ f(x_i)\right\}$$

subject to [FC], [MT], and [PC]. We will show that at the optimal solution, the participation constraints hold with equality. By contradiction, suppose that for some license k, the [PC] holds with strict inequality at the optimum:

$$\sum_{i=1}^{2} \tau^k(x_i)\lambda^k(x_i)\pi^p(\theta_i) > \sum_{i=1}^{2} c(x_i,y_k)\lambda^k(x_i)\pi^p(\theta_i)$$

Then, the designer can decrease $\tau^k(x_1)$ and $\tau^k(x_2)$ by a small $\varepsilon > 0$ satisfying the constraint while increasing the objective function. A contradiction. Therefore, the optimal transfers can be pinned-down by the [PC] which hold with equality.

D Appendix: Analysis of Private Information

First, it is useful to understand who has incentives to mimic whom under the first best menu of licenses. Recall the incentive compatibility constraint [IC] for $k \neq k' = 1, 2$:

$$\sum_{i=1}^{2} \big[\tau^k(x_i) \; - \; c(x_i, y_k)\big] \lambda^k(x_i)\pi^p(\theta_i) \;\; \geq \;\; \sum_{i=1}^{2} \big[\tau^{k'}(x_i) \; - \; c(x_i, y_k)\big] \lambda^{k'}(x_i)\pi^p(\theta_i)$$

and the participation constraint [PC] for $k = 1, 2$:

$$\sum_{i=1}^{2} \Big[\tau^k(x_i) - c(x_i, y_k)\Big] \lambda^k(x_i)\pi^p(\theta_i) \geq 0$$

In the complete information case, [PC]s holds with equality. Now, plugging [PC](k) and [PC](k') into [IC](k) yields the following inequality:

$$0 \geq \Big[c(x_1, y_{k'}) - c(x_1, y_k)\Big] \lambda^{k'}(x_1)\pi^p(\theta_1) + \Big[c(x_2, y_{k'}) - c(x_2, y_k)\Big] \lambda^{k'}(x_2)\pi^p(\theta_2)$$

Since $c(x, y)$ is decreasing in y, the inequality holds for $k = 1$ but not for $k = 2$. Thus, under the first best, type-y_2 parents have incentives to mimic type-y_1 parents.

Next, we know that the [IC] for high-ability and the [PC] for low-ability parents hold with equality in equilibrium (see Proof of Proposition 4):

$$[\text{PC1}] \quad \sum_{i=1}^{2} \Big[\tau^1(x_i) - c(x_i, y_1)\Big] \lambda^1(x_i)\pi^p(\theta_i) = 0$$

$$[\text{IC2}] \quad \sum_{i=1}^{2} \big[\tau^2(x_i) - c(x_i, y_2)\big] \lambda^2(x_i)\pi^p(\theta_i) = \sum_{i=1}^{2} \big[\tau^1(x_i) - c(x_i, y_2)\big] \lambda^1(x_i)\pi^p(\theta_i)$$

Replacing [PC1] in [IC2]:

$$\tau^2(x_1)\lambda^2(x_1)\pi^p(\theta_1) + \tau^2(x_2)\lambda^2(x_2)\pi^p(\theta_2) = c(x_1, y_2)\lambda^2(x_1)\pi^p(\theta_1) + c(x_2, y_2)\lambda^2(x_2)\pi^p(\theta_2) + \Big[c(x_1, y_1) - c(x_1, y_2)\Big] \lambda^1(x_1)\pi^p(\theta_1) + \Big[c(x_2, y_1) - c(x_2, y_2)\Big] \lambda^1(x_2)\pi^p(\theta_2)$$

Now, replacing the restrictions into the objective function, the designer solves:

$$\max_{\{\lambda^k(x_1),\lambda^k(x_2)\}_{k=1}^2} \Bigg\{ \sum_{i=1}^{2} \pi^p(\theta_i) \left[\sum_{k=1}^{2} \Big(\underbrace{u(x_i,y_k) - c(x_i,y_k)}_{S(x,y)} \Big) \lambda^k(x_i)\, g(y_k) \right]$$

$$- \Big[c(x_1,y_1) - c(x_1,y_2)\Big] \lambda^1(x_1) \pi^p(\theta_1) g(y_2) - \Big[c(x_2,y_1) - c(x_2,y_2)\Big] \lambda^1(x_2)\pi^p(\theta_2) g(y_2) \Bigg\}$$

subject to [FC], [MT], and

$$[\text{AC}] = \begin{cases} \frac{c(x_2,y_2)-c(x_2,y_1)}{c(x_1,y_2)-c(x_1,y_1)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)} & \text{if } \lambda^2(x_2) > \lambda^1(x_2) \\ \frac{c(x_1,y_2)-c(x_1,y_1)}{c(x_2,y_2)-c(x_2,y_1)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_1)}\right)} & \text{if } \lambda^2(x_2) < \lambda^1(x_2) \end{cases} \tag{D.1}$$

This additional constraint [AC] ensures that the [IC] for low-ability parents is satisfied when the [IC] for high ability parents holds (see Proof of Proposition 4).

Corollary D.1. *With private information, the randomization device* $\left\{\lambda^k(x_1), \lambda^k(x_2)\right\}_{k=1}^2$ *is independent of whether interim or ex-post participation constraints are implemented.*

Proof. This follows from the fact that the objective function is independent of the transfers after incorporating the participation constraints. □

D.1 Proof of Lemma 1 under Private Information

We can establish Lemma 1 for the private information case.

Lemma D.1. *In the private information setting, for at least one of the licenses, the optimal randomization rule yields a corner solution.*

For each (x,k), let $\lambda^k(x_1) \in (0,1)$ be an arbitrary-feasible interior probability that generates a total welfare equal to:

$$\hat{W}\big(\lambda^1(x_1), \lambda^2(x_1)\big) = \pi^p(\theta_1) \cdot \Big[g(y_1)\, \lambda^1(x_1)\, S(x_1,y_1) + \big(1-g(y_1)\big)\, \lambda^2(x_1)\, S(x_1,y_2)\Big]$$

$$+ \pi^p(\theta_2) \cdot \Big[g(y_1)\, \big(1-\lambda^1(x_1)\big)\, S(x_2,y_1) + \big(1-g(y_1)\big)\, \big(1-\lambda^2(x_1)\big)\, S(x_2,y_2)\Big]$$

$$- \Big[c(x_1,y_1) - c(x_1,y_2)\Big] \lambda^1(x_1)\pi^p(\theta_1) g(y_2) - \Big[c(x_2,y_1) - c(x_2,y_2)\Big] \big(1-\lambda^1(x_1)\big)\pi^p(\theta_2) g(y_2)$$

where:

$$\theta_1 = \frac{g(y_1)\, \lambda^1(x_1) + \big(1-g(y_1)\big)\, \lambda^2(x_1)}{f(x_1)} \quad \text{and} \quad \theta_2 = \frac{g(y_1)\, \big(1-\lambda^1(x_1)\big) + \big(1-g(y_1)\big)\big(1-\lambda^2(x_1)\big)}{1-f(x_1)} \tag{D.2}$$

As in the complete information, we tremble $\lambda^1(x_1)$ by ε_1 and $\lambda^2(x_1)$ by ε_2 such that $\varepsilon_2 \equiv -\frac{\varepsilon_1 g(y_1)}{1-g(y_1)}$ ensuring that the market tightness in each submarket remains constant. The new total welfare is:

$$\begin{aligned}
\hat{W}\big(\lambda^1(x_1)+\varepsilon_1, \lambda^2(x_1)+\varepsilon_2\big) &= \pi^p(\theta_1)\cdot\Big[g(y_1)\,\lambda^1(x_1)\,S(x_1,y_1)+\big(1-g(y_1)\big)\,\lambda^2(x_1)\,S(x_1,y_2)\Big] \\
&+ \pi^p(\theta_2)\cdot\Big[g(y_1)\,\big(1-\lambda^1(x_1)\big)\,S(x_2,y_1)+\big(1-g(y_1)\big)\,\big(1-\lambda^2(x_1)\big)\,S(x_2,y_2)\Big] \\
-\Big[c(x_1,y_1)-c(x_1,y_2)\Big]&\lambda^1(x_1)\pi^p(\theta_1)g(y_2)-\Big[c(x_2,y_1)-c(x_2,y_2)\Big]\big(1-\lambda^1(x_1)\big)\pi^p(\theta_2)g(y_2) \\
&+\varepsilon_1 g(y_1)\Bigg\{\pi^p(\theta_2)\Big[S(x_2,y_2)-S(x_2,y_1)+\frac{g(y_2)}{g(y_1)}\Big(c(x_2,y_1)-c(x_2,y_2)\Big)\Big] \\
&\qquad -\pi^p(\theta_1)\Big[S(x_1,y_2)-S(x_1,y_1)+\frac{g(y_2)}{g(y_1)}\Big(c(x_1,y_1)-c(x_1,y_2)\Big)\Big]\Bigg\}
\end{aligned}$$

Thus, the change in welfare is equal to:

$$\Delta_{\hat{W}} = \varepsilon_1 g(y_1)\underbrace{\Bigg\{\pi^p(\theta_2)\Big[S(x_2,y_2)-S(x_2,y_1)+\frac{g(y_2)}{g(y_1)}\Big(c(x_2,y_1)-c(x_2,y_2)\Big)\Big] -\pi^p(\theta_1)\Big[S(x_1,y_2)-S(x_1,y_1)+\frac{g(y_2)}{g(y_1)}\Big(c(x_1,y_1)-c(x_1,y_2)\Big)\Big]\Bigg\}}_{Z^{PI}(\theta_1)}$$

where θ_1 and θ_2 are defined as in Equation D.2. As earlier, $Z^{PI}(\theta_1)$ is strictly increasing in θ_1. Therefore, $Z^{PI}(\theta_1^{\max}) \geq Z^{PI}(\theta_1) \geq Z^{PI}(0)$ for any $\theta_1 \in [0, \theta_1^{\max}]$ where $\theta_1^{\max} = \frac{1}{f(x_1)}$. Now, we analyze three cases:

1. Suppose $Z^{PI}(\theta_1) > 0$. Then, pick $\varepsilon_1 > 0$ with $\varepsilon_2 \equiv -\frac{\varepsilon_1 g(y_1)}{1-g(y_1)}$ such that either $\hat{\lambda}^1(x_1) \equiv \lambda^1(x_1)+\varepsilon_1 = 1$ or $\hat{\lambda}^2(x_1) \equiv \lambda^2(x_1)+\varepsilon_2 = 0$. In the former case, $\hat{\lambda}^1(x_2) = 0$ and $\hat{\lambda}^2(x_2) \in (0,1)$; and in the latter case, $\hat{\lambda}^1(x_2) \in (0,1)$ and $\hat{\lambda}^2(x_2) = 1$. In both cases, the definition of PAM is satisfied.

2. Suppose $Z^{PI}(\theta_1) < 0$. Then, pick $\varepsilon_1 < 0$ with $\varepsilon_2 \equiv -\frac{\varepsilon_1 g(y_1)}{1-g(y_1)}$ such that either $\hat{\lambda}^1(x_1) \equiv \lambda^1(x_1)+\varepsilon_1 = 0$ or $\hat{\lambda}^2(x_1) \equiv \lambda^2(x_1)+\varepsilon_2 = 1$. In the former case, $\hat{\lambda}^1(x_2) = 1$ and $\hat{\lambda}^2(x_2) \in (0,1)$; and in the latter case, $\hat{\lambda}^1(x_2) \in (0,1)$ and $\hat{\lambda}^2(x_2) = 0$. In both cases, the definition of NAM is satisfied.

3. Suppose $Z^{PI}(\theta) = 0$. We show that an interior randomization device can not be an equilibrium. To see this, first tremble $\lambda^1(x_1)$ by ε_1, and calculate

welfare:

$$\begin{aligned}
\hat{W}\big(\lambda^1(x_1)+\varepsilon_1, \lambda^2(x_1)\big) &= \pi^p(\hat{\theta}_1)\cdot\Big[g(y_1)\,\lambda^1(x_1)\,S(x_1,y_1)+\big(1-g(y_1)\big)\,\lambda^2(x_1)\,S(x_1,y_2)\Big] \\
&+ \pi^p(\hat{\theta}_2)\cdot\Big[g(y_1)\,\big(1-\lambda^1(x_1)\big)\,S(x_2,y_1)+\big(1-g(y_1)\big)\,\big(1-\lambda^2(x_1)\big)\,S(x_2,y_2)\Big] \\
-\Big[c(x_1,y_1)-c(x_1,y_2)\Big]&\lambda^1(x_1)\pi^p(\hat{\theta}_1)g(y_2)-\Big[c(x_2,y_1)-c(x_2,y_2)\Big]\big(1-\lambda^1(x_1)\big)\pi^p(\hat{\theta}_2)g(y_2) \\
&+ \varepsilon_1 g(y_1)\Big[\pi^p(\hat{\theta}_1)\,S(x_1,y_1)-\pi^p(\hat{\theta}_2)\,S(x_2,y_1)\Big] \\
&+ \varepsilon_1 g(y_2)\Big\{\pi^p(\hat{\theta}_2)\Big[c(x_2,y_1)-c(x_2,y_2)\Big]-\pi^p(\hat{\theta}_1)\Big[c(x_1,y_1)-c(x_1,y_2)\Big]\Big\}
\end{aligned}$$

where $\hat{\theta}_1 = \theta_1 + \frac{\varepsilon_1 g(y_1)}{f(x_1)}$, $\hat{\theta}_2 = \theta_2 - \frac{\varepsilon_1 g(y_1)}{1-f(x_1)}$, and θ_1, θ_2 are defined as in Equation D.2. Now, let's tremble $\lambda^2(x_1)$ by ε_2, and calculate welfare:

$$\begin{aligned}
\hat{W}(\lambda^2(x_1), \lambda^2(x_1)+\varepsilon_2) &= \pi^p(\tilde{\theta}_1)\cdot\Big[g(y_1)\,\lambda^1(x_1)\,S(x_1,y_1)+\big(1-g(y_1)\big)\,\lambda^2(x_1)\,S(x_1,y_2)\Big] \\
&+ \pi^p(\tilde{\theta}_2)\cdot\Big[g(y_1)\,\big(1-\lambda^1(x_1)\big)\,S(x_2,y_1)+\big(1-g(y_1)\big)\,\big(1-\lambda^2(x_1)\big)\,S(x_2,y_2)\Big] \\
-\Big[c(x_1,y_1)-c(x_1,y_2)\Big]&\lambda^1(x_1)\pi^p(\tilde{\theta}_1)g(y_2)-\Big[c(x_2,y_1)-c(x_2,y_2)\Big]\big(1-\lambda^1(x_1)\big)\pi^p(\tilde{\theta}_2)g(y_2) \\
&+ \varepsilon_2\big(1-g(y_1)\big)\Big[\pi^p(\tilde{\theta}_1)\,S(x_1,y_2)-\pi^p(\tilde{\theta}_2)\,S(x_2,y_2)\Big]
\end{aligned}$$

where $\tilde{\theta}_1 = \theta_1 + \frac{\varepsilon_2(1-g(y_1))}{f(x_1)}$, $\tilde{\theta}_2 = \theta_2 - \frac{\varepsilon_2(1-g(y_1))}{1-f(x_1)}$, and θ_1, θ_2 are defined as in Equation D.2.

For any small ε_1 with $\varepsilon_2 \equiv \frac{\varepsilon_1 g(y_1)}{1-g(y_1)}$, it follows that $\hat{\theta}_1 = \tilde{\theta}_1$ and $\hat{\theta}_2 = \tilde{\theta}_2$. Pick such ε_2. Then, increasing $\lambda^1(x_1)$ is marginally more profitable than increasing $\lambda^2(x_1)$ if and only if

$$\begin{aligned}
&\pi^p(\theta_2)\Big[S(x_2,y_2)-S(x_2,y_1)+\frac{g(y_2)}{g(y_1)}\Big(c(x_2,y_1)-c(x_2,y_2)\Big)\Big] \\
&\underbrace{-\pi^p(\theta_1)\Big[S(x_1,y_2)-S(x_1,y_1)+\frac{g(y_2)}{g(y_1)}\Big(c(x_1,y_1)-c(x_1,y_2)\Big)\Big]}_{Z^{PI}(\theta_1)} \geq 0
\end{aligned}$$

Since $Z^{PI}(\hat{\theta}_1) > Z^{PI}(\theta_1) = 0$, then the inequality holds. Therefore, at least one of the partial derivatives of W at $\big(\lambda^1(x_1), \lambda^2(x_1)\big)$ is non-zero, meaning that $\big(\lambda^1(x_1), \lambda^2(x_1)\big)$ at $Z^{PI}(\theta_1) = 0$ is not an equilibrium. This finishes the proof.

D.2 Proof of Proposition 1 under Private Information

We can establish Proposition 1 for the private information case. Let $\widehat{\theta}_1$ be such that $Z^{PI}(\widehat{\theta}_1) = 0$, then the following result holds:

Proposition D.1. *In the private information setting, let θ_1^{**} be the equilibrium market tightness. **(i)** If $\theta_1^{**} > \widehat{\theta}_1$ then the equilibrium sorting exhibits PAM. **(ii)** If $\theta_1^{**} < \widehat{\theta}_1$ then the equilibrium sorting exhibits NAM. **(iii)** $\theta_1^{**} = \widehat{\theta}_1$ is never optimal.*

By assumption $S(x, y)$ is increasing in y, thus $Z^{PI}(\theta_1)$ is increasing in θ_1. Therefore, items (i) to (iii) from the previous proof applies here.

D.3 Proof of Corollary 2

$Z^{PI}(\theta_1)$ is increasing in θ_1 reaching its minimum value at $\theta_1 = 0$, and when $\theta_1 = 0$ it follows that $\pi^p(0) = 1$ and $\theta_2 = \frac{1}{1-f(x_1)}$. Therefore, from Proposition D.1, we can ensure PAM by imposing that the following inequality must hold:

$$\pi^p\left(\frac{1}{f(x_2)}\right)\Big[S(x_2,y_2) - S(x_2,y_1) + \frac{g(y_2)}{g(y_1)}\Big(c(x_2,y_1) - c(x_2,y_2)\Big)\Big] \\ - \Big[S(x_1,y_2) - S(x_1,y_1) + \frac{g(y_2)}{g(y_1)}\Big(c(x_1,y_1) - c(x_1,y_2)\Big)\Big] \geq 0$$

Now, $Z^{PI}(\theta_1)$ reaches its maximum value at $\theta_1 = \frac{1}{f(x_1)}$.Therefore, from Proposition D.1, we can ensure NAM by imposing that the following inequality must hold:

$$\Big[S(x_2,y_2) - S(x_2,y_1) + \frac{g(y_2)}{g(y_1)}\Big(c(x_2,y_1) - c(x_2,y_2)\Big)\Big] \\ - \pi^p\left(\frac{1}{f(x_1)}\right)\Big[S(x_1,y_2) - S(x_1,y_1) + \frac{g(y_2)}{g(y_1)}\Big(c(x_1,y_1) - c(x_1,y_2)\Big)\Big] \leq 0$$

D.4 Assortative Matching in Equilibrium

Lemmas C.1 and C.2 carry over to the case of private information:

Lemma D.2. *The rate of change in Welfare $\hat{W}\left(\lambda^1(x_1), \lambda^2(x_1)\right)$ monotonically decreases in $\lambda^k(x_1)$ for each $k = 1, 2$.*

Proof. Recall the welfare of children:

$$\hat{W}\big(\lambda^1(x_1),\lambda^2(x_1)\big) = \pi^p(\theta_1)\cdot \underbrace{\Big\{g(y_1)\,\lambda^1(x_1)\,S(x_1,y_1) + \big(1-g(y_1)\big)\,\Big[\lambda^2(x_1)\,S(x_1,y_2) - \lambda^1(x_1)\Big(c(x_1,y_1)-c(x_1,y_2)\Big)\Big]\Big\}}_{\mathbb{E}\hat{U}_1}$$

$$+\,\pi^p(\theta_2)\cdot \underbrace{\Big\{g(y_1)\,\big(1-\lambda^1(x_1)\big)\,S(x_2,y_1) + \big(1-g(y_1)\big)\,\Big[\big(1-\lambda^2(x_1)\big)\,S(x_2,y_2) - \big(1-\lambda^1(x_1)\big)\Big(c(x_2,y_1)-c(x_2,y_2)\Big)\Big]\Big\}}_{\mathbb{E}\hat{U}_2}$$

where

$$\theta_1 = \frac{g(y_1)\lambda^1(x_1) + \big(1-g(y_1)\big)\lambda^2(x_1)}{f(x_1)} \quad \text{and} \quad \theta_2 = \frac{1-\theta_1 f(x_1)}{1-f(x_1)}$$

Fix $\lambda^{-k}(x_1)$. Increasing $\lambda^k(x_1)$ by a small amount $\varepsilon > 0$, increases $\mathbb{E}\hat{U}_1$ and θ_1, and decreases $\mathbb{E}\hat{U}_2$ and θ_2 linearly. Since $\pi^p(\cdot)$ is a decreasing and convex function, the rate of increase through $\pi^p(\theta_1)\cdot\mathbb{E}\hat{U}_1$ decreases, while the rate of decrease through $\pi^p(\theta_2)\cdot\mathbb{E}\hat{U}_2$ increases in $\lambda^k(x_1)$, for any $k = 1, 2$. □

Lemma D.2 implies that $\frac{\partial\hat{W}\left(\lambda^1(x_1),\lambda^2(x_1)\right)}{\partial\lambda^k(x_1)}$ is monotonically decreasing. Now, another useful lemma follows:

Lemma D.3. *Fix $\big(\hat{\lambda}^1(x_1),\hat{\lambda}^2(x_1)\big)$. For any $\big(\tilde{\lambda}^1(x_1),\tilde{\lambda}^2(x_1)\big)$ such that $\theta_1\big(\hat{\lambda}^1(x_1),\hat{\lambda}^2(x_1)\big) = \theta_1\big(\tilde{\lambda}^1(x_1),\tilde{\lambda}^2(x_1)\big)$ and $\hat{\lambda}^1(x_1) \geq \tilde{\lambda}^1(x_1)$, the following holds:*

$$\frac{\partial\hat{W}\big(\lambda^1(x_1),\lambda^2(x_1)\big)}{\partial\lambda^k(x_1)}\Big|_{\left(\hat{\lambda}^1(x_1),\hat{\lambda}^2(x_1)\right)} \leq \frac{\partial\hat{W}\big(\lambda^1(x_1),\lambda^2(x_1)\big)}{\partial\lambda^k(x_1)}\Big|_{\left(\tilde{\lambda}^1(x_1),\tilde{\lambda}^2(x_1)\right)}$$

Proof. Taking partial derivative on welfare under private information yields the followings:

$$\frac{\partial\hat{W}\big(\lambda^1(x_1),\lambda^2(x_1)\big)}{\partial\lambda^1(x_1)} = g(y_1)\cdot V(\lambda^1(x_1),\lambda^2(x_1)) + g(y_1)\cdot\Big[\pi^p(\theta_1)S(x_1,y_1) - \pi^p(\theta_2)S(x_2,y_1)\Big] - \big(1-g(y_1)\big)\cdot\Big\{\pi^p(\theta_1)\big[c(x_1,y_1)-c(x_1,y_2)\big] - \pi^p(\theta_2)\big[c(x_2,y_1)-c(x_2,y_2)\big]\Big\}$$

$$\frac{\partial \hat{W}\left(\lambda^1(x_1), \lambda^2(x_1)\right)}{\partial \lambda^2(x_1)} =$$
$$\left(1-g(y_1)\right) \cdot V(\lambda^1(x_1), \lambda^2(x_1)) + \left(1-g(y_1)\right) \cdot \left[\pi^p(\theta_1) S(x_1, y_2) - \pi^p(\theta_2) S(x_2, y_2)\right]$$

with

$$V(\lambda^1(x_1), \lambda^2(x_1)) = \frac{\pi^{p\prime}(\theta_1)}{f(x_1)} \cdot \mathbb{E}\hat{U}_1(\lambda^1(x_1), \lambda^2(x_1)) - \frac{\pi^{p\prime}(\theta_2)}{1-f(x_1)} \cdot \mathbb{E}\hat{U}_2(\lambda^1(x_1), \lambda^2(x_1))$$

where $\mathbb{E}\hat{U}_1(\lambda^1(x_1), \lambda^2(x_1))$ and $\mathbb{E}\hat{U}_2(\lambda^1(x_1), \lambda^2(x_1))$ are defined as in Lemma D.2. Notice, plugging $V(\lambda^1(x_1), \lambda^2(x_1))$ into $\frac{\partial \hat{W}\left(\lambda^1(x_1), \lambda^2(x_1)\right)}{\partial \lambda^1(x_1)}$ yields the following:

$$\frac{\partial \hat{W}\left(\lambda^1(x_1), \lambda^2(x_1)\right)}{\partial \lambda^1(x_1)} = \frac{g(y_1)}{\left(1-g(y_1)\right)} \cdot \frac{\partial \hat{W}\left(\lambda^1(x_1), \lambda^2(x_1)\right)}{\partial \lambda^2(x_1)} + g(y_1) \cdot Z^{PI}(\theta_1)$$

It is easy to see that $V(\lambda^1(x_1), \lambda^2(x_1))$ is the same as the complete information case, and thus, it decreases as we move down on the market tightness θ_1. In other words, as we increase $\lambda^1(x_1)$ while decreasing $\lambda^2(x_1)$, $V(\lambda^1(x_1), \lambda^2(x_1))$ decreases. This implies that the rate of change with respect to $\lambda^1(x_1)$ decreases as one moves down on the same market tightness, which finishes the proof. □

Now, by using Lemmas D.2 and D.3, we characterize the equilibrium allocation of parents across submarkets as in the complete information case.

Proposition D.2 (**Positive Assortative Matching(PAM)**). *Suppose that* $\frac{S(x_2,y_2)-S(x_2,y_1)+\frac{g(y_2)}{g(y_1)}\cdot[c(x_2,y_1)-c(x_2,y_2)]}{S(x_1,y_2)-S(x_1,y_1)+\frac{g(y_2)}{g(y_1)}\cdot[c(x_1,y_1)-c(x_1,y_2)]} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)}$ *holds. The equilibrium exhibits:*

i. low-type PAM with $\lambda^{1\star}(x_1) \in (0,1)$ *and* $\lambda^{2\star}(x_1) = 0$ *if*

$$\frac{\partial \hat{W}(\lambda^1(x_1), \lambda^2(x_1))}{\partial \lambda^1(x_1)}\Big|_{\{\lambda^2(x_1)=0\}} = 0 \text{ for some } \lambda^{1\star}(x_1) \in (0,1) \tag{D.3}$$

ii. perfect PAM with $\lambda^{1\star}(x_1) = 1$ *and* $\lambda^{2\star}(x_1) = 0$ *if*

$$\frac{\partial \hat{W}(\lambda^1(x_1), \lambda^2(x_1))}{\partial \lambda^1(x_1)}\Big|_{\{\lambda^1(x_1)=1,\lambda^2(x_1)=0\}} \geq 0 \geq \frac{\partial \hat{W}(\lambda^1(x_1), \lambda^2(x_1))}{\partial \lambda^2(x_1)}\Big|_{\{\lambda^1(x_1)=1,\lambda^2(x_1)=0\}} \tag{D.4}$$

iii. high-type PAM with $\lambda^{1\star}(x_1) = 1$ *and* $\lambda^{2\star}(x_1) \in (0,1)$ *if*

$$\frac{\partial \hat{W}(\lambda^1(x_1), \lambda^2(x_1))}{\partial \lambda^2(x_1)}\Big|_{\{\lambda^1(x_1)=1\}} = 0 \text{ for some } \lambda^2(x_1)^* \in (0,1) \tag{D.5}$$

Proof. By assumption, $\frac{S(x_2,y_2)-S(x_2,y_1)+\frac{g(y_2)}{g(y_1)}\cdot[c(x_2,y_1)-c(x_2,y_2)]}{S(x_1,y_2)-S(x_1,y_1)+\frac{g(y_2)}{g(y_1)}\cdot[c(x_1,y_1)-c(x_1,y_2)]} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)}$ holds, which implies that $Z^{PI}(\theta_1) \geq 0$ for any θ_1. Therefore, starting from an initial allocation $\lambda^1(x_1) = 0$ and $\lambda^2(x_1) = 0$, the designer first allocated y_1-parents into submarket x_1 until either parents are exhausted or itis not profitable anymore. Accordingly, perfect PAM and high-type PAM follows. □

Proposition D.2 characterizes the equilibrium sorting patterns when the conditions specified in Corollary 2(i) hold. We leave the analysis of the cases where the sufficient conditions are violated to the readers.

D.5 Proof of Proposition 4

The designer solves the following problem:

$$\max_{\left\{\left(\lambda^k(x_i),\tau^k(x_i)\right)_{i=1}^2\right\}_{k=1}^2} \left\{ \sum_{i=1}^2 \pi^c(\theta_i) \frac{\sum_{k=1}^2 \left[u(x_i,y_k) - \tau^k(x_i)\right]\lambda^k(x_i) g(y_k)}{\sum_{k=1}^2 \lambda^k(x_i)\, g(y_k)} f(x_i) \right\}$$

subject to [FC], [MT],[PC], and [IC]. We will analyze the constraints in this maximization problem.

First, consider the [IC]s for low- and high-ability parents, respectively:

$$\sum_{i=1}^2 c(x_i,y_1)\left[\lambda^2(x_i) - \lambda^1(x_i)\right]\pi^p(\theta_i) \geq \sum_{i=1}^2 \left[\tau^2(x_i)\,\lambda^2(x_i) - \tau^1(x_i)\,\lambda^1(x_i)\right]\pi^p(\theta_i)$$

$$\sum_{i=1}^2 \left[\tau^2(x_i)\,\lambda^2(x_i) - \tau^1(x_i)\,\lambda^1(x_i)\right]\pi^p(\theta_i) \geq \sum_{i=1}^2 c(x_i,y_2)\left[\lambda^2(x_i) - \lambda^1(x_i)\right]\pi^p(\theta_i)$$

From these two inequalities, we get the following expression:

$$\sum_{i=1}^2 c(x_i,y_1)\left[\lambda^2(x_i) - \lambda^1(x_i)\right]\pi^p(\theta_i) \geq \sum_{i=1}^2 c(x_i,y_2)\left[\lambda^2(x_i) - \lambda^1(x_i)\right]\pi^p(\theta_i)$$

$$\Rightarrow c(x_1,y_1)\left[\lambda^2(x_1) - \lambda^1(x_1)\right]\pi^p(\theta_1) + c(x_2,y_1)\left[\lambda^2(x_2) - \lambda^1(x_2)\right]\pi^p(\theta_2) \geq$$
$$c(x_1,y_2)\left[\lambda^2(x_1) - \lambda^1(x_1)\right]\pi^p(\theta_1) + c(x_2,y_2)\left[\lambda^2(x_2) - \lambda^1(x_2)\right]\pi^p(\theta_2)$$

$$\Rightarrow \left[c(x_2,y_1) - c(x_2,y_2)\right]\cdot\left[\lambda^2(x_2) - \lambda^1(x_2)\right]\pi^p(\theta_2) \geq$$
$$\left[c(x_1,y_2) - c(x_1,y_1)\right]\cdot\left[\lambda^2(x_1) - \lambda^1(x_1)\right]\pi^p(\theta_1)$$

Note that $\lambda^2(x_1) - \lambda^1(x_1) = 1 - \lambda^2(x_2) - [1 - \lambda^1(x_2)] = \lambda^1(x_2) - \lambda^2(x_2)$, hence replacing in the previous inequality yields:

$$\begin{aligned}\big[c(x_2, y_1) - c(x_2, y_2)\big] \cdot \big[\lambda^2(x_2) - \lambda^1(x_2)\big]\pi^p(\theta_2) \geq \\ \big[c(x_1, y_1) - c(x_1, y_2)\big] \cdot \big[\lambda^2(x_2) - \lambda^1(x_2)\big]\pi^p(\theta_1) \quad \text{(D.6)}\end{aligned}$$

This inequality depends on the sign of the term $[\lambda^2(x_2) - \lambda^1(x_2)]$, which defines PAM and NAM. Hence, consider the following cases:

- **Case 1:** Suppose $\lambda^2(x_2) - \lambda^1(x_2)$ is positive. Then, Equation D.6 reduces to: $\big[c(x_2, y_1) - c(x_2, y_2)\big] \cdot \pi^p(\theta_2) \geq \big[c(x_1, y_1) - c(x_1, y_2)\big] \cdot \pi^p(\theta_1)$ which is satisfied if the following holds:

$$\frac{c(x_2, y_2) - c(x_2, y_1)}{c(x_1, y_2) - c(x_1, y_1)} \geq \frac{1}{\pi^p\left(1/f(x_2)\right)} \tag{D.7}$$

- **Case 2:** Suppose $\lambda^2(x_2) - \lambda^1(x_2)$ is negative. Then, Equation D.6 reduces to: $\big[c(x_1, y_1) - c(x_1, y_2)\big] \cdot \pi^p(\theta_1) \geq \big[c(x_2, y_1) - c(x_2, y_2)\big] \cdot \pi^p(\theta_2)$ which is satisfied if the following holds:

$$\frac{c(x_1, y_2) - c(x_1, y_1)}{c(x_2, y_2) - c(x_2, y_1)} \geq \frac{1}{\pi^p\left(1/f(x_1)\right)} \tag{D.8}$$

Now, we show that the [PC] for low-ability parents, and the [IC] for high-ability parents imply the [PC] for high-ability parents:

$$\begin{aligned}\sum_{i=1}^{2}\big[\tau^2(x_i) - c(x_i, y_2)\big]\lambda^2(x_i)\pi^p(\theta_i) &\geq \sum_{i=1}^{2}\big[\tau^1(x_i) - c(x_i, y_2)\big]\lambda^1(x_i)\pi^p(\theta_i) \\ &\geq \sum_{i=1}^{2}\big[\tau^1(x_i) - c(x_i, y_1)\big]\lambda^1(x_i)\pi^p(\theta_i) \geq 0\end{aligned}$$

Thus, we can ignore the [PC] for high-ability parents.

Next, suppose that the [IC] for high-ability parents holds with strict inequality:

$$\sum_{i=1}^{2}\big[\tau^2(x_i) - c(x_i, y_2)\big]\lambda^2(x_i)\pi^p(\theta_i) > \sum_{i=1}^{2}\big[\tau^1(x_i) - c(x_i, y_2)\big]\lambda^1(x_i)\pi^p(\theta_i)$$

Then, the designer can decrease $\tau^2(x_1)$ and $\tau^2(x_2)$ by a small $\varepsilon > 0$ satisfying the constraint while increasing the objective function. A contradiction. Therefore, the [IC] for high-ability parents holds with equality at the optimum.

Similarly, suppose that the [PC] for low-ability parents holds with strict inequality:

$$\sum_{i=1}^{2}\left[\tau^1(x_i)-c(x_i,y_1)\right]\lambda^1(x_i)\pi^p(\theta_i)>0$$

Then, the designer can decrease $\tau^1(x_1)$ and $\tau^1(x_2)$ by a small $\varepsilon>0$ satisfying the constraint while increasing the objective function. A contradiction. Therefore, the [PC] for low-ability parents holds with equality at the optimum.

Lastly, we show that the [IC] for high-ability parents combined with Equations D.7 and D.8 imply the [IC] for low-ability parents. Thus, consider the [IC] for high-ability parents:

$$\sum_{i=1}^{2}\left[\tau^2(x_i)-c(x_i,y_2)\right]\lambda^2(x_i)\pi^p(\theta_i)=\sum_{i=1}^{2}\left[\tau^1(x_i)-c(x_i,y_2)\right]\lambda^1(x_i)\pi^p(\theta_i)$$
$$\Rightarrow\sum_{i=1}^{2}\left[\tau^2(x_i)\lambda^2(x_i)-\tau^1(x_i)\lambda^1(x_i)\right]\pi^p(\theta_i)=\sum_{i=1}^{2}c(x_i,y_2)\left[\lambda^2(x_i)-\lambda^1(x_i)\right]\pi^p(\theta_i)$$

The right-hand side of the previous equation can be written as:

$$c(x_1,y_2)\left[\lambda^2(x_1)-\lambda^1(x_1)\right]\pi^p(\theta_1)+c(x_2,y_2)\left[\lambda^2(x_2)-\lambda^1(x_2)\right]\pi^p(\theta_2)$$
$$\Rightarrow c(x_2,y_2)\left[\lambda^2(x_2)-\lambda^1(x_2)\right]\pi^p(\theta_2)-c(x_1,y_2)\left[\lambda^2(x_2)-\lambda^1(x_2)\right]\pi^p(\theta_1)$$
$$\Rightarrow\left[c(x_2,y_2)\pi^p(\theta_2)-c(x_1,y_2)\pi^p(\theta_1)\right]\cdot\left[\lambda^2(x_2)-\lambda^1(x_2)\right]$$

Thus, the [IC] for high-ability parents can be written as:

$$\sum_{i=1}^{2}\left[\tau^2(x_i)\lambda^2(x_i)-\tau^1(x_i)\lambda^1(x_i)\right]\pi^p(\theta_i)=$$
$$\left[c(x_2,y_2)\pi^p(\theta_2)-c(x_1,y_2)\pi^p(\theta_1)\right]\cdot\left[\lambda^2(x_2)-\lambda^1(x_2)\right]$$

As previously, we need to consider the following cases:

- **Case 1:** Suppose $\lambda^2(x_2)-\lambda^1(x_2)$ is positive. Equation D.7 ensures that the following inequality holds:

$$\left[c(x_2,y_1)-c(x_2,y_2)\right]\pi^p(\theta_2)\cdot\left[\lambda^2(x_2)-\lambda^1(x_2)\right]\geq$$
$$\left[c(x_1,y_1)-c(x_1,y_2)\right]\pi^p(\theta_1)\cdot\left[\lambda^2(x_2)-\lambda^1(x_2)\right]$$

After some algebra:

$$\Big[c(x_2, y_1)\pi^p(\theta_2) - c(x_1, y_1)\pi^p(\theta_1)\Big] \cdot \big[\lambda^2(x_2) - \lambda^1(x_2)\big] \geq \Big[c(x_2, y_2)\pi^p(\theta_2) - c(x_1, y_2)\pi^p(\theta_1)\Big] \cdot \big[\lambda^2(x_2) - \lambda^1(x_2)\big]$$

Which implies the [IC] for low-ability parents:

$$\Big[c(x_2, y_1)\pi^p(\theta_2) - c(x_1, y_1)\pi^p(\theta_1)\Big] \cdot \big[\lambda^2(x_2) - \lambda^1(x_2)\big] \geq \sum_{i=1}^{2} \big[\tau^2(x_i)\lambda^2(x_i) - \tau^1(x_i)\lambda^1(x_i)\big]\pi^p(\theta_i)$$

- **Case 2:** Suppose $\lambda^2(x_2) - \lambda^1(x_2)$ is negative. Equation D.8 ensures that the following inequality holds:

$$\Big[c(x_1, y_1) - c(x_1, y_2)\Big]\pi^p(\theta_1) \cdot \big[\lambda^1(x_2) - \lambda^2(x_2)\big] \geq \Big[c(x_2, y_1) - c(x_2, y_2)\Big]\pi^p(\theta_2) \cdot \big[\lambda^1(x_2) - \lambda^2(x_2)\big]$$

After some algebra:

$$\Big[c(x_2, y_1)\pi^p(\theta_2) - c(x_1, y_1)\pi^p(\theta_1)\Big] \cdot \big[\lambda^1(x_2) - \lambda^2(x_2)\big] \geq \Big[c(x_2, y_2)\pi^p(\theta_2) - c(x_1, y_2)\pi^p(\theta_1)\Big] \cdot \big[\lambda^1(x_2) - \lambda^2(x_2)\big]$$

Which implies the [IC] for low-ability parents:

$$\Big[c(x_2, y_1)\pi^p(\theta_2) - c(x_1, y_1)\pi^p(\theta_1)\Big] \cdot \big[\lambda^2(x_2) - \lambda^1(x_2)\big] \geq \sum_{i=1}^{2} \big[\tau^2(x_i)\lambda^2(x_i) - \tau^1(x_i)\lambda^1(x_i)\big]\pi^p(\theta_i)$$

Therefore, we can drop the [IC] for low-ability parents.

E Appendix: Analysis of Extension

E.1 Continuous Type of Parents

First, let's rewrite the constraints the designer faces when solving Equation 6:

$$[\text{FC}] \quad \tau(x,y) \geq 0 \text{ and } \lambda(x,y) \geq 0 \text{ for all } (x,y), \text{ and} \sum_{i=1}^{2} \lambda(x_i,y) = 1 \text{ for all } y \in Y.$$

$$[\text{MT}] \quad \theta_i = \frac{1}{f(x)} \cdot \int_{\underline{y}}^{\overline{y}} \lambda(x,y) g(y)\, dy\,, \text{ for all } i \in \{1,2\}$$

$$[\text{PC}] \quad \sum_{i=1}^{2} \Big[\tau(x_i,y) - c(x_i,y)\Big] \lambda(x_i,y) \pi^p(\theta_i) \geq 0\,, \text{ for all } y \in Y.$$

E.1.1 Proof of Lemma 2

Let $\lambda(x,y)$ be an arbitrary interior allocation for any y, that is, $\lambda(x_1,y) \in (0,1)$, and thus $\lambda(x_2,y) = 1 - \lambda(x_1,y) \in (0,1)$ for any y by [FC]. Define a perturbation function $\varepsilon : Y \to (0,1)$ such that $\int\limits_{y \in Y} \varepsilon(y) g(y) dy = 0$.

Consider the allocations $\lambda(y) \equiv \big(\lambda(x_1,y), \lambda(x_2,y)\big)$ and $\tilde{\lambda}(y) \equiv \big(\lambda(x_1,y) - \varepsilon(y), \lambda(x_2,y) + \varepsilon(y)\big)$. Notice, the market tightness derived by allocations $\lambda(y)$ and $\tilde{\lambda}(y)$ is the same: $\theta_i = \frac{1}{f(x_i)} \cdot \int\limits_{y \in Y} \lambda(x_i,y) g(y) dy$ for $i = 1,2$. The change in welfare between these two allocations is:

$$W(\tilde{\lambda}) - W(\lambda) \equiv \Delta_W = \int_{\underline{y}}^{\overline{y}} \underbrace{\big[\pi^p(\theta_2) S(x_2,y) - \pi^p(\theta_1) S(x_1,y)\big]}_{\hat{Z}^{CI}(y|\theta)} \varepsilon(y) g(y) dy,$$

where $\theta = (\theta_1, \theta_2)$. Note that $\hat{Z}^{CI}(y|\theta)$ is continuous but not necessarily monotone, provided that $S(x,y)$ is continuous in y for all x.

Now, let Figure E.1 be an arbitrary representation of $\hat{Z}^{CI}(y|\theta)$, and consider $\varepsilon(y)$ defined as follows:

$$\int\limits_{y \in [\underline{y}, y_1]} \varepsilon(y) g(y) dy = 0, \quad \int\limits_{y \in [y_1, y_2]} \varepsilon(y) g(y) dy = 0, \text{ and} \int\limits_{y \in [y_2, \overline{y}]} \varepsilon(y) g(y) dy = 0$$

and more importantly, $\partial\varepsilon(y)/\partial y > 0$ for $y \in Y \backslash [y_1, y_2]$ and $\partial\varepsilon(y)/\partial y < 0$ for $y \in [y_1, y_2]$.[38]

[38] Please refer to y_1 and y_2 defined in Figure E.1, here and henceforth.

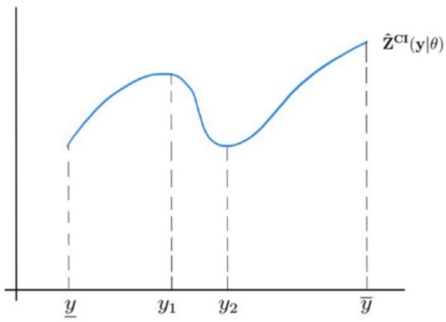


Figure E.1: Change in Welfare - Extension

Thus, the change in welfare is:

$$\int_{y\in[\underline{y},y_1]} \hat{Z}^{CI}(y|\theta)\varepsilon(y)g(y)dy + \int_{y\in[y_1,y_2]} \hat{Z}^{CI}(y|\theta)\varepsilon(y)g(y)dy + \int_{y\in[y_2,\overline{y}]} \hat{Z}^{CI}(y|\theta)\varepsilon(y)g(y)dy.$$

Since $\hat{Z}^{CI}(y|\theta)$ is monotonically increasing over the interval $[\underline{y}, y_1]$ choosing $\varepsilon(y)$ to be monotonic increasing ensures that the first term above is positive. Similarly, each term can be shown to be positive, which collectively guarantees a welfare improvement over the interior allocation $\lambda(x, y)$.[39] The analysis holds for any interior allocation $\lambda(x, y) \in (0, 1)$ over any arbitrary subset $Y' \subseteq Y$.

Lemma 2 implies the following: "Given an equilibrium market tightness (θ_1^*, θ_2^*), the optimal allocation is always on the corner, that is, $\lambda(x_1, y) \in \{0, 1\}$". Specifically, the following yields the optimal allocation:

Corollary E.1. *Let $\theta^* \equiv (\theta_1^*, \theta_2^*)$ be the equilibrium market tightness. Suppose $\hat{Z}^{CI}(y|\theta^*)$ is as in Figure E.1. Then the optimal allocation $\lambda^*(x, y)$ is as follows: $\lambda^*(x_1, y) = 1 - \lambda^*(x_2, y)$ and:*

$$\lambda^*(x_2, y) = \begin{cases} 0 & y \in [\underline{y}, z_1] \cup [z_2, z_3] \\ 1 & y \in [z_1, z_2] \cup [z_3, \overline{y}] \end{cases}$$

for some z_1, z_2, z_3 such that $\underline{y} < z_1 < y_1 < z_2 < y_2 < z_3 < \overline{y}$ as in Figure E.2.

[39] Notice, one can always find a such an $\varepsilon(y)$ through a very small perturbation around interior $\lambda(x, y)$.

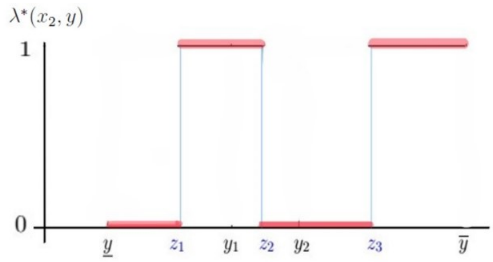


Figure E.2: Optimal Allocation - Extension

Proof. By Lemma 2, a monotone increasing perturbation of $\lambda(y)$ over the interval $[\underline{y}, y_1]$ such that $\int\limits_{y\in[\underline{y},y_1]} \varepsilon(y)g(y)dy = 0$ guarantees that $\int\limits_{y\in[\underline{y},y_1]} \hat{Z}^{CI}(y|\theta)\varepsilon(y)g(y)dy > 0$. Since $\varepsilon(y)$ monotone increases and sums up to 0, there exists $z \in (\underline{y}, y_1)$ such that $\varepsilon(y) \leq 0$ if and only if $y \leq z$. This implies $\lambda(x_1, y) \leq \tilde{\lambda}(x_1, y)$ if and only if $y \leq z$. Thus, moving towards PAM only in the interval $[\underline{y}, y_1]$ increases the welfare. Therefore, one can keep increasing the welfare only over the region $[\underline{y}, y_1]$ by trembling as much as possible, which proves that there exists $z_1 \in (\underline{y}, y_1)$ such that $\lambda^*(x_1, y) = 1$ for $y \in [\underline{y}, z_1]$, and $\lambda^*(x_1, y) = 0$ for $y \in [z_1, y_1]$. Analogously, the optimal allocation for other regions follows. □

E.1.2 Proof of Proposition 5

Notice that, if $\pi^p(\theta^*)S(x, y)$ is super-modular given equilibrium $\theta^* = (\theta_1^*, \theta_2^*)$, then $\hat{Z}^{CI}(y|\theta^*)$ is increasing everywhere. Thus, Corollary E.1 implies that the optimal allocation is such that $\lambda(x_1, y) = 1$ for $y \leq \hat{y}^{PAM}$ for some $\hat{y}^{PAM} \in (\underline{y}, \overline{y})$, and $\lambda(x_1, y) = 0$ otherwise. Moreover, given θ_1^* is the equilibrium market tightness in submarket x_1, $\hat{y}^{PAM}$ is such that $\theta_1^* = G(\hat{y}^{PAM})/1-f(x_2)$. Proof of part (ii) analogously follows.

E.1.3 Proof of Corollary 4

Recall $\hat{Z}^{CI}(y|\theta) = \pi^p(\theta_2)S(x_2, y) - \pi^p(\theta_1)S(x_1, y)$. One can easily see that $\hat{Z}^{CI}(y|\theta)$ increases in θ_1. Taking derivative of $\hat{Z}^{CI}(y|\theta)$ with respect to y yields the following:

$$\frac{\partial \hat{Z}^{CI}(y|\theta)}{\partial y} \equiv \hat{Z}_y^{CI}(y|\theta) = \pi^p(\theta_2)S_y(x_2, y) - \pi^p(\theta_1)S_y(x_1, y).$$

Since $S_y(y|\theta) > 0$, it follows that $\hat{Z}_y^{CI}(y|\theta)$ also increases in θ_1, and thus assigns its minimum value at $\theta_1 = 0$ and $\theta_2 = 1/f(x_2)$. That is, $\hat{Z}_y^{CI}(y|\theta_1 = 0, \theta_2 = 1/f(x_2)) \leq \hat{Z}_y^{CI}(y|\theta)$ for any θ and any y.

Let $\grave{y} := \arg\min_{y \in Y} S_y(x_1, y) - \pi^p\left(1/1-f(x_1)\right) S_y(x_2, y)$, that is, $\grave{y}$ is the argument at which $\hat{Z}_y^{CI}(y|\theta_1 = 0, \theta_2 = 1/f(x_2))$ assigns its minimum value. Now, notice the following: $\pi^p\left(1/f(x_2)\right) S_y(x_2, \grave{y}) - S_y(x_1, \grave{y}) \geq 0$ implies that $\hat{Z}_y^{CI}(y|\theta) \geq 0$ for any θ and any $y \in [\underline{y}, \overline{y}]$. As a result, $\pi^p(\theta^*) S(x, y)$ is super-modular at equilibrium θ^*, and the optimal sorting exhibits PAM by Proposition 5(i). Therefore, the planner simply optimizes the welfare by solving the following problem:

$$\max_{\hat{y} \in Y} \pi^p\left(\frac{G(\hat{y})}{f(x_1)}\right) \int_{\underline{y}}^{\hat{y}} S(x_1, y) g(y) dy + \pi^p\left(\frac{1 - G(\hat{y})}{1 - f(x_1)}\right) \int_{\hat{y}}^{\overline{y}} S(x_2, y) g(y) dy,$$

which finishes the proof of Corollary 4 (i). The proof of (ii) follows analogously.

E.2 Improvement in the Meeting Technology

Recall the sorting condition under complete information for a given meeting technology π^p:

$$Z^{CI}(\theta_1|\pi^p) = \pi^p(\theta_2) \underbrace{[S(x_2, y_2) - S(x_2, y_1)]}_{\Delta S_2} - \pi^p(\theta_1) \underbrace{[S(x_1, y_2) - S(x_1, y_1)]}_{\Delta S_1}.$$

If the equilibrium sorting is PAM, then the equilibrium θ_1^* is such that $Z^{CI}(\theta_1^*|\pi^p) > Z^{CI}(\overline{\theta}_1|\pi^p) = 0$ and $\theta_1^* > \overline{\theta}_1$. Note that $\left(\pi^p(\overline{\theta}_1)/\pi^p(\overline{\theta}_2)\right) = \left(\Delta S_2/\Delta S_1\right)$.

E.2.1 Proof of Proposition 6

Take an arbitrary $\hat{\pi}^p \in \mathbf{\Pi}^p$ such that $\pi^p \rhd \hat{\pi}^p$. Notice $Z^{CI}(\overline{\overline{\theta}}_1|\hat{\pi}^p) = 0$ if and only if $\left(\hat{\pi}^p(\overline{\overline{\theta}}_1)/\hat{\pi}^p(\overline{\overline{\theta}}_2)\right) = \left(\Delta S_2/\Delta S_1\right) = \left(\pi^p(\overline{\theta}_1)/\pi^p(\overline{\theta}_2)\right)$. It is easy to verify that:

$$\frac{\hat{\pi}^p(\theta_1)}{\hat{\pi}^p(\theta_2)} \leq \frac{\pi^p(\theta_1)}{\pi^p(\theta_2)} \text{ if and only if } \theta_1 \leq 1,$$

which holds with equality if $\theta_1 = \theta_2 = 1$. In short, given any market tightness θ_1 and $\theta_2 \equiv \left(1 - f(x_1)\theta_1\right)/\left(1 - f(x_1)\right)$, the ratio of meeting probabilities in submarkets x_1 and x_2, gets flatter as the technology improves (as can be seen in Figure E.3). Thus, a super-modular $S(x, y)$ implies $\overline{\overline{\theta}}_1 < \overline{\theta}_1 < \theta_1^*$.

Recall that the meeting technology gets flatter *everywhere* at an improved technology, that is, let $\partial\pi^p(\theta)/\partial\theta \leq \partial\hat{\pi}^p(\theta)/\partial\theta$ for any $\theta \in [0, \min\{1/f(x_1), 1/1-f(x_1)\}]$ given $f(x_1)$. Notice, Lemma C.1 also implies that $\partial W\left(\lambda^1(x_1), \lambda^2(x_1)\right)/\partial\lambda^k(x_i)$ is monotonically decreasing for any $i = 1, 2$ and any $k = 1, 2$. Given π^p, let $\theta_i^* = \min\{\theta_1^*, \theta_2^*\}$ and

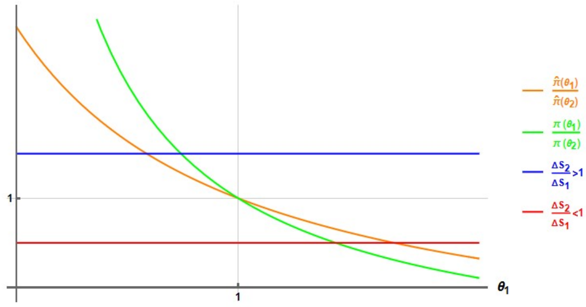


Figure E.3: Monotone Comparative Statics for the Meeting Technology

thus $\theta_i^* \leq 1 \leq \theta_j^*$. Therefore,

$$\frac{\partial W\left(\lambda^1(x_1), \lambda^2(x_1)\right)}{\partial \lambda^k(x_i)}|_{\theta_i=\theta_i^*} = 0 \geq \frac{\partial W\left(\lambda^1(x_1), \lambda^2(x_1)\right)}{\partial \lambda^k(x_i)}|_{\theta_i=1}$$

for some $k = 1, 2$.[40] The inequality becomes strict unless $\min\{\theta_1^*, \theta_2^*\} = \max\{\theta_1^*, \theta_2^*\}$. Suppose that is the case from now on.

Thus, at equal market tightness where the parents-to-children ratio is equal to 1 in both market, the designer would like to allocate some of type-k parents into submarket x_j because $\partial W(\lambda)/\partial\lambda^k(x_i)|_{\theta_i=1} < 0$. Doing so will have two effects as in the proof of Lemma C.1: congestion and decongestion effects, which link to the probability of meeting given the technology, and the surplus effect. Notice the surplus effect is linear, whereas the meeting technology is convex. Hence, decreasing $\lambda^k(x_i)$ at $\theta_i = 1$ increases the probability of meeting in submarket-x_i and decreases in submarket-x_j at different and non-constant rates.

Now, for an improved technology $\hat{\pi}^p$ defined above, the congestion and decongestion effects become less pronounced, leading to less divergence from the equal market tightness $\theta_i = 1 = \theta_j$. Let θ_i^{**} be the equilibrium market tightness with $\hat{\pi}^p$. Therefore, $|1 - \theta_i^{**}| < |1 - \theta_i^*|$, which simply implies $\overline{\overline{\theta_1}} < \theta_1^{**}$. The proof follows analogously for a submodular $S(x, y)$ and NAM.

[40]For a supermodular $S(x, y)$, consider the corner $(\lambda^1(x_1), \lambda^2(x_1))$ that exhibits PAM and $\theta_1 = 1 = \theta_2$; that is, either $\lambda^1(x_1) = 1$ and $\lambda^2(x_1) \in [0, 1)$, or $\lambda^1(x_1) \in (0, 1)$ and $\lambda^2(x_1) = 0$. Consider analogous λ's for submodular $S(x_i, y_j)$.