Designing the Menu of Licenses for Foster Care *

Altınok, Ahmet † MacDonald, Diana‡

January 2024

Abstract

In the United States, prospective foster parents must become licensed by a child welfare agency before a foster child can be placed in their care. This paper contributes by developing a theoretical matching model to study the optimal menu of licenses designed to screen foster parents. We construct a two-sided matching model with heterogeneous agents, adverse selection, search frictions, and a designer who coordinates match formation through a menu of contracts. We focus on incentive compatible contracts, examine optimal allocations and transfers, and analyze equilibrium sorting patterns. There are three main results: **(i)** optimal allocation calls for a segregation of the market, **(ii)** a simple transfer schedule does the job, **(iii)** complementarities do not ensure that Positive Assortative Matching (PAM) will arise in equilibrium, thus we provide an additional condition that guarantees it. Our results suggest that the menu of licenses used in practice exhibits some of the properties of the optimal solution. However, the menu might not be reaching its screening objective.

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

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

*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, EPCS Annual Meeting, ITAM, LACEA Annual Meeting, Nazarbayev University, Stony Brook International Conference on Game Theory, and Universidad Popular Autónoma del Estado de Puebla. 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 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 the parents' attribute, search frictions, and a designer who coordinates match formation through a menu of contracts. 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 allocation calls for a segregation of the market. Second, we show that a simple transfer schedule does the job, that is, parents holding different licenses and providing care for the same type of child can receive the same monetary transfer. However,

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

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. 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. It is important to highlight that our theoretical predictions over the optimal sorting pattern conveys information that cannot be obtained empirically due to data restrictions.

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). 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.[2] 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 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 assume 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 market-tightness. Thus, we introduce a search friction assumption into the model.

It is important to highlight that the search friction assumption is a key aspect of our model, since it introduces non-trivial effects on the analysis. In particular, when a mass of type-y parents is re-allocated from one submarket to the other, there are three effects taking place: **(i)** *surplus effect* is the change in total expected

[2] 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.

surplus of the market, **(ii)** *congestion effect* is the increase in the market tightness of the *submarket to* where the parents are reallocated (submarket becomes thicker), and finally **(iii)** *decongestion effect* is the decrease in the market tightness of the *submarket from* where the parents are reallocated (submarket becomes thinner). These effects not only introduce challenges in the analysis, but also make our model predictions richer.

We start by analyzing the complete information case and establish results for a super- and a sub-modular surplus function.[3] First, we find that, if the surplus of a match is super-modular (i.e. there is complementarities between level of care needed and ability to provide care) then it is never optimal for the designer to allocate both type-y parents with strictly positive probability into submarkets x_1 and x_2.[4] 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.[5] 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. But, 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.

[3]We discuss the case of super-modularity and leave sub-modularity to be discussed in the body of the paper.

[4]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.

[5]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).

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 parent 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 foster parents' observable characteristics are not associated with their willingness to foster children with higher-needs, but the license they hold does (Cox et al., 2011).[6] This motivates our next analysis relaxing the assumption over the observability of a parent' attribute.

With the private information, our results from complete information carry on, except the additional condition for PAM. 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. 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.[7] 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.

[6]Using a sample of 297 foster mothers and a linear multiple regression analyses, Cox et al. (2011) found that foster mothers' observable characteristics (such as race, marital status, highest level of education completed, and income) are not associated with the willingness to foster children with emotional and behavioral problems. Moreover, they find that foster mothers who hold a therapeutic license were significantly more willing, than foster mothers holding a traditional license, to foster children with emotional and behavioral problem.

[7]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. There are a few papers analyzing foster care as a matching market. Slaugh et al. (2015), using a reduced form approach, studies the Pennsylvania Adoption Exchange program, a computational tool created to facilitate the adoption of children in foster care, and makes several recommendations to improve the success of adoptions. Robinson-Cortés (2019) presents an empirical framework to study how children are assigned to foster homes using a confidential dataset, and uses the estimates to study different policy interventions. Olberg et al. (2021) constructs a dynamic search and matching model, where agents' attributes are perfectly observable, to compare two different search processes used by child welfare agencies to identify potential adoption matches between parents and children. Lastly, MacDonald (2022) conducts an empirical analysis that yields four new facts related to match transitions of children, and develops a dynamic search and matching model where parents and children can form reversible (foster) or irreversible (adoption) matches to rationalize these empirical facts. Our paper is the first one to incorporate the menu of licenses into the analysis, and to tailor a model to fit the main features of foster care under the presence of information frictions.

This paper is related to the literature on assortative matching under asymmetric information. In a principal-agent setting with adverse selection, several papers have studied sorting patterns arising from microfinance loan contracts where a population of heterogeneous borrowers optimally matches into pairs (Ghatak, 1999; Van Tassel, 1999; Ghatak, 2000; Guttman, 2008; Altinok, 2023). As in our framework, the lender can induce PAM or NAM, but in these papers there is only one side of the market, and no match coordinator. In a principal-agent setting with moral hazard, less related to our model but relevant, Serfes (2005) analyzes equilibrium sorting patterns between heterogeneous principals and agents restricting attention to linear contracts and a CARA utility function. His results rationalize the empirical finding of Ackerberg and Botticini (2002) who document a positive relationship between the degree of risk aversion of tenants and landlords with the riskiness of a crop. Franco et al. (2011) and Kaya and Vereshchagina (2014) examine a framework where a manager assigns heterogeneous workers to teams in the presence of moral hazard, and show that even in the presence of complementarities, the equilibrium sorting might exhibit NAM.

Lastly, our paper relates to the search and matching literature. The randomization device that allocates parents into a particular submarket has a flavor of directed search as in Menzio and Shi (2010a) and Menzio and Shi (2010b). In their labor market framework, they define submarkets, directed search into these submarkets, and use the notion of market tightness. Shi (2001) was the first to highlight that super-modularity in the match value is not sufficient to ensure PAM when considering a specific directed search technology. Later on, Eeckhout and Kircher (2010) provide a stronger complementarity condition necessary and sufficient for PAM using a general directed search technology. Our model differs from the aforementioned papers because we introduce private information and transfers from the designer to one-side of the market. In addition, Shimer and Smith (2000) and Smith (2006) analyze a two-sided matching setting with random search and complete information, show that PAM fails and provide stronger complementarity conditions to ensure it. Thus, by incorporating search frictions, greater complementarity in attributes is needed, such as log-supermodularity and even more, depending on the specific assumption over the search friction. Therefore, our results are in line with the literature.

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 extends the analysis to the case of private information. Lastly, Section 5 concludes. Appendix A presents a description of foster care in the US. All omitted proofs and examples are in the Appendices B and C.

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 $f(x_2) = 1 - f(x_1)$. For the purpose of exposition, 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 $y_2 > y_1$. 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, and one-to-one. 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. Throughout the paper, we restrict attention to the menu of licenses with the following features: (i) allocations are non-wasteful, that is, $\sum_{x \in X} \lambda(x) = 1$,[8] and (ii) parents have limited liability, that is, $\tau(x) \geq 0$ for any $x \in X$.

Figure 1 presents two examples of licenses. In Panel 1a, parents holding license $\mathcal{L}$ are allocated to submarket x_1 with probability one, and to submarket x_2 with probability zero. In Panel 1b, parents holding license $\mathcal{L}'$ are allocated to submarket x_1 with probability $\frac{1}{4}$, and to submarket x_2 with probability $\frac{3}{4}$.

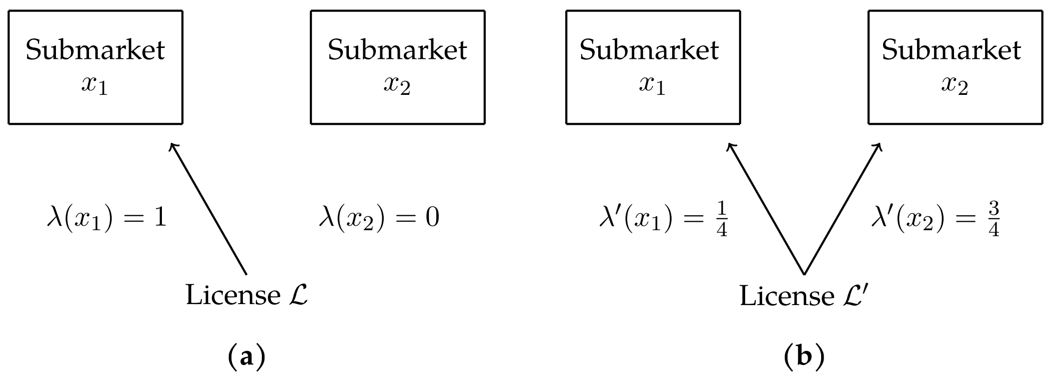


Figure 1: Examples of Licenses

All agents are risk-neutral. The designer maximizes children 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 $c(x, y)$.

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 in institutional care, parents always incur a cost when provid-

[8]For the complete information case, this assumption 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 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 shortage of foster parents, parents who have 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's objectives. In addition, relaxing this assumption would make the problem intractable for the private information case.

ing 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 $\{\mathcal{L}^k\}_{k=1}^2 \equiv \left\{\left\{\left(\lambda^k(x_i), \tau^k(x_i)\right)\right\}_{i=1}^2\right\}_{k=1}^2$.

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.

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 $\pi^p(\theta_x) = \frac{\pi^c(\theta_x)}{\theta_x}$ and $\pi^p(0) = 1$.

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 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:

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

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

$$W(x)=\frac{\sum_{k=1}^{2}\Big[\sum_{j=1}^{2}\big(\big[u(x,y_j)-\tau^k(x)\big]\lambda^k(x)\cdot h^k(y_j)\Big]}{\sum_{k=1}^{2}\lambda^k(x)\cdot\sum_{j=1}^{2}h^k(y_j)}.$$

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:

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

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

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

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

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 using the randomization device of each license, $\{\lambda^1(x_i), \lambda^2(x_i)\}_{i=1}^2$, and establish sorting patterns based on the randomization device.

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**).*

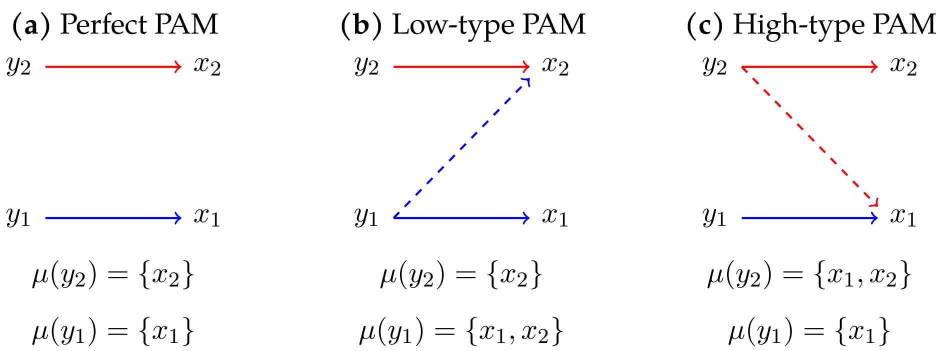


Figure 2: Examples of Positive Assortative Matching (PAM)

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 follows: We say j**–type sorting** if type-y_j parents are allocated into both submarkets x_1 and x_2, while type-y_{-j} parents are allocated only into submarket $x_i \in \{x_1, x_2\}$.[9] To avoid confusion for the rest of the paper, we call it *low*–type if $j = 1$ and *high*–type if $j = 2$. Figure 2 illustrates examples using our notion 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 submarket 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 submarket with strictly positive probability and y_1-parents are allocated into submarket x_2 with probability zero, thus $\lambda^2(x_2) \in (0, 1) \geq \lambda^1(x_2) = 0$. Note that, the randomization device in Panel 2c

[9] Here, $-j$ denotes parents of type that is not j. Formally, we say j–type sorting if $\lambda^j(x) \in (0, 1)$ while $\lambda^{-j}(x) \in \{0, 1\}$.

could represent the menu of licenses used in practice as described in the introduction, where low-needs children can be fostered by two types of parents, while high-needs children are fostered only by one type of parent.

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. First, note that by incorporating the [PC] constraints into the objective function in Equation 1, reduces the designer's problem to:

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

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 $S(x,y) = u(x,y) - c(x,y)$ denote the surplus of a match (x,y) where $S(x,y)$ is increasing in y following Assumption 1, and by Assumption 2, the surplus is greater for matches involving children with low needs:

Assumption 2. *$S(x,y)$ is decreasing in x.*[10]

Lemma 1 states that if the surplus of a match is super- or sub-modular then it is never optimal for the designer to allocate both type of y-parents with strictly positive probability into submarkets x_1 and x_2.

Lemma 1. *For at least one of the licenses, the optimal randomization rule (allocation) yields a corner solution whenever $S(x,y)$ is super- or sub-modular.*

Proof. See Appendix B.1. □

To prove Lemma 1, we start by assuming that the designer allocates both type of y-parents into both submarkets with strictly positive probabilities. Note 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 $\big(\lambda^1(x_1), \lambda^2(x_1)\big)$. Thus, meeting probabilities along that line are constant. Hence, super- or sub-modularity of the surplus function ensures a greater (expected) welfare on the corners than it does in the

[10] This assumption allows the following over the children's payoff function: (i) $u(x,y)$ constant in x, (ii) $u(x,y)$ increasing in x, and (iii) $u(x,y)$ decreasing in x but satisfying that $u(x_1,y) - u(x_2,y) > c(x_1,y) - c(x_2,y)$ for all y .

interior. 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:

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

where:

$$\theta_1 = \frac{g(y_1)\, \lambda^1(x_1) + \big(1 - g(y_1)\big)\, \lambda^2(x_1)}{f(x_1)} \quad \text{and} \quad \theta_2 = \frac{g(y_1)\, \big(1 - \lambda^1(x_1)\big) + \big(1 - g(y_1)\big)\big(1 - \lambda^2(x_1)\big)}{1 - f(x_1)} \tag{3}$$

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 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 3. Note that, $\theta_2 = \frac{1-f(x_1)\theta_1}{1-f(x_1)}$, thus $Z^{CI}(\theta_1)$ 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 of 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 B.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, optimality requires 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.

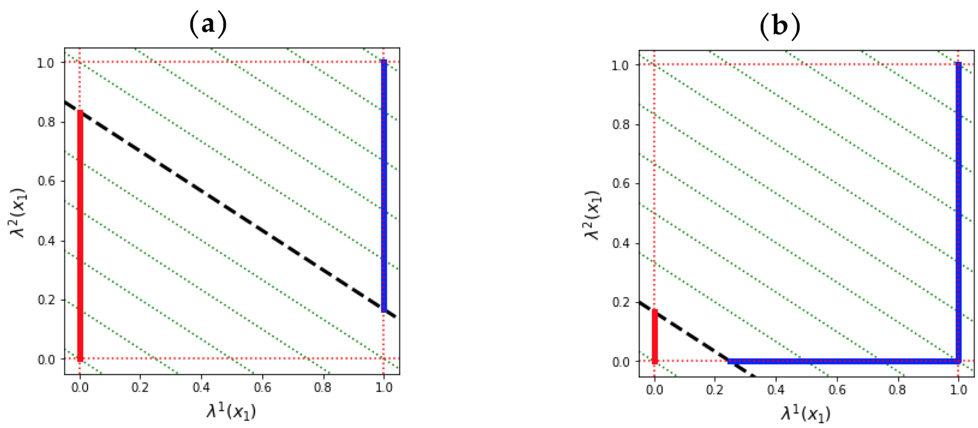


Figure 3: Illustration of PAM and NAM given $Z^{CI}(\bar{\theta}_1)$

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 $(\lambda^1(x_1), \lambda^2(x_1))$ 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-dotted line corresponds to the values of $(\lambda^1(x_1), \lambda^2(x_1))$ such that $Z^{CI}(\bar{\theta}_1) = 0$, each blue-line shows the feasible allocations that can be an equilibrium when $Z^{CI}(\theta_1) > 0$ (above the black-dotted line), and each red-line shows the feasible allocations that can be an equilibrium when $Z^{CI}(\theta_1) < 0$ (below the black-dotted 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.[11]

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 B.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 dominates the adversary effect of the search friction, and becomes sufficient to induce PAM at the optimum. From another point of view, 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 bellow. This is in line with the literature in dynamic search and matching, which imposes stronger complementarity conditions to ensure that more desirable partner have incentives to wait for more desirable partner from the other side of the market.[12] 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 $Z^{CI}(\overline{\theta}) = 0$ is located in the left-bottom

[11]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 B.4 for a detail characterization.

[12]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 give up to 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.

corner. Analogously, in Panel 4b, the equilibrium sorting can only exhibit NAM, since $Z^{CI}(\overline{\theta}) = 0$ is located in the right-top corner.

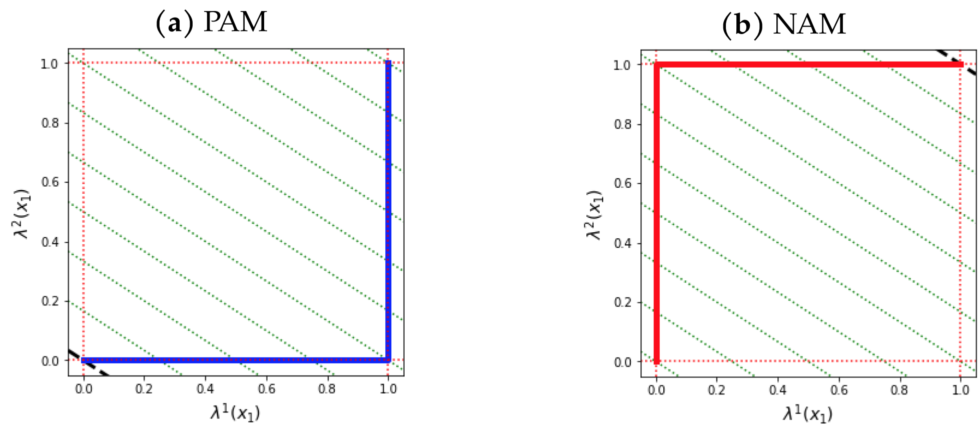


Figure 4: Illustration of Sufficient Conditions for Monotone Sorting

Next, we study the optimal transfer scheme. By fixing the optimal allocations $\{\lambda^{k*}(x_1), \lambda^{k*}(x_2)\}_{k=1}^{2}$ from Equation 2, the designer solves the following:

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

subject to [FC], [MT], and [PC] from Equation 1. The following proposition states that the transfer scheme is characterize 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 B.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 in practice.

In case of an interior solution for at least one license, the optimal transfers 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 that, 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:

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

Therefore, as we can see, equilibrium transfers depend on other features of the market such as number of children, number of parents, and the resulting meeting probabilities.

Lastly, before concluding this sections, we present two examples. One illustrating an environment where, in equilibrium, super-modularity does not imply PAM. Another example presents an environment where the sufficient conditions for monotone sorting described in Corollary 1 hold.

Example 1. (***Positive Assortative Matching Fails***)**.** Figure 5 illustrates an environment where super-modularity in the surplus function $S(x, y)$ is not a sufficient condition for PAM. Here, 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 $S(x_1, y_1) = 51$.[13] Here, the condition over primitives presented in Corollary 1(i) is violated:

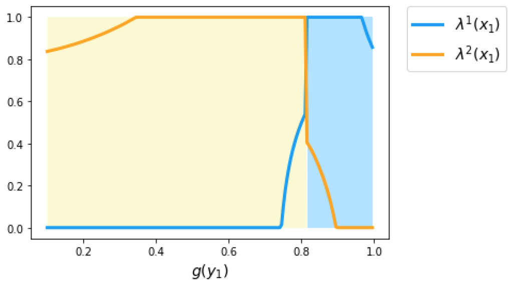
$$1 = \frac{S(x_2, y_2) - S(x_2, y_1)}{S(x_1, y_2) - S(x_1, y_1)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)} = 5.99$$

Panel 5a 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 5b, 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

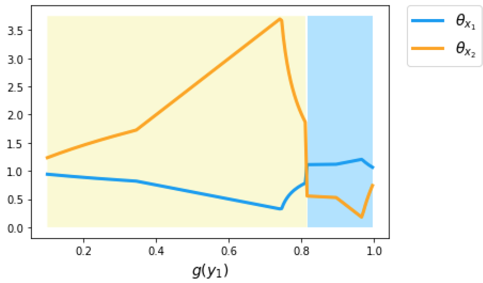
[13] The share of low-needs children is similar to the one observed in practice. See Appendix A.

the market tightness. In Panel 5c, 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- shared regions correspond to PAM and NAM, respectively.

(**a**) Randomization Device



(**b**) Market Tightness



(**c**) Expected Transfers

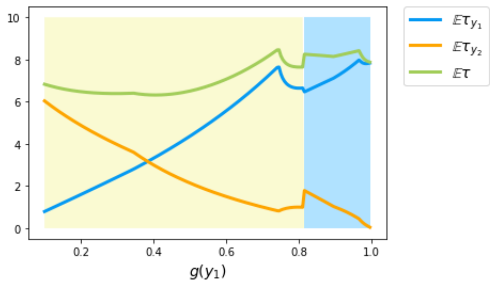


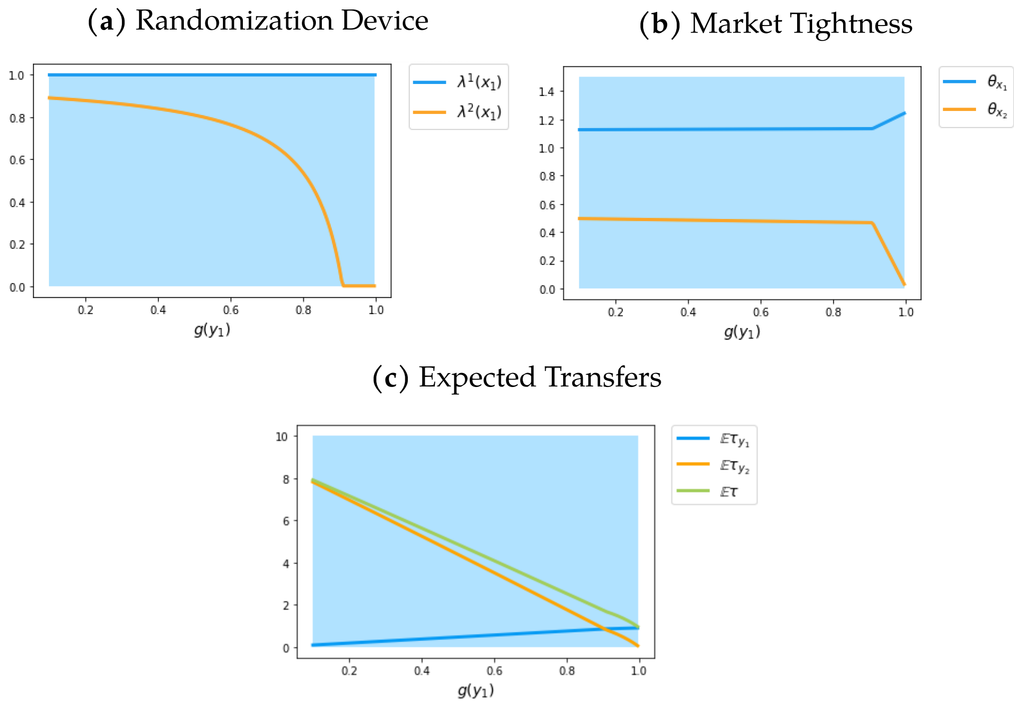
Figure 5: Monotone Sorting Fails

As Panel 5a illustrates, for small enough values of $g(y_1)$, the equilibrium sorting exhibits NAM, even when the surplus function is super-modular. Thus, supermodularity is not a sufficient condition for PAM to hold in equilibrium. For the same interval of $g(y_1)$, Panel 5b 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 5c that the total expected cost of imposing NAM increases as the share of low-ability parents increases. This is intuitive, since low-ability parents incur in a greater cost of providing care than high-ability parents, the designer must pay greater transfers to low-ability parents to satisfy the [PC]. □

Example 2. (***Positive Assortative Matching Holds***)**.** Figure 6 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 primitives satisfy the following:

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

As Panel 6a illustrates the equilibrium sorting exhibits PAM for any value of $g(y_1)$. Moreover, for sufficiently high values of $g(y)$, there is a perfect segregation of the market, where all type-y_1 (y_2) parents are allocated into submarket x_1 (x_2). In Panel 6b, we observe 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. Here, we have 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. Note that, market tightness is larger in submarket x_1 than in x_2 for any value of $g(y_1)$, resulting in parents being more likely to meet a child in submarket x_2. Interestingly, in Panel 6c, the total expected cost of implementing PAM is decreasing in the share of low-ability parents, unlike the intuition presented in the previous example. Here, we have assumed that $c(x_2, y_2) = 13$, $c(x_1, y_2) = 1$, $c(x_2, y_1) = 20$ and $c(x_1, y_1) = 2$. Thus, y_1-parents receive a transfer equal to 2 while y_2-parents receive $13 - \frac{\pi^p(\theta_1^\star)\lambda^{2\star}(x_1)}{\pi^p(\theta_2^\star)\lambda^{2\star}(x_2)}$. Note that, the expected transfers receive by y_2-parents (orange-line) exhibits two effect when $g(y_1)$ increases: **(i)** since the expected transfer depends on the share of y_2-parents, it decreases as $g(y_1)$ increases, and **(ii)** since the expected transfer depends on the transfer received by each y_2-parent, it increases as $\lambda^{2\star}(x_1)$ decreases. In this environment, (i) overcomes (ii). Hence, given our cost function assumption, even though the expected transfers received by all y_1-parents (blue-line) are increasing with $g(y)$, this is not enough to compensate the decrease of the expected transfers received by all y_2-parents (orange-line).



(a) Randomization Device

(b) Market Tightness

(c) Expected Transfers

Figure 6: Monotone Sorting Holds

□

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. Note that, by assumption, high-ability parents incur in a smaller cost when providing care than low-ability parents do, regardless of the disability status of the child. This translates to high-ability parents receiving a smaller expected transfer under the menu of licenses specified under complete information. Thus, in the presence of private information, high-ability parents have incentives to mimic low-ability parents in order to receive greater transfers, regardless of the sorting pattern. Therefore, in equilibrium, [PC] holds with equality for low-ability parents, and [IC] holds with equality for high-ability parents. After incorporating these two constraints into the objective function in Equation 1, the designer's problem reduces to:

$$\max_{\{\lambda^k(x_1),\lambda^k(x_2)\}_{k=1}^2} \Bigg\{ \sum_{i=1}^{2} \pi^p(\theta_{x_i}) \left[\sum_{k=1}^{2} \Big(\underbrace{u(x_i, y_\mathbf{k}) - c(x_i, y_\mathbf{k})}_{S(x,y)} \Big) \lambda^k(x_i)\, g(y_\mathbf{k}) \right]$$
$$- \Big[c(x_1, y_1) - c(x_1, y_2)\Big] \lambda^1(x_1) \pi^p(\theta_1) g(y_2) - \Big[c(x_2, y_1) - c(x_2, y_2)\Big] \lambda^1(x_2) \pi^p(\theta_2) g(y_2) \Bigg\} \tag{4}$$

subject to [FC], [MT], and [IC] for low-ability parents.[14]

We start our analysis by establishing that Lemma 1 and Proposition 1 presented in the previous section hold under private information (See Appendix C.1 and C.2). Following the same arguments as we did previously, the term that characterizes the optimal allocation of parents across submarkets becomes:

$$Z^{PI}(\theta_1) = \pi^p(\theta_2) \left(\left[u(x_2, y_2) - \frac{c(x_2, y_2)}{g(y_1)} \right] - \left[u(x_2, y_1) - \frac{c(x_2, y_1)}{g(y_1)} \right] \right)$$
$$- \pi(\theta_1) \left(\left[u(x_1, y_2) - \frac{c(x_1, y_2)}{g(y_1)} \right] - \left[u(x_1, y_1) - \frac{c(x_1, y_1)}{g(y_1)} \right] \right) \tag{5}$$

where $Z^{PI}(\theta_1)$ is analogous to $Z^{CI}(\theta_1)$, adjusted by the cost due to information friction. Recall,

$$Z^{CI}(\theta_1) = \pi^p(\theta_2) \Big(\big[u(x_2, y_2) - c(x_2, y_2)\big] - \big[u(x_2, y_1) - c(x_2, y_1)\big] \Big)$$
$$- \pi^p(\theta_1) \Big(\big[u(x_1, y_2) - c(x_1, y_2)\big] - \big[u(x_1, y_1) - c(x_1, y_1)\big] \Big) \tag{6}$$

A couple of remarks worth mentioning. First, if $g(y_1) = 1$ then Equations 5 and 6 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. Lastly, as $g(y_2)$ approaches to one (equivalently as $g(y_1)$ approaches to zero), then the cost function becomes the parameter of interest determining the equilibrium sorting pattern, unlike the case of complete information. Therefore, the sign of $Z^{PI}(\theta_1)$ will determine the equilibrium sorting.

Now, we present sufficient conditions for monotone sorting under private in-

[14] Notice, [PC] for low-ability parents and [IC] for high-ability parents imply [PC] for high-ability parents, see Proposition 3 in Appendix C.5.

formation, analogous to Corollary 1:[15]

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 C.3. □

Unlike complete information, the surplus of a match is not a sufficient statistics to ensure monotone sorting under the presence of private information. Here, we need to take into account the cost of a match as well as the exogenous distribution of parents. In Corollary 2(i), the second condition over the cost function ensures that the incentive-compatibility constrains are satisfied. We require for $c(x,y)$ to be a sub-modular function, that is, the difference in the cost of providing care for a child x_2 and child x_1 must be greater for low-ability than for high-ability parents. A reasonable condition since we would expect to see that in practice. Note that, if the cost function is sub-modular, the informational rents paid to high-ability parents are lower under PAM than NAM. In other words, it is easier for the designer to incentivize high-ability parents to report truthfully under PAM. The first condition in Corollary 2(i), ensures that we will observe PAM in equilibrium. Under both conditions, even if the surplus function is sub-modular we will observe PAM. Thus, by imposing sub-modularity in $c(x,y)$, we can relax the super-modularity assumption over $S(x,y)$, which is intuitive since this condition over $c(x,y)$ moves the sorting towards PAM for the reason mentioned above. Analogous for (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 the share of low-ability parents in the market:

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(i) requires $S(x,y)$ to be a *strong* super-modular function as in Corollary 1(i). This condition is ensuring that PAM will maximize the total expected surplus. Also, we impose the condition of *strong* sub-modularity in $c(x,y)$

[15] See Appendix C.4 for a detailed characterization.

to ensure incentive-compatibility. Thus, we are adding an extra condition to the complete information result. Similar arguments and intuition follows for NAM.

Now, 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 4, the designer solves the following:

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

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 3. *Suppose conditions in Corollary 2 holds. 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 constraint of y_1-type parent as well as the incentive-compatibility constraint of y_2-type parent hold by equality is an equilibrium.*

Proof. See Appendix C.5. □

When the equilibrium sorting exhibits perfect PAM, the optimal transfer scheme is $\tau^1(x_1) = c(x_1, y_1)$ 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, when information frictions are introduced, it is no longer optimal to just transfer parents the cost of providing care, but screening requires to compensate high-ability parents to disclose their type truthfully. Now, suppose that the equilibrium sorting exhibits high-type PAM. In this case, the optimal transfers are $\tau^1(x_1) = c(x_1, y_1)$ and $\tau^2(x_2) = c(x_2, y_2) - [\tau^2(x_1) - c(x_1, y_2)]\frac{\pi^p(\theta_1)\lambda^2(x_1)}{\pi^p(\theta_2)\lambda^2(x_2)} + [c(x_1, y_1) - c(x_1, y_2)]\frac{\pi^p(\theta_1)}{\pi^p(\theta_2)\lambda^2(x_2)}$. Compared to the complete information setting, we also observe a positive extra term in the transfer for high-ability parents who provide care in submarket x_2, reflecting the incentive for high-ability parents to reveal their true type. Therefore, we can see that high-ability parents receive information rents.

Next, we present two examples illustrating the same environment as in Examples 1 and 2, incorporating private information.

Example 3. (***Positive Assortative Matching Fails***). Figure 7 illustrates the environment in Example 1, where super-modularity in the surplus function is not a sufficient condition for PAM. In both panels, the solid-lines represent the equilibrium objects under the complete information, while the dash-lines corresponds to the private information. We assume that the cost function is super-modular with 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.

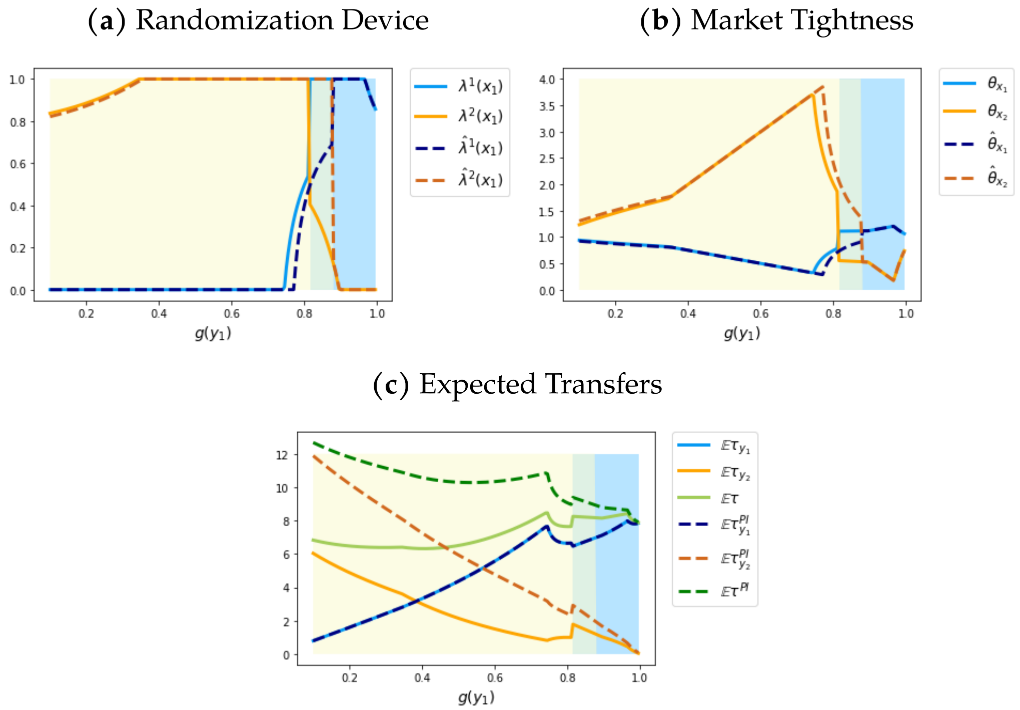


Figure 7: Monotone Sorting Fails under Private Information

Almost for any value of $g(y_1)$, the sorting patterns are the same in the complete information case as in the private information one. However, when $g(y_1)$ is approximately in $(0.8, 0.9)$, the equilibrium sorting pattern is PAM under the complete information, whereas it is NAM with the private information. To see the intuition, consider an equilibrium menu of licenses that implements perfect sorting under complete information, such that $c(x, y) = \tau^y(x)$. If the menu implements NAM, y_2 mimics y_1 and matches with x_2 instead of x_1, and if it implements PAM, y_2 mimics y_1 and matches with x_1 instead of x_2.[16] The former misreport allows y_2 to (ex-post) gain as much as $\tau^{y_1}(x_2) - c(x_2, y_2) = 5$ whereas the latter does $\tau^{y_1}(x_1) - c(x_1, y_2) = 14$. That is, y_2 has *stronger* incentives to misreport if the equilibrium sorting is PAM than when is NAM under complete information. Thus, it is cheaper for the designer to switch the equilibrium sorting from PAM to NAM for the (roughly) specified region of $g(y_1)$.

[16]For exposition, y_i represents type-y_i parent and x_i represents type-x_i child for each $i = 1, 2$.

It is important to highlight that, in this environment, the optimal randomization device, $\lambda^1(x_1)$ and $\lambda^2(x_2)$, is very similar for the complete and private information cases. As a result, the optimal market tightness coincides for a fair big interval of $g(y)$. However, note that implementing PAM is more expensive for private than for incomplete information cases. Thus, the designer is able to reach the same allocation as with complete information but at a greater cost.

□

Example 4. (***Positive Assortative Matching Holds***)**.** Figure 8 illustrates an environment as in Example 2 that satisfies the additional conditions presented in Corollary 3(i). We assume $c(x,y)$ is a sub-modular function with values $c(x_2,y_2) = 13$, $c(x_1,y_2) = 1$, $c(x_2,y_1) = 20$ and $c(x_1,y_1) = 2$.

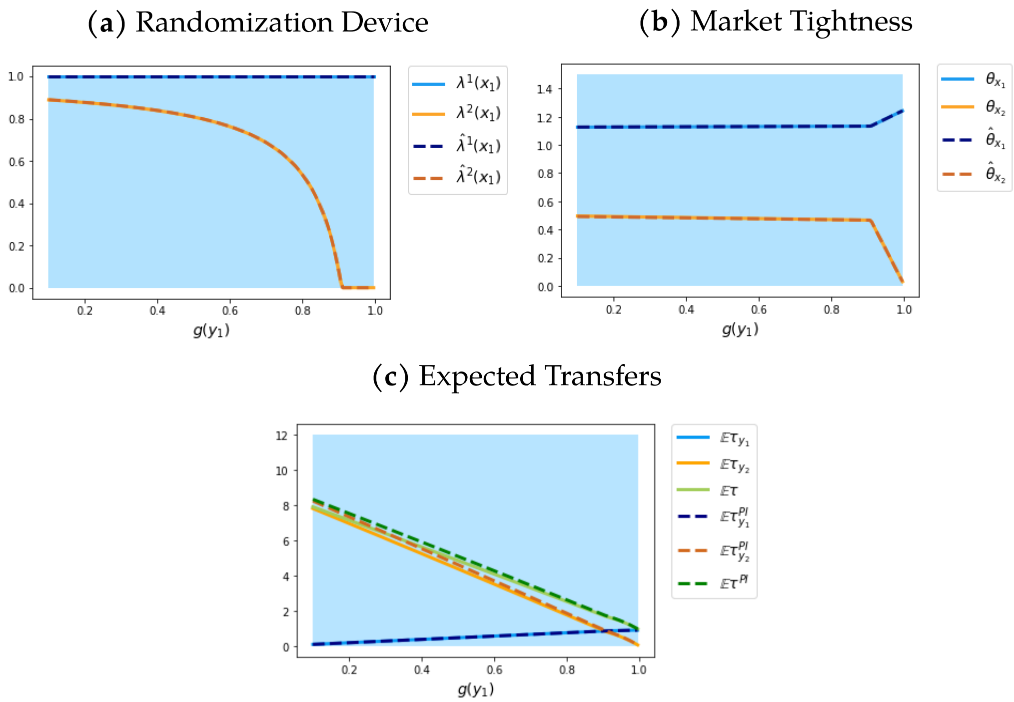


Figure 8: Monotone Sorting Holds under Private Information

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:

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

As Panel 8a illustrates the equilibrium sorting exhibits PAM for any value of $g(y_1)$. In Panel 8b, we observe that the market tightness in both submarkets remains flat for a fair range of values of $g(y_1)$. As the cost function is sub-modular, it is easier for the designer to incentivize high-ability parents to report truthfully under PAM. Thus, the equilibrium sorting (PAM) is robust to information friction here, unlike it is in Example 3. Lastly, Panel 8c we observe that its more expensive to impose PAM under private information. □

5 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 from each other in their ability to take care of a child. We solve for the optimal menu of licenses which specifies allocation of parents across submarkets of children as well as corresponding transfers, under the presence of search and information frictions.

There are two main results of the paper 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 rationalizes the nested licenses in the foster care system 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 rent 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 high, then not only the allocation diverge from the first-best, but also the optimal sorting may reverse as in Example 3.

References

ACKERBERG, D. AND M. BOTTICINI (2002): "Endogenous Matching and the Empirical Determinants of Contract Form," *The Journal of Political Economy*, 110, 564–592.

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

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.

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.

FRANCO, A., M. MITCHELL, AND G. VERESHCHAGINA (2011): "Incentives and the Structure of Teams," *The Journal of Economic Theory*, 146, 2307–2332.

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.

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

KAYA, A. AND G. VERESHCHAGINA (2014): "Partnerships versus Corporations: Moral Hazard, Sorting, and Ownership Structure," *American Economic Review*, 104, 291–307.

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*.

Serfes, K. (2005): "Risk Sharing vs. Incentives: Contract Design Under Two-sided Heterogeneity," *Economics Letters*, 88, 343–349.

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),[17] 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,[18] 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.[19] Thus, based on the disability variable, we can infer that at least 24 percent of children in the US foster care are special needs.[20] 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 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 is 44.1 years old, 88.2 percent are

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

[18] 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.

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

[20] 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.

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 a year 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 foster parent who are certified, though 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: 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:

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

subject to:

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

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

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

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

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

Next, by rearranging terms from the objective function:

$$\sum_{i=1}^{2} \pi^p\big(\theta_i\big) \, \sum_{k=1}^{2} u(x_i,y_k)\lambda^k(x_i) g(y_k) - \sum_{i=1}^{2} \pi^p\big(\theta_i\big) \sum_{k=1}^{2} \tau^k(x_i)\lambda^k(x_i) g(y_k)$$
$$\Rightarrow \sum_{i=1}^{2} \pi^p\big(\theta_i\big) \, \sum_{k=1}^{2} u(x_i,y_k)\lambda^k(x_i) g(y_k)$$
$$- \left[\sum_{i=1}^{2} \tau^1(x_i)\lambda^1(x_i)\pi^p\big(\theta_i\big) g(y_1) + \sum_{i=1}^{2} \tau^2(x_i)\lambda^2(x_i)\pi^p\big(\theta_i\big) g(y_2) \right]$$

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

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

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

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

subject to:

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

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

The following corollary is immediate:

Corollary B.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. □

B.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:

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

where:

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

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:

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

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 B.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:

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

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

B.2. Now, let's tremble $\lambda^2(x_1)$ by ε_2, and calculate welfare:

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

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 B.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.

B.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.

B.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:

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

B.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 B.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

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

Fix $\lambda^{-k}(x_1)$. Increasing $\lambda^k(x_1)$ by a small amount $\varepsilon > 0$, increases $\mathbb{E}U_1$ 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 B.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 B.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 B.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 B.1 and B.2, we characterize the equilibrium allocation of parents across submarket step by step. Initially, we establish the equilibrium allocation of parents when the sufficient conditions of Corollary 1 hold. Later, we extend the analysis to the case where the sufficient conditions are violated.

Proposition B.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*

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

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

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

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

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

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. ☐

One can easily characterize the equilibrium distribution of parents across sub-markets for NAM, with a parallel argument.

Proposition B.2 (**Negative Assortative Matching (NAM)**). *Suppose* $\frac{S(x_1,y_2)-S(x_1,y_1)}{S(x_2,y_2)-S(x_2,y_1)} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_1)}\right)}$ *holds. The equilibrium sorting exhibits:*

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

$$\frac{\partial W\left(\lambda^1(x_1), \lambda^2(x_1)\right)}{\partial \lambda^1(x_1)}|_{\{\lambda^2(x_1)=1\}} = 0 \text{ for some } \lambda^{1\star}(x_1) \in (0,1) \tag{B.6}$$

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

$$\frac{\partial W\left(\lambda^1(x_1), \lambda^2(x_1)\right)}{\partial \lambda^2(x_1)}|_{\{\lambda^1(x_1)=0,\lambda^2(x_1)=1\}} \geq 0 \geq \frac{\partial W\left(\lambda^1(x_1), \lambda^2(x_1)\right)}{\partial \lambda^1(x_1)}|_{\{\lambda^1(x_1)=0,\lambda^2(x_1)=1\}} \tag{B.7}$$

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

$$\frac{\partial W\left(\lambda^1(x_1), \lambda^2(x_1)\right)}{\partial \lambda^2(x_1)}|_{\{\lambda^1(x_1)=0\}} = 0 \text{ for some } \lambda_1^{2^*} \in (0,1) \tag{B.8}$$

Proof. QED following the same arguments as in Proposition B.1. ☐

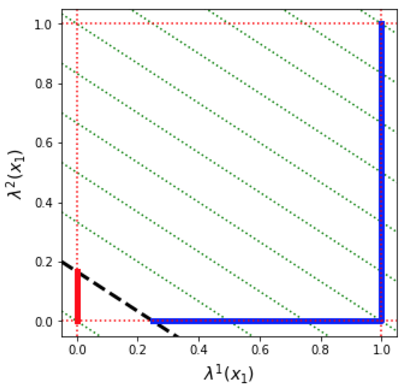
Propositions B.1 and B.2 characterize the equilibrium sorting patterns when the conditions specified in Corollary 1 hold. Now, we extend the analysis to the case where the sufficient conditions are violated. As the dashed-lines in Figure B.1, suppose that $Z^{CI}(\bar{\theta}_1) = 0$ for some $\bar{\theta}_1 \in \left(0, \frac{1}{f(x_1)}\right)$. Then, for NAM (red-lines in Figure B.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 B.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 B.1.

Case 1. Fix $\tilde{\lambda}^1(x_1) = 0$ and $\tilde{\lambda}^2(x_1) \leq 1$. We consider sub-cases of $\hat{\lambda}^1(x_1)$:

(1A) Consider $\hat{\lambda}^1(x_1) \leq 1$ and $\hat{\lambda}^2(x_1) = 0$ [Panel B.1a]. By starting from a particular corner $\lambda^1(x_1) = 0$ and $\lambda^2(x_1) = 0$, we initially increase $\lambda^2(x_1)$ until it is either not profitable any more or we switch from NAM to PAM by increasing $\lambda^1(x_1)$ instead. Now, the following characterizes the equilibrium sorting for this particular case:

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

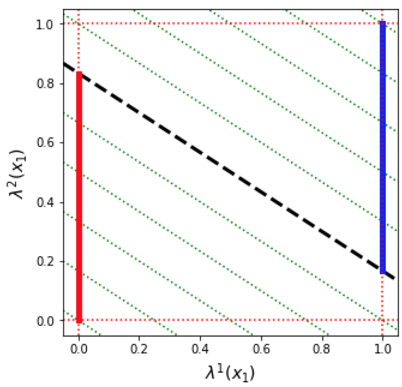
(a) Case 1A



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

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

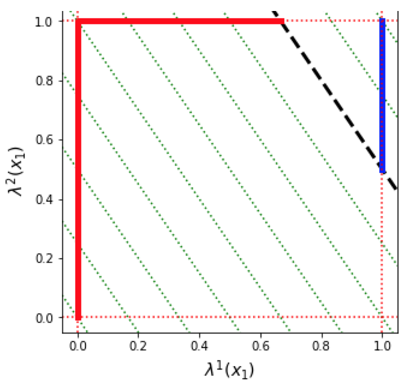
(b) Case 1B



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

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

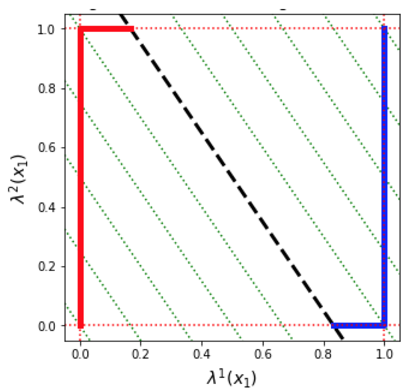
(c) Case 2A



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

Proposition B.3. *Consider* $\tilde{\lambda}^1(x_1) = 0$ *and* $\tilde{\lambda}^2(x_1) \leq 1$, *and* $\hat{\lambda}^1(x_1) \leq 1$ *and* $\hat{\lambda}^2(x_1) = 0$.

i. Suppose $\lambda^{1\star}(x_1) = 0$ *and* $\lambda^{2\star}(x_1) < \tilde{\lambda}^2(x_1)$, *then the equilibrium sorting exhibits high-type NAM if* $\lambda^{2\star}(x_1) \in [0, \tilde{\lambda}^2(x_1))$ *solves Equation B.8*

ii. Suppose $\lambda^{2\star}(x_1) = 0$ *and* $\lambda^{1\star}(x_1) > \hat{\lambda}^1(x_1)$, *then the equilibrium sorting exhibits:*

a. low-type PAM if $\lambda^{1\star}(x_1) \in (\hat{\lambda}^1(x_1), 1)$ *solves Equation B.3*

b. perfect PAM if $\lambda^{1\star}(x_1) = 1$ *solves Equation B.4*

iii. Suppose $\lambda^{1\star}(x_1) = 1$ *and* $\lambda^{2\star}(x_1) > 0$, *then the equilibrium sorting exhibits high-type PAM if* $\lambda^{2\star}(x_1) \in (0, 1]$ *solves Equation B.5*

(1B) Consider $\hat{\lambda}^1(x_1) = 1$ and $\hat{\lambda}^2(x_1) \geq 0$ [Panel B.1b]. Following a symmetric argument, the characterization of this case follows:

Proposition B.4. *Consider* $\tilde{\lambda}^1(x_1) = 0$ *and* $\tilde{\lambda}^2(x_1) \leq 1$, *and* $\hat{\lambda}^1(x_1) = 1$ *and* $\hat{\lambda}^2(x_1) \geq 0$.

i. *Suppose* $\lambda^{1\star}(x_1) = 0$ *and* $\lambda^{2\star}(x_1) < \tilde{\lambda}^2(x_1)$, *then the equilibrium sorting exhibits high-type NAM if* $\lambda^{2\star}(x_1) \in [0, \tilde{\lambda}^2(x_1))$ *solves Equation B.8*

ii. *Suppose* $\lambda^{1\star}(x_1) = 1$ *and* $\lambda^{2\star}(x_1) > \hat{\lambda}^2(x_1)$, *then the equilibrium sorting exhibits high-type PAM if* $\lambda^{2\star}(x_1) \in (\hat{\lambda}^2(x_1), 1]$ *solves Equation B.5*

Case 2. Fix $\tilde{\lambda}^1(x_1) \geq 0$ and $\tilde{\lambda}^2(x_1) = 1$. We consider sub-cases of $\hat{\lambda}^1(x_1)$:

(2A) Consider $\hat{\lambda}^1(x_1) = 1$ and $\hat{\lambda}^2(x_1) \geq 0$ [Panel B.1c]. Following a symmetric argument, the characterization of this case follows:

Proposition B.5. *Consider* $\tilde{\lambda}^1(x_1) \geq 0$ *and* $\tilde{\lambda}^2(x_1) = 1$, *and* $\hat{\lambda}^1(x_1) = 1$ *and* $\hat{\lambda}^2(x_1) \geq 0$.

i. *Suppose* $\lambda^{1\star}(x_1) = 1$ *and* $\lambda^{2\star}(x_1) > \hat{\lambda}^2(x_1)$, *then the equilibrium sorting exhibits high-type PAM if* $\lambda^{2\star}(x_1) \in (\hat{\lambda}^2(x_1), 1]$ *solves Equation B.5*

ii. *Suppose* $\lambda^{2\star}(x_1) = 1$ *and* $\lambda^{1\star}(x_1) < \tilde{\lambda}^1(x_1)$, *then the equilibrium sorting exhibits:*
   a. *low-type NAM if* $\lambda^{1\star}(x_1) \in (0, \tilde{\lambda}^1(x_1))$ *solves Equation B.6*
   b. *perfect NAM if* $\lambda^{1\star}(x_1) = 0$ *solves Equation B.7*

iii. *Suppose* $\lambda^{1\star}(x_1) = 0$ *and* $\lambda^{2\star}(x_1) < 1$, *then the equilibrium sorting exhibits high-type NAM if* $\lambda^{2\star}(x_1) \in [0, 1)$ *solves Equation B.8.*

(2B) Consider $\hat{\lambda}^1(x_1) \leq 1$ and $\hat{\lambda}^2(x_1) = 0$ [Panel B.1d]. The following characterizes the equilibrium sorting for this particular case:

Proposition B.6. *Consider* $\tilde{\lambda}^1(x_1) \geq 0$ *and* $\tilde{\lambda}^2(x_1) = 1$, *and* $\hat{\lambda}^1(x_1) \leq 1$ *and* $\hat{\lambda}^2(x_1) = 0$.

i. *Suppose* $\lambda^{2\star}(x_1) = 1$ *and* $\lambda^{1\star}(x_1) < \tilde{\lambda}^1(x_1)$, *then the equilibrium sorting exhibits:*
   a. *low-type NAM if* $\lambda^{1\star}(x_1) \in (0, \tilde{\lambda}^1(x_1))$ *solves Equation B.6*
   b. *perfect NAM if* $\lambda^{1\star}(x_1) = 0$ *solves Equation B.7*

ii. *Suppose* $\lambda^{1\star}(x_1) = 0$ *and* $\lambda^{2\star}(x_1) < 1$, *then the equilibrium sorting exhibits high-type NAM if* $\lambda^{2\star}(x_1) \in [0, 1)$ *solves Equation B.8*

iii. *Suppose* $\lambda^{2\star}(x_1) = 0$ *and* $\lambda^{1\star}(x_1) > \hat{\lambda}^1(x_1)$, *then the equilibrium sorting exhibits:*
   a. *low-type PAM if* $\lambda^{1\star}(x_1) \in (\hat{\lambda}^1(x_1), 1)$ *solves Equation B.3*
   b. *perfect PAM if* $\lambda^{1\star}(x_1) = 1$ *solves Equation B.4*

iv. *Suppose* $\lambda^{1\star}(x_1) = 1$ *and* $\lambda^{2\star}(x_1) > 0$, *then the equilibrium sorting exhibits high-type PAM if* $\lambda^{2\star}(x_1) \in (0, 1]$ *solves Equation B.5*

B.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.

C Appendix: 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 $k \neq k' = 1, 2$:

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

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

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

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 3):

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

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

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:

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

subject to [FC], [MT], and

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

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 3).

Corollary C.1. *In the private information setting, 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 under Private Information

We can establish Lemma 1 for the private information case.

Lemma C.1. *In the private information setting, for at least one of the licenses, the optimal randomization rule yields a corner solution whenever* $S(x,y)$ *is super- or sub-modular.*

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:

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

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:

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

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 C.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:

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

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:

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

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

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

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.

C.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 C.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.

C.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 C.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 C.1, we can ensure NAM by imposing that the following inequality must hold:

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

C.4 Assortative Matching in Equilibrium

First, we show that Lemmas B.1 and B.2 carry over to the case of private information.

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

Proof. Recall the 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 C.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 C.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 C.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:

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

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 C.2 and C.3, we characterize the equilibrium allocation of parents across submarkets as in the complete information case. Initially, we establish the equilibrium allocation of parents when the sufficient conditions of Corollary 2 hold. Later, we extend the analysis to the case where sufficient conditions in Corollary 2 are violated.

Proposition C.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*

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

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

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

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

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

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 C.3 (**Negative Assortative Matching (NAM)**)**.** *Suppose that* $\frac{S(x_1,y_2)-S(x_1,y_1)+\frac{g(y_2)}{g(y_1)}\cdot[c(x_1,y_1)-c(x_1,y_2)]}{S(x_2,y_2)-S(x_2,y_1)+\frac{g(y_2)}{g(y_1)}\cdot[c(x_2,y_1)-c(x_2,y_2)]} \geq \frac{1}{\pi^p\left(\frac{1}{f(x_1)}\right)}$ *holds. The equilibrium exhibits:*

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

$$\frac{\partial \hat{W}(\lambda^1(x_1), \lambda^2(x_1))}{\partial \lambda^1(x_1)}|_{\{\lambda^2(x_1=1\}} = 0 \ \textit{for some} \ \lambda^{1\star}(x_1) \in (0,1) \tag{C.6}$$

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

$$\frac{\partial \hat{W}(\lambda^1(x_1), \lambda^2(x_1))}{\partial \lambda^2(x_1)}|_{\{\lambda^1(x_1)=0,\lambda^2(x_1)=1\}} \geq 0 \geq \frac{\partial \hat{W}(\lambda^1(x_1), \lambda^2(x_1))}{\partial \lambda^1(x_1)}|_{\{\lambda^1(x_1)=0,\lambda^2(x_1)=1\}} \tag{C.7}$$

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

$$\frac{\partial \hat{W}(\lambda^1(x_1), \lambda^2(x_1))}{\partial \lambda^2(x_1)}|_{\{\lambda^1(x_1=1\}} = 0 \ \textit{for some} \ \lambda^{2\star}(x_1) \in (0,1) \tag{C.8}$$

Proof. QED following the same arguments in the proof of Proposition C.2 □

Propositions C.2 and C.3 characterize the equilibrium sorting patterns when the conditions specified in Corollary 2 hold. Now, we extend the analysis to the case where the sufficient conditions are violated.

Case 1. Fix $\tilde{\lambda}^1(x_1) = 0$ and $\tilde{\lambda}^2(x_1) \leq 1$. We consider sub-cases of $\hat{\lambda}^1(x_1)$:

(1A) Consider $\hat{\lambda}^1(x_1) \leq 1$ and $\hat{\lambda}^2(x_1) = 0$ [Panel B.1a]. By starting from a particular corner $\lambda^1(x_1) = 0$ and $\lambda^2(x_1) = 0$, we initially increase $\lambda^2(x_1)$ until it is either not profitable any more or we switch from NAM to PAM by increasing $\lambda^1(x_1)$ instead. Now, the following characterizes the equilibrium sorting for this particular case:

Proposition C.4. *Consider* $\tilde{\lambda}^1(x_1) = 0$ *and* $\tilde{\lambda}^2(x_1) \leq 1$*, and* $\hat{\lambda}^1(x_1) \leq 1$ *and* $\hat{\lambda}^2(x_1) = 0$*.*

i. Suppose $\lambda^{1\star}(x_1) = 0$ *and* $\lambda^{2\star}(x_1) < \tilde{\lambda}^2(x_1)$*, then the equilibrium sorting exhibits high-type NAM if* $\lambda^{2\star}(x_1) \in [0, \tilde{\lambda}^2(x_1))$ *solves Equation C.8*

ii. Suppose $\lambda^{2\star}(x_1) = 0$ *and* $\lambda^{1\star}(x_1) > \hat{\lambda}^1(x_1)$*, then the equilibrium sorting exhibits:*

a. low-type PAM if $\lambda^{1\star}(x_1) \in (\hat{\lambda}^1(x_1), 1)$ *solves Equation C.3*

b. perfect PAM if $\lambda^{1\star}(x_1) = 1$ *solves Equation C.4*

iii. Suppose $\lambda^{1\star}(x_1) = 1$ *and* $\lambda^{2\star}(x_1) > 0$*, then the equilibrium sorting exhibits high-type PAM if* $\lambda^{2\star}(x_1) \in (0, 1]$ *solves Equation C.5*

(1B) Consider $\hat{\lambda}^1(x_1) = 1$ and $\hat{\lambda}^2(x_1) \geq 0$ [Panel B.1b]. Following a symmetric argument, the characterization of this case follows:

Proposition C.5. *Consider* $\tilde{\lambda}^1(x_1) = 0$ *and* $\tilde{\lambda}^2(x_1) \leq 1$*, and* $\hat{\lambda}^1(x_1) = 1$ *and* $\hat{\lambda}^2(x_1) \geq 0$*.*

i. Suppose $\lambda^{1\star}(x_1) = 0$ *and* $\lambda^{2\star}(x_1) < \tilde{\lambda}^2(x_1)$*, then the equilibrium sorting exhibits high-type NAM if* $\lambda^{2\star}(x_1) \in [0, \tilde{\lambda}^2(x_1))$ *solves Equation C.8*

ii. Suppose $\lambda^{1\star}(x_1) = 1$ *and* $\lambda^{2\star}(x_1) > \hat{\lambda}^2(x_1)$*, then the equilibrium sorting exhibits high-type PAM if* $\lambda^{2\star}(x_1) \in (\hat{\lambda}^2(x_1), 1]$ *solves Equation C.5*

Case 2. Fix $\tilde{\lambda}^1(x_1) \geq 0$ and $\tilde{\lambda}^2(x_1) = 1$. We consider sub-cases of $\hat{\lambda}^1(x_1)$:

(2A) Consider $\hat{\lambda}^1(x_1) = 1$ and $\hat{\lambda}^2(x_1) \geq 0$ [Panel B.1c]. Following a symmetric argument, the characterization of this case follows:

Proposition C.6. *Consider* $\tilde{\lambda}^1(x_1) \geq 0$ *and* $\tilde{\lambda}^2(x_1) = 1$*, and* $\hat{\lambda}^1(x_1) = 1$ *and* $\hat{\lambda}^2(x_1) \geq 0$*.*

i. Suppose $\lambda^{1\star}(x_1) = 1$ *and* $\lambda^{2\star}(x_1) > \hat{\lambda}^2(x_1)$*, then the equilibrium sorting exhibits high-type PAM if* $\lambda^{2\star}(x_1) \in (\hat{\lambda}^2(x_1), 1]$ *solves Equation C.5*

ii. Suppose $\lambda^{2\star}(x_1) = 1$ *and* $\lambda^{1\star}(x_1) < \tilde{\lambda}^1(x_1)$*, then the equilibrium sorting exhibits:*

a. low-type NAM if $\lambda^{1\star}(x_1) \in (0, \tilde{\lambda}^1(x_1))$ *solves Equation C.6*

b. perfect NAM if $\lambda^{1\star}(x_1) = 0$ *solves Equation C.7*

iii. Suppose $\lambda^{1\star}(x_1) = 0$ *and* $\lambda^{2\star}(x_1) < 1$*, then the equilibrium sorting exhibits high-type NAM if* $\lambda^{2\star}(x_1) \in [0, 1)$ *solves Equation C.8.*

(2B) Consider $\hat{\lambda}^1(x_1) \leq 1$ and $\hat{\lambda}^2(x_1) = 0$ [Panel B.1d]. The following characterizes the equilibrium sorting for this particular case:

Proposition C.7. *Consider $\tilde{\lambda}^1(x_1) \geq 0$ and $\tilde{\lambda}^2(x_1) = 1$, and $\hat{\lambda}^1(x_1) \leq 1$ and $\hat{\lambda}^2(x_1) = 0$.*

i. Suppose $\lambda^{2\star}(x_1) = 1$ and $\lambda^{1\star}(x_1) < \tilde{\lambda}^1(x_1)$, then the equilibrium sorting exhibits:

a. low-type NAM if $\lambda^{1\star}(x_1) \in (0, \tilde{\lambda}^1(x_1))$ solves Equation C.6

b. perfect NAM if $\lambda^{1\star}(x_1) = 0$ solves Equation C.7

ii. Suppose $\lambda^{1\star}(x_1) = 0$ and $\lambda^{2\star}(x_1) < 1$, then the equilibrium sorting exhibits high-type NAM if $\lambda^{2\star}(x_1) \in [0, 1)$ solves Equation C.8

iii. Suppose $\lambda^{2\star}(x_1) = 0$ and $\lambda^{1\star}(x_1) > \hat{\lambda}^1(x_1)$, then the equilibrium sorting exhibits:

a. low-type PAM if $\lambda^{1\star}(x_1) \in (\hat{\lambda}^1(x_1), 1)$ solves Equation C.3

b. perfect PAM if $\lambda^{1\star}(x_1) = 1$ solves Equation C.4

iv. Suppose $\lambda^{1\star}(x_1) = 1$ and $\lambda^{2\star}(x_1) > 0$, then the equilibrium sorting exhibits high-type PAM if $\lambda^{2\star}(x_1) \in (0, 1]$ solves Equation C.5

C.5 Proof of Proposition 3

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:

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

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

From these two inequalities, we get the following expression:

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

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

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

Note the following:

$$\lambda^2(x_1)-\lambda^1(x_1)=1-\lambda^2(x_2)-[1-\lambda^1(x_2)]=\lambda^1(x_2)-\lambda^2(x_2)$$

Therefore, replacing in the previous inequality:

$$\big[c(x_2,y_1)-c(x_2,y_2)\big]\cdot\big[\lambda^2(x_2)-\lambda^1(x_2)\big]\pi^p(\theta_2)\geq \big[c(x_1,y_1)-c(x_1,y_2)\big]\cdot\big[\lambda^2(x_2)-\lambda^1(x_2)\big]\pi^p(\theta_1) \quad \text{(C.9)}$$

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 C.9 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 it is satisfied if the following holds:

  $$\frac{c(x_2,y_2)-c(x_2,y_1)}{c(x_1,y_2)-c(x_1,y_1)}\geq\frac{1}{\pi^p\left(\frac{1}{f(x_2)}\right)} \quad \text{(C.10)}$$

- **Case 2:** Suppose $\lambda^2(x_2)-\lambda^1(x_2)$ is negative. Then, Equation C.9 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 it is satisfied if the following holds:

  $$\frac{c(x_1,y_2)-c(x_1,y_1)}{c(x_2,y_2)-c(x_2,y_1)}\geq\frac{1}{\pi^p\left(\frac{1}{f(x_1)}\right)} \quad \text{(C.11)}$$

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

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

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

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

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

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:

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

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 C.10 and C.11 imply the [IC] for low-ability parents. Thus, consider the [IC] for high-ability parents:

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

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

$$\begin{aligned}
c(x_1,y_2)\big[\lambda^2(x_1)-\lambda^1(x_1)\big]\pi^p(\theta_1)+c(x_2,y_2)\big[\lambda^2(x_2)-\lambda^1(x_2)\big]\pi^p(\theta_2)\\
\Rightarrow c(x_2,y_2)\big[\lambda^2(x_2)-\lambda^1(x_2)\big]\pi^p(\theta_2)-c(x_1,y_2)\big[\lambda^2(x_2)-\lambda^1(x_2)\big]\pi^p(\theta_1)\\
\Rightarrow \Big[c(x_2,y_2)\pi^p(\theta_2)-c(x_1,y_2)\pi^p(\theta_1)\Big]\cdot\big[\lambda^2(x_2)-\lambda^1(x_2)\big]
\end{aligned}$$

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

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

As previously, we need to consider the following cases:

- **Case 1:** Suppose $\lambda^2(x_2)-\lambda^1(x_2)$ is positive. Equation C.10 ensures that the following inequality holds:

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

After some algebra:

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

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

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

- **Case 2:** Suppose $\lambda^2(x_2)-\lambda^1(x_2)$ is negative. Equation C.11 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:

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

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