

Matching Multidimensional Types: Theory and Application

Veli Safak[†]

6/10/2020

Abstract

Becker (1973) presents a bilateral matching model in which scalar types describe agents. For this framework, he establishes the conditions under which positive sorting between agents' attributes is the unique market outcome. Becker's celebrated sorting result has been applied to address many economic questions. However, recent empirical studies in the fields of health, household, and labor economics suggest that agents have multiple outcome-relevant attributes. In this paper, I study a matching model with multidimensional types. I offer multidimensional generalizations of concordance and supermodularity to construct three multidimensional sorting patterns and two classes of multidimensional complementarities. For each of these sorting patterns, I identify the sufficient conditions which guarantee its optimality. In practice, we observe sorting patterns between observed attributes that are aggregated over unobserved characteristics. To reconcile theory with practice, I establish the link between production complementarities and the aggregated sorting patterns. Finally, I examine the relationship between agents' health status and their spouses' education levels among U.S. households within the framework for multidimensional matching markets. Preliminary analysis reveals a weak positive association between agents' health status and their spouses' education levels. This weak positive association is estimated to be a product of three factors: (a) an attraction between better-educated individuals, (b) an attraction between healthier individuals, and (c) a weak positive association between agents' health status and their education levels.

^{*}Carnegie Mellon University Qatar, e-mail: vsafak@andrew.cmu.edu

[†]I would like to express my gratitude to my advisor Axel Z. Anderson for his guidance, strong and continuous support, and always having faith in my capability. I would also like to thank James W. Albrecht and Luca Anderlini for their constructive comments. I would further like to thank Laurent Bouton, Dan Cao, Chris Chambers, Roger Lagunoff, Arik M. Levinson, Yusufcan Masatlioglu, Franco Peracchi, and John Rust for helpful discussions.

The attraction channel suggests that the insurance risk associated with a two-person family plan is higher than the aggregate risk associated with two individual policies.

Introduction

Becker (1973) proposes a general framework for two-sided frictionless matching models in which scalar types represent the agents on each side of the market, i.e. each agent has only one outcome-relevant attribute. A match between two agents (one from each side) generates a type-dependent matching output. A social planner¹ maximizes the aggregate output by matching the agents in pairs. There are two essential components of Becker's theory: complementarity and sorting. The matching output exhibits *strictly positive complementarity* when the marginal product of an agent strictly increases in his/her partner's type. Similarly, the matching output exhibits *strictly negative complementarity* when the marginal product of an agent strictly decreases in his/her partner's type.

Becker (1973) shows that if the matching output exhibits strictly positive complementarity, then the unique solution to the planner's problem is *positive sorting*, i.e. the highest types are matched together, then the next highest types, etc. Likewise, negative sorting is the unique solution when the matching output exhibits strictly negative complementarity. Economists have applied Becker's assortative matching results² to address several questions. For example, Kremer (1993) sheds light on the positive correlation between wages of the workers within a firm. Gabaix and Landier (2008) explain the rise in the CEOs' salaries and its connection to the increase in firms' sizes over time.

In many applications, the agents may have multiple outcome-relevant attributes. For instance, education and race in the dating/marriage market, workers' social and cognitive skills in the labor market, and doctors' listening skills for diagnosis and fostering the doctor-patient relationship in the healthcare market are some well-documented outcome-relevant attributes in literature. In the next section, I survey additional recent studies that support the presence of multiple outcome-relevant attributes. If a single index can capture all outcome-relevant information, then a unidimensional model may be suitable. However, the single index assumption is implausible in many applications. For example, Chiappori et al. (2012) analyze the U.S. marriage market by using a multidimensional matching model with an index restriction, i.e. two agents with different attributes are identical if they have the same index value calculated by an exogenous index function. Fletcher and Padron (2015) provide empirical evidence against the implications of Chiappori et al.'s (2012)

¹A decentralized version of this model with perfectly transferable utilities can easily be constructed by using the dual version of the planner's problem.

²For a recent literature review on the matching markets; I refer the readers to Chade et al. (2017).

single index assumption.

The empirical support in health, household and labor economics for multidimensional types highlight the practical importance of the multidimensional matching theory. In this paper, I present a matching model with multidimensional types and examine the link between output complementarities and sorting patterns. The only difference between Becker’s framework and the framework presented in this paper is that I allow the agents to have multiple outcome-relevant attributes. Although the proposed model is general, to ease the exposition throughout the introduction, I consider a particular labor market model in which firms and workers have only two scalar productive skills: cognitive and social. In this context, a matching distribution satisfies *global positive sorting* if and only if it exhibits positive sorting (a) between firms’ and their workers’ cognitive skills, and (b) between firms’ and their workers’ social skills. Similarly, a matching distribution satisfies *global negative sorting* if and only if it exhibits negative sorting (a) between firms’ and their workers’ cognitive skills, and (b) between firms’ and their workers’ social skills.

A naive application of Becker’s sorting result implies positive sorting between firms’ and their workers’ cognitive skills when the marginal product of each firm’s cognitive skill strictly increases in its worker’s cognitive skill. Similarly, positive sorting between firms’ and their workers’ social skills is obtained when the marginal product of each firm’s social skill strictly increases in its worker’s social skill according to Becker’s sorting result. In a multidimensional matching market, one may observe simultaneous positive complementarities between cognitive skills and social skills. However, simultaneous positive sorting between cognitive skills and social skills may not be feasible.

Consider two firms, $x = (x_c, x_s)$ and $x' = (x'_c, x'_s)$, such that $x = (10, 10)$ and $x' = (20, 20)$. Furthermore, suppose that there are two workers, $y = (y_c, y_s)$ and $y' = (y'_c, y'_s)$, such that $y = (10, 20)$ and $y' = (20, 10)$. Notice that matching x with y and x' with y' satisfies positive sorting between cognitive skills and violates positive sorting between social skills. Similarly, matching x with y' and x' with y satisfies positive sorting between social skills and violates positive sorting between cognitive skills. In this paper, I show that, conditioning on the existence, Becker’s sorting results apply: when the output function exhibits strictly positive complementarities between cognitive skills and between manual skills, if there exists a matching scheme which satisfies global positive sorting, then (a) every optimal matching scheme satisfies global positive sorting, and (b) every matching scheme satisfying global positive sorting solves the planner’s problem. I establish the optimality of global sorting for a general global sorting class in Proposition 1.

Since global positive sorting may not be feasible, I examine an alternative sorting pattern inspired by Chiappori et al. (2017). Chiappori et al. (2017) study a marriage model in which one continuous variable

(socioeconomic status) and one binary variable (smoking habit) represent the agents on each side of the market. They categorize couples into two main groups. In the first group, both men and women are non-smokers. In the second group, at least one of the spouses smokes. They assume that the matching output of a couple is the multiplication of spouses' socioeconomic status. If there is a smoker in the household, then the output is scaled down by a constant. Under this complementarity structure, Chiappori et al. (2017) predict positive sorting between agents' and their spouses' socioeconomic status within each group. Notice that one can easily apply the idea of splitting the sample into different groups and studying the sorting patterns for each group in a more general setting.

For the previous labor market example, a matching satisfies *within-group positive sorting between cognitive skills* if it exhibits positive sorting between firms' and their workers' cognitive skills for all social skill pairs of firms and workers (x_s, y_s) . Consider four firms and four workers: $\{(10, 10), (10, 20), (20, 10), (20, 20)\}$. Here, matching the $(10, \mathbf{10})$ firm with the $(10, \mathbf{10})$ worker and the $(20, \mathbf{10})$ firm with the $(20, \mathbf{10})$ worker is consistent with within-group positive sorting between cognitive skills for the $(\mathbf{10}, \mathbf{10})$ social skill combination. Within-group sorting solves the feasibility problem: for arbitrary distributions of agents, there exists a matching scheme which satisfies (a) within-group positive sorting between cognitive skills and (b) within-group positive sorting between social skills. Furthermore, I show that, when the matching output exhibits strictly positive complementarities (\clubsuit) between cognitive skills and (\spadesuit) between social skills, every optimal matching distribution satisfies (a) within-group positive sorting between cognitive skills and (b) within-group positive sorting between social skills. I establish the optimality of within-group sorting for a general within-group sorting class in Proposition 1.

Within-group sorting has two major drawbacks as a sorting concept. First of all, there may be multiple ways to match agents without violating within-group sorting.

Firms	(10, 10)	(10, 20)	(20, 10)	(20, 20)
Matched Worker	(20, 20)	(10, 20)	(20, 10)	(10, 10)
Matching scheme-1				

Firms	(10, 10)	(10, 20)	(20, 10)	(20, 20)
Matched Worker	(10, 10)	(10, 20)	(20, 10)	(20, 20)
Matching scheme-2				

Note that for each social skill combination of firms and workers, there is only one firm-worker couple under these two matching schemes. The same is also true for each cognitive skill combination of firms and workers. Consequently, these matching schemes satisfy within-group positive sorting between cognitive skills and

social skills.

Secondly, a matching scheme may satisfy within-group positive sorting and cannot be optimal for any matching output that exhibits strictly positive complementarities between cognitive skills and between social skills. For matching scheme-1, a swap between the first and the last firm-worker couples, i.e. $((10, 10), (20, 20))$ and $((20, 20), (10, 10))$, strictly increases the aggregate output for any matching output which exhibits strictly positive complementarities between cognitive skills and between social skills. Therefore, matching scheme-1 can never be an optimal matching scheme when the matching output exhibits strictly positive complementarities between cognitive skills and between social skills.

To obtain a finer characterization of optimal matching schemes, I consider another extension of Becker's sorting concepts and propose a weak sorting notion: a matching scheme satisfies *weak positive sorting* if there does not exist a pair of matched couples that (a) is consistent with global negative sorting, and (b) violates global positive sorting. Notice that the first and the last firm-worker couples in matching scheme-1 violate global positive sorting. Indeed, the set of matching schemes that satisfy weak positive sorting is a subset of the set of matching schemes that satisfy within-group positive sorting. More importantly, I show that the set of optimal matching schemes for any matching output that exhibits strictly positive complementarities between cognitive skills and between social skills is a subset of the set of matching schemes that satisfies weak positive sorting. I present a general version of this result in Proposition 1.

Similar to within-group sorting, a matching scheme may satisfy weak positive sorting and can never be optimal when the matching output exhibits strictly positive complementarities between cognitive skills and between social skills. Note that matching scheme-3 below satisfies weak positive sorting. At the same time, the following swap sequence strictly increases the aggregate output for any matching output that exhibits strictly positive complementarities between cognitive skills and between social skills:

Swap-1: Between the first and the second couples

Swap-2: Between the third and the fourth couples

Swap-3: Between the first and the last couples matched after swap-1 and swap-2

Firms	(10, 10)	(10, 20)	(20, 10)	(20, 20)
Matched Worker	(10, 20)	(20, 20)	(10, 10)	(20, 10)

Matching scheme-3

Firms	(10, 10)	(10, 20)	(20, 10)	(20, 20)
Matched Worker	(20, 20)	(10, 20)	(20, 10)	(10, 10)

Swap-2

Firms	(10, 10)	(10, 20)	(20, 10)	(20, 20)
Matched Worker	(20, 20)	(10, 20)	(10, 10)	(20, 10)

Swap-1

Firms	(10, 10)	(10, 20)	(20, 10)	(20, 20)
Matched Worker	(10, 10)	(10, 20)	(20, 10)	(20, 20)

Swap-3

These examples demonstrate that global, within-group and weak sorting concepts are not adequate to obtain a fine characterization of the set of optimal matching distributions. In Section 1, I lay out the statistical logic behind Becker’s sorting result and restate it by using the upper-set properties of the supermodular order. By devising multidimensional generalizations of supermodularity, supermodular order, and concordance, I characterize Pareto improving swaps for a large set of complementarity structures that allow for negative complementarities between some skills along with positive complementarities between some other skills.

Although these results can be applied to wide-ranging matching markets, they lack predictive power: the set of optimal matching distributions may not be a singleton. Lindenlaub (2017) offers a multidimensional sorting theory with higher predictive power. She adopts three key assumptions: (a) the agents on each side of the market have the same number of outcome-relevant attributes, (b) each attribute complements one and only one attribute on the other side of the market, and (c) the matching output exhibits either strictly positive complementarity in all attributes or strictly negative complementarity in all attributes. More specifically, she considers matching output functions that have the following form: $Q(x, y) = Q_c(x_c, y_c) + Q_s(x_s, y_s)$ where the cross-partial derivative of Q_i is either strictly positive for all $i \in \{c, s\}$ or strictly negative for all $i \in \{c, s\}$.

For this framework, she shows that the optimal matching is unique. In addition, she proves that the optimal matching is a smooth function under additional restrictions: the agents are distributed with infinitely many times continuously differentiable probability distribution functions; Q_i is four times continuously differentiable; and Q_{x_i, y_i} is supermodular and log-supermodular. Under these additional restrictions, she establishes that $\partial y_i^* / \partial x_i > 0$ for all $i \in \{c, s\}$, where $y^* := m^*(x)$ denotes the matched worker of type- x firm under optimal matching function $m^* : \mathbb{R}^2 \rightarrow \mathbb{R}^2$.

However, these assumptions on the output function may not be suitable for many applications. As an

example, consider the following matching output: $Q(x, y) = \alpha x_c y_c + \beta x_s y_s + \gamma x_s y_c$. The results presented by Lindenlaub (2017) apply only if $\alpha\beta > 0$ and $\gamma = 0$. In this case, once one assumes strictly positive complementarity between cognitive skills, one must also assume strictly positive complementarity between social skills, and vice versa ($\alpha\beta > 0$). Also, each attribute of one side can interact with one and only one attribute of the other side of the market ($\gamma = 0$). Dupuy and Galichon (2014, Table 3) provide empirical evidence suggesting that (a) positive and negative complementarities between different attributes exist and (b) some attributes simultaneously complement multiple attributes in the Dutch marriage market.

One other drawback of the sorting theorems presented in Proposition 1 is that they predict extreme sorting patterns as optimal. However, observing extreme sorting in practice is improbable. A potential reason³ is that we do not observe every outcome-relevant attribute. For the previous example, assume that (a) firms and workers observe each other's cognitive and social skills, and (b) econometricians observe firms' and workers' cognitive skills but do not observe their social skills. In this case, econometricians observe a sorting pattern between firms' and their workers' cognitive skills aggregated over social skills. The aggregated matching pattern between cognitive skills may exhibit mismatch, i.e. deviations from extreme sorting. The framework presented in Section 1 allows me to explore this source of mismatch. Consider four types of firms (x) and four types of workers (y): $\{(10, 10), (10, 20), (20, 10), (20, 20)\}$. Assume that for skill vectors $(10, 10)$ and $(20, 20)$, there are four firms. In addition, suppose that there is one firm for each skill vector $(10, 20)$ and $(20, 10)$. Similarly, let there be four workers for each skill vector $(10, 20)$ and $(20, 10)$; and one worker for each skill vector $(10, 10)$ and $(20, 20)$. The unique optimal matching scheme between these firms and workers is given below for the following output function: $Q(x, y) = x_c y_c + 2x_s y_s$.

		Workers			
		(10, 10)	(10, 20)	(20, 10)	(20, 20)
Firms	(10, 10)	1	0	3	0
	(10, 20)	0	1	0	0
	(20, 10)	0	0	1	0
	(20, 20)	0	3	0	1

Optimal matching counts between firms and workers

		Workers	
		10	20
Firms	10	2	3
	20	3	2

Aggregated matching counts
by cognitive skills

In line with Proposition 1, the optimal matching scheme (left panel) satisfies weak positive sorting. However, the aggregated matching between firms' and their workers' cognitive skills (right panel) exhibits neither

³Search and informational frictions may also cause deviations from extreme sorting.

positive nor negative sorting.

In Section 2, I examine a framework introduced by Choo and Siow (2006) that allows agents to have unobserved characteristics that are outcome-relevant. In this context, obtaining a tractable aggregation over unobserved characteristics requires additional assumptions on the distributions of the unobserved characteristics. In addition, complementarities between unobserved characteristics cannot be allowed as they affect the aggregated match between observed attributes in a way that cannot be controlled by observed attributes. Under these additional restrictions, I obtain empirically robust sorting results that (a) suggest milder sorting patterns between observed attributes and (b) allow us to infer the underlying complementarities between the observed attributes given empirical matching distribution.

In this context, the changes in complementarities affect the matching outcome in a sophisticated way. In Section 2, I propose a non-parametric notion of increasing complementarities and establish comparative static results regarding the changes in complementarities without making assumptions on either the matching output function or the distributions of the observable attributes.

To demonstrate an application of the empirically robust sorting results, I examine the association between agents' health status and their spouses' education levels among U.S. households in Section 3 by using the IPUMS-CPS data series for 2010-2017. In literature, many studies report a positive association between agents' health status and their spouses' education levels. From the actuarial point of view, decomposing this association is essential. If one's spouse's education level is a strong predictor of one's health status, then insurance companies can reduce the risk that they carry by taking one's spouse's education level into account. I show that one's spouse's education level is not a strong predictor of one's health status and identify an attraction channel which explains the positive association between agents' health status and their spouses' education levels. It is estimated that the association is a product of three factors: (a) an attraction between better-educated individuals, (b) an attraction between healthier individuals, and (c) a positive association between agents' health status and their own education levels. This decomposition implies that the insurers' risk associated with a two-person family plan is higher than the aggregate risk associated with two individual policies.

Empirical Support for Multidimensional Types

In this section, I survey the recent empirical studies that support the presence of multiple outcome-relevant attributes in the healthcare, labor, and marriage/dating markets.

In the healthcare market, the relationship between doctors and patients is known to be multidimensional. Jagosh et al. (2011) show that effective communication enhances patient recovery. The authors argue that three listening skills of health professionals (listening as an essential component of clinical data gathering and diagnosis; listening as a healing and therapeutic agent; and listening as a means of fostering and strengthening the doctor–patient relationship) are central to successful clinical outcomes. Stavropoulou (2011) indicates that six aspects of the relationship between doctors and patients affect nonadherence to medication using the European Social Survey. Nonadherence to medication is also proven to be a complex and multidimensional healthcare problem by Hugtenburg et al. (2013). Similarly, Mazzi et al. (2018) identify four attributes of doctors and three characteristics of patients that are essential to successful clinical outcomes by using an integrated survey of thirty-one European countries. Belasen and Belasen (2018) provide evidence suggesting that different aspects of the relationships between doctors and patients affect not only the clinical outcome but also patients’ rankings of hospitals.

The empirical findings in labor economics literature also suggest that workers and firms have multidimensional types. Deming (2017) shows that workers’ cognitive and social skills are important determinants of wages in the U.S. labor market. Girsberger et al. (2018) add manual skill to that list for the Swiss labor market by using data from the Social Protection and Labour Market (SESAM) panel. Guvenen et al. (2018) analyze the skill mismatch between workers and firms in the U.S. labor market. Their analysis indicates that verbal and math skills have statistically significant effects on workers’ wages.

Hitsch et al. (2010) study mating behavior in the U.S. dating market by using a large dataset provided by an online dating website. They show that the differences in age, educational attainment, and body mass index decrease the probability of dating. Belot and Francesconi (2012) confirm these findings by studying the speed dating patterns of individuals based on a British dataset. They find that physical attractiveness factors (age, height, and body mass index) play an essential role in the earlier stages of the relationship. Klostad et al. (2013) add political views to that list. By analyzing a large dataset provided by another online dating website, the authors show that couples tend to share the same political preferences.

Gemici and Laufer (2010) study cohabitation, marriage and separation patterns in the U.S. mating market. They report that age, educational attainment, and race are key variables in explaining agents’ choices. Dupuy and Galichon (2014) add other important variables to that list by studying the Dutch marriage market. They show that personality traits, such as emotional stability and conscientiousness, are also important determinants of the Dutch household formation. Domingue et al. (2014) analyze the genetic similarities between married couples by using the Health and Retirement Study and information from 1.7

million single-nucleotide polymorphisms. Their results demonstrate that similar genetic types attract each other. They also find that educational similarities between spouses are stronger than genetic similarities. The thorough examination of the household formation allows researchers to explain changes in the household income inequality. For example, Greenwood et al. (2014) argue that similarities between spouses' education levels contributed to the increasing household income inequality in the U.S. between 1960 and 2005.

1 The Matching Model with Multidimensional Types

In this section, I study a general model of two-sided matching markets. The only difference between the framework examined by Becker (1973) and the one presented in this section is that I allow the agents to have multiple outcome-relevant attributes while Becker (1973) assumes that each agent has only one outcome-relevant attribute. I start by outlining the general framework in detail.

Agents: There are two sides of the matching market, namely firms and workers. A generic firm is denoted by x , and a generic worker is denoted by y . Every firm has K productive attributes, i.e. $x \in \mathbb{R}^K$, and each worker has L productive attributes, i.e. $y \in \mathbb{R}^L$. It is assumed that the overall measures of firms and workers coincide. The firms and workers are distributed according to cumulative distribution functions $F : \mathbb{R}^K \rightarrow [0, 1]$ and $G : \mathbb{R}^L \rightarrow [0, 1]$, respectively.

Matching Distribution: Matching distribution $M : \mathbb{R}^K \times \mathbb{R}^L \rightarrow [0, 1]$ is a cumulative distribution function associated with a particular matching scheme. More specifically, $M(x, y)$ represents the fraction of matched firm-worker couples with attributes less than or equal to (x, y) .

Given F and G , matching distribution M satisfies no-single property if and only if

$$(a) \lim_{y \rightarrow \infty^L} M(x, y) = F(x) \text{ for all } x \in \mathbb{R}^K, \text{ and } (b) \lim_{x \rightarrow \infty^K} M(x, y) = G(y) \text{ for all } y \in \mathbb{R}^L.$$

Let $\mathcal{M}(F, G)$ denote the set of matching distributions that satisfy no-single property given F and G .

Output Function: A match between a firm and a worker with attributes x and y generates a matching output. The matching output is determined by exogenously specified output function $Q : \mathbb{R}^K \times \mathbb{R}^L \rightarrow \mathbb{R}_{++}$, and denoted by $Q(x, y)$. For any unmatched agent, the output is assumed to be 0.

Planner's Problem: Given F , G , and Q , the social planner maximizes the aggregate output by choosing a matching distribution that satisfies no-single property:

$$\max_{M \in \mathcal{M}(F, G)} \int Q dM.$$

Two key concepts of the matching theory are complementarity and sorting. In this context, *positive(negative) complementarity* between a firm's i^{th} attribute and its worker's j^{th} attribute means that

the marginal product of the firm's i^{th} attribute is increasing(decreasing) in its worker's j^{th} attribute. In unidimensional matching literature, complementarities are formulated by using supermodularity. Function $Q : \mathbb{R}^2 \rightarrow \mathbb{R}$ is called *(strictly) supermodular* if for all $x' > x$ and $y' > y$, it holds that

$$\{Q(x', y') - Q(x, y')\} - \{Q(x', y) - Q(x, y)\} \geq (>) 0.$$

Similarly, Q is *(strictly) submodular* if $-Q$ is (strictly) supermodular; and Q satisfies *modularity* if Q is both supermodular and submodular. In the multidimensional setting, supermodularity can also be used with a slight modification to formulate multidimensional complementarities. Towards this end, I introduce *i,j pairwise supermodularity*.

Definition 1. Function $Q : \mathbb{R}^K \times \mathbb{R}^L \rightarrow \mathbb{R}_{++}$ is

- (a) (strictly) i,j pairwise supermodular if $Q(x_i, x_{-i}, y_j, y_{-j})$ is a (strictly) supermodular function of (x_i, y_j) for all $(x_{-i}, y_{-j}) \in \mathbb{R}^{K-1} \times \mathbb{R}^{L-1}$;
- (b) (strictly) i,j pairwise submodular if $-Q$ is (strictly) i,j pairwise supermodular; and
- (c) i,j pairwise modular if Q is both i,j pairwise supermodular and i,j pairwise submodular.

By construction, the pairwise modularity concepts can be used to formulate the relationships between the marginal product of a firm's i^{th} attribute and its worker's j^{th} attribute. This aspect is easy to observe when the output function is smooth: $Q : \mathbb{R}^K \times \mathbb{R}^L \rightarrow \mathbb{R}$ satisfies i,j pairwise supermodularity if and only if its cross-partial derivative with respect to x_i and y_j is positive. Through the use of pairwise modularity concepts, I define two multidimensional complementarity classes. Consider two disjoint subsets of $\{1, \dots, K\} \times \{1, \dots, L\}$: P and N .

Definition 2. Function $Q : \mathbb{R}^K \times \mathbb{R}^L \rightarrow \mathbb{R}_{++}$ exhibits (strict) P,N modularity if and only if $Q(x, y)$ is

- (a) (strictly) i,j pairwise supermodular for all $(i, j) \in P$;
- (b) (strictly) p,q pairwise submodular for all $(p, q) \in N$; and
- (c) m,n pairwise modular for all $(m, n) \notin P \cup N$.

Let $\mathbb{C}(P, N)$ and $\mathbb{C}_+(P, N)$ denote the sets of functions which satisfy P,N modularity and strict P,N modularity for set parameters P and N .

Here P represents the set of firm-worker attribute pairs for which the output function exhibits positive complementarity. Similarly, N represents the set of firm-worker attribute pairs for which the output function exhibits negative complementarity. The strongest element of these complementarity classes is that they do not allow for complementarity between a firm's i^{th} attribute and its worker's j^{th} attribute to change

signs (from positive to negative). For example, an output function which exhibits (a) strictly positive complementarity between a firm's and its worker's cognitive skills for some levels of its worker's social skill, and (b) strictly negative complementarity between the firm's and its worker's cognitive skills for some levels of its worker's social skill cannot be formulated by using a P,N modular function. Despite this shortcoming, P,N modular functions can capture various output function forms which have been frequently used in the matching literature.

Table 1: Some examples of P,N modular functions frequently used in matching literature

Output function	Strict P,N modularity	Reference
$Q(x, y) = x' Ay = \sum_{i=1}^K \sum_{j=1}^L a_{ij} x_i y_j$	$P = \{(i, j) : a_{ij} > 0\}$ and $N = \{(i, j) : a_{ij} < 0\}$	Dupuy and Galichon (2014)
$Q(x, y) = \sum_{i=1}^K Q_i(x_i, y_i)$	$P = \{1, \dots, K\}$ and $N = \emptyset$ if $\partial^2 Q_i(x_i, y_i) / \partial x_i \partial y_i > 0$	Lindenlaub (2017)

Due to the simplicity of their interpretation and their frequent use, understanding the optimal sorting patterns for P,N modular output functions is of theoretical and empirical interest. To achieve this goal, I introduce three multidimensional sorting patterns. The first sorting pattern is a direct extension of the global sorting pattern presented by Becker (1973).

Definition 3. Matching distribution $M : \mathbb{R}^K \times \mathbb{R}^L \rightarrow [0, 1]$ satisfies

(a) positive sorting between firms' i^{th} and their workers' j^{th} attributes if and only if

$$(x'_i - x_i)(y'_j - y_j) \geq 0 \text{ for all } (x, y), (x', y') \in \text{supp}(M), \text{ and}$$

(b) negative sorting between firms' i^{th} and their worker's j^{th} attributes if and only if

$$(x'_i - x_i)(y'_j - y_j) \leq 0 \text{ for all } (x, y), (x', y') \in \text{supp}(M).$$

Similar to P,N modularity, I combine pairwise sorting patterns to construct a multidimensional sorting class.

Definition 4 (Global P,N sorting). Matching distribution $M : \mathbb{R}^K \times \mathbb{R}^L \rightarrow [0, 1]$ satisfies global P,N sorting if and only if the following conditions hold for all $(x, y), (x', y') \in \text{supp}(M)$,

(a) $(x'_i - x_i)(y'_j - y_j) \geq 0$ for all $(i, j) \in P$, and

(b) $(x'_p - x_p)(y'_q - y_q) \leq 0$ for all $(p, q) \in N$.

In other words, a matching distribution exhibits global P,N sorting if and only if it exhibits (a) positive sorting between firms' i^{th} and their workers' j^{th} attributes for all $(i, j) \in P$ and (b) negative sorting between firms' p^{th} and their workers' q^{th} attributes for all $(p, q) \in N$.

Although global P,N sorting is a clear sorting pattern between two multidimensional agents, it requires simultaneous sorting between different attributes. Consequently, its existence is tied to the distributions of the agents.

Example 1. Consider a labor market with equal numbers of two types of firms: $(10, 10)$ and $(20, 20)$; and equal numbers of two types of workers: $(10, 20)$ and $(20, 10)$. Notice that matching a $(10, 20)$ worker with a $(10, 10)$ firm and a $(20, 10)$ worker with a $(20, 20)$ firm violates positive sorting between the second attributes. Similarly, the swap between these two couples, i.e. matching a $(10, 20)$ worker with a $(20, 20)$ firm and a $(20, 10)$ worker with a $(10, 10)$ firm, violates positive sorting between the first attributes. Therefore, it is not possible to observe simultaneous positive sorting between the first attributes and the second attributes without inefficiently leaving some agents unmatched.

Due to the existence issue, I study two alternative sorting patterns that exist for arbitrary F, G, P , and N . The next sorting pattern is inspired by Chiappori et al. (2017). The authors analyze a matching model in which each side of the market is represented by one continuous variable (socioeconomic status) and one binary variable (smoking habit). They categorize couples into two main groups. In the first group, both men and women are non-smokers. In the second group, at least one spouse is a smoker. They assume that the matching output of a couple is the multiplication of spouses' socioeconomic status. If there is a smoker in the household, then the output is scaled down by a constant. Under this complementarity structure, they predict positive sorting between agents' socioeconomic status *within* each group. Although this sorting result immediately follows from Becker's (1973) unidimensional sorting theory, the idea of within-group sorting can be deployed in a more general setting.

Definition 5. Matching distribution $M : \mathbb{R}^K \times \mathbb{R}^L \rightarrow [0, 1]$ satisfies

(a) within-group positive sorting between firms' i^{th} and their workers' j^{th} attributes if and only if for all $(x, y), (x', y') \in \text{supp}(M)$ such that $(\clubsuit) x_k = x'_k$ for all $k \neq i$ and $(\spadesuit) y_l = y'_l$ for all $l \neq j$, it holds that

$$(x'_i - x_i)(y'_j - y_j) \geq 0; \text{ and}$$

(b) within-group negative sorting between firms' i^{th} and their workers' j^{th} attributes if and only if for all $(x, y), (x', y') \in \text{supp}(M)$ such that $(\clubsuit) x_k = x'_k$ for all $k \neq i$ and $(\spadesuit) y_l = y'_l$ for all $l \neq j$, it holds that

$$(x'_i - x_i)(y'_j - y_j) \leq 0.$$

Similar to P,N sorting, I combine within-group sorting patterns to define a within-group sorting class.

Definition 6 (Within-group P,N Sorting). Matching distribution $M : \mathbb{R}^K \times \mathbb{R}^L \rightarrow [0, 1]$ satisfies within-group P,N sorting if and only if it satisfies

- (a) within-group positive sorting between firms' i^{th} and their workers' j^{th} attributes for all $(i, j) \in P$; and
- (b) within-group negative sorting between firms' p^{th} and their workers' q^{th} attributes for all $(p, q) \in N$.

Within-group P,N sorting is not a very informative sorting pattern as some of the nonoptimal matching distributions may satisfy within-group P,N sorting.

Example 2. Consider a matching market with equal numbers of two types of workers: $(10, 20)$ and $(20, 10)$; and equal numbers of two types of firms: $(10, 20)$ and $(20, 10)$. Notice that matching $(10, 20)$ workers with $(20, 10)$ firms, and $(20, 10)$ workers with $(10, 20)$ firms, satisfies within-group P,N sorting for $P = \{(1, 1), (2, 2)\}$ and $N = \{\}$. However, it is clear that a swap of the partners between firm-worker couples $\{(10, 20), (20, 10)\}$ and $\{(20, 10), (10, 20)\}$ strictly improves the aggregate output for any strictly P,N modular output function. For example, matching same types with each other is associated with strictly higher aggregate output for output function $Q(x, y) = x_1y_1 + x_2y_2$:

$$100 + 400 + 400 + 100 = 1000 > 800 = 200 + 200 + 200 + 200.$$

The right-hand side of the equation above equals the total output produced by firm-worker couples $\{(10, 20), (20, 10)\}$ and $\{(20, 10), (10, 20)\}$; and the left-hand side of the equation is the total output produced by firm-worker couples $\{(10, 20), (10, 20)\}$ and $\{(20, 10), (20, 10)\}$.

As it is demonstrated in Example 2, a fine characterization of the set of optimal matching distributions cannot be obtained by using within-group P,N sorting. As such, I adopt a more constructive approach to obtain a fine description of the set of optimal matching distributions. To motivate the idea behind this approach, I first lay out a statistical course to obtain Becker's (1973) sorting results.

Theorem 1 (Becker's (1973) sorting results). *In a unidimensional matching market, i.e. $K = L = 1$,*

- (a) *positive sorting is an optimal matching distribution when the output function is supermodular;*
- (b) *negative sorting is an optimal matching distribution when the output function is submodular;*
- (c) *positive sorting is the unique optimal matching distribution when the output function is strictly supermodular; and*
- (d) *negative sorting is the unique optimal matching distribution when the output function is strictly submodular.*

It is easy to establish these sorting results by using the upper-set properties in the supermodular order. Distribution $M : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$ dominates $M' : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$ in the *supermodular order* if and only if $\int Q dM \geq \int Q dM'$ for all supermodular $Q : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$. Muller and Scarsini (2010), and Meyer and Strulovici (2013) offer an alternative characterization of the supermodular order.

Definition 7.

- (a) A pair of couples $(x, y), (x', y') \in \mathbb{R}^2$ is (weakly) concordant if and only if $(x - x')(y - y') > (\geq) 0$.
- (b) A pair of couples $(x, y), (x', y') \in \mathbb{R}^2$ is (weakly) discordant if and only if $(x - x')(y - y') < (\leq) 0$.
- (c) A concordance improving transfer is a uniform probability transfer from a discordant bivariate pair to the concordant bivariate pair that is obtained via the swap between the discordant bivariate pair. Let $\tau(x, y, x', y'; \alpha)$ denote the concordance improving transfer with $\alpha \geq 0$ weight which increases densities by α at (x, y) and (x', y') such that $(x - x')(y - y') > 0$; and decreases densities at (x, y') and (x', y) by α .

Theorem 2 (Muller and Scarsini (2010), and Meyer and Strulovici (2013)). $M : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$ dominates $M' : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$ in the *supermodular order* if and only if M can be obtained from M' via concordance improving transfers.

Based on this alternative characterization, it is easy to see that positive sorting is the *dominant* matching distribution in the supermodular order. Similarly, negative sorting is strictly *dominated* by any other matching distribution in the supermodular order. Consequently, Becker's (1973) sorting result is obtained. Following a similar logic, a fine characterization of the set of optimal matching distributions for the multi-dimensional setting can be established by identifying the changes in the matching distribution that increase the aggregate output when the matching output is P,N modular. On this note, I define P,N modular order in Definition 8.

Definition 8. Distribution $M : \mathbb{R}^K \times \mathbb{R}^L \rightarrow [0, 1]$ dominates $M' : \mathbb{R}^K \times \mathbb{R}^L \rightarrow [0, 1]$ in P,N modular order, denoted by $M \succeq_{P,N} M'$, if and only if $\int Q dM \geq \int Q dM'$ for all $Q \in \mathbb{C}(P, N)$. Distribution M strictly dominates M' in P,N modular order if and only if (a) M dominates M' in P,N modular order; and (b) M is not dominated by M' in P,N modular order.

Next, I formally define the sets of P,N dominant and P,N undominated distributions that are essential to characterizing the set of optimal matching distributions.

Definition 9. Matching distribution $M \in \mathcal{M}(F, G)$ is P,N dominant if it dominates every $M' \in \mathcal{M}(F, G)$ in P,N modular order. Similarly, $M \in \mathcal{M}(F, G)$ is P,N undominated if there does not exist $M' \in \mathcal{M}(F, G)$ which strictly dominates M in P,N modular order.

The next theorem restates Becker's (1973) sorting result by using multidimensional concepts. By doing so, it makes the theoretical differences between unidimensional and multidimensional matching markets clear.

Theorem 3 (Unidimensional Sorting - Multidimensional Concepts). *Let $K = L = 1$ and $P \cup N = \{(1, 1)\}$.*

- a) For any P and N , a matching distribution is P, N dominant if and only if it is P, N undominated.*
- b) For any P and N , there is only one P, N dominant distribution.*
- c) For any $Q \in \mathbb{C}(P, N)$, the P, N dominant distribution solves the planner's problem.*
- d) For any $Q \in \mathbb{C}_+(P, N)$, the P, N dominant distribution is the unique solution to the planner's problem.*
- e) The P, N dominant matching distribution is positive assortative matching when $P = \{(1, 1)\}$ and $N = \{\}$:*

$$\Lambda(x, y) = \min \{F(x), G(y)\}.$$

- f) The P, N dominant matching distribution is negative assortative matching when $P = \{\}$ and $N = \{(1, 1)\}$:*

$$\Omega(x, y) = \max \{F(x) + G(y) - 1, 0\}.$$

There are two key aspects of optimality in matching markets with unidimensional agents. First of all, the set of P, N undominated distributions is singleton when $P \cup N \neq \emptyset$. Secondly, there exists a P, N dominant distribution. These two features make it possible to fully characterize the solution to the planner's problem for strictly P, N modular output functions in unidimensional case. However, they do not apply to the multidimensional setting.

Example 3. Consider a matching market with equal numbers of two types of workers: $(10, 20)$ and $(20, 10)$, equal numbers of two types of firms: $(10, 10)$ and $(20, 20)$, and a class of matching output functions with parameter $\gamma \in [0, 1]$: $Q(x, y; \gamma) = \gamma x_1 y_1 + (1 - \gamma) x_2 y_2$. Note that for all values of γ , the output function exhibits P, N modularity for $P = \{(1, 1), (2, 2)\}$ and $N = \{\}$.

The set of P, N undominated distributions is not singleton: Consider matching distribution M under which every $(10, 20)$ worker is matched with a $(10, 10)$ firm; and every $(20, 10)$ worker is matched with a $(20, 20)$ firm. It immediately follows from Becker's sorting result that M is P, N undominated since it is the unique solution to the planner's problem when $\gamma = 1$. Alternative matching distribution M' under which every $(20, 10)$ worker is matched with a $(10, 10)$ firm, and every $(10, 20)$ worker is matched with a $(20, 20)$ firm is also P, N undominated since it is the unique solution to the planner's problem when $\gamma = 0$.

The set of P, N dominant distributions is empty: Consider a matching distribution under which some $(10, 20)$ workers are matched with $(10, 10)$ firms; and some $(20, 10)$ workers are matched with $(20, 20)$ firms. This matching distribution cannot be P, N dominant as M' is associated with strictly higher aggregate output

when $\gamma = 0$. Similarly, a matching distribution under which some $(20, 10)$ workers are matched with $(10, 10)$ firms, and some $(10, 20)$ workers are matched with $(20, 20)$ firms cannot be P,N dominant as M is associated with strictly higher aggregate output when $\gamma = 1$.

Due to the fact that the set of P,N dominant distributions is neither singleton nor non-empty for arbitrary model parameters, characterization of the optimal matching distributions is a non-trivial task. I obtain a fine description of the optimal matching distributions, by characterizing Pareto improving swaps in Lemma 1 below.

Definition 10. A pair of firm-worker couples $(x, y), (x', y') \in \mathbb{R}^K \times \mathbb{R}^L$ is P,N weak concordant if

(a) $(x_i - x'_i)(y_j - y'_j) \geq 0$ for all $(i, j) \in P$; and

(b) $(x_p - x'_p)(y_q - y'_q) \leq 0$ for all $(p, q) \in N$.

A P,N concordant pair is a P,N weak concordant pair such that (\clubsuit) some of the inequalities in (a) hold with strict inequality for some $(i, j) \in P$; or (\spadesuit) some of the inequalities in (b) hold with strict inequality for some $(p, q) \in N$.

Definition 11. A P,N concordance improving transfer is a uniform probability transfer from an N,P weak concordant pair of couples to the P,N weak concordant pair of couples that is obtained via the swap between the N,P weak concordant pair of couples.

Lemma 1. *Distribution $M : \mathbb{R}^K \times \mathbb{R}^L \rightarrow [0, 1]$ dominates $M' : \mathbb{R}^K \times \mathbb{R}^L \rightarrow [0, 1]$ in P,N modular order if and only if M can be obtained from M' via a sequence of P,N concordance improving transfers.*

This alternative characterization of dominance in P,N modular order allows me to obtain a finer description of the set of optimal matching distributions. Consider matching distribution $M \in \mathcal{M}(F, G)$ under which there exists an N,P concordant pair of matched couples. Due to Lemma 1, it is easy to see that a swap between the N,P concordant pair of couples (strictly) increases the aggregate output when the output function is (strictly) P,N modular. Consequently, a matching distribution under which there exists an N,P concordant pair of matched couples cannot be obtained as optimal for strictly P,N modular output functions. This observation rules out the nonoptimal matching distribution illustrated in Example 2, and allows me to define a new sorting pattern.

Definition 12. Matching distribution $M \in \mathcal{M}(F, G)$ satisfies weak P,N sorting if there does not exist an N,P concordant pair of couples with a positive mass under M .

The next proposition establishes the link between the proposed sorting patterns and the set of optimal matching distributions for the proposed complementarity structures.

Proposition 1 (Multidimensional Sorting).

1. For arbitrary distributions of the agents F and G , and two disjoint sets of firm-worker attribute pairs P and N ,
 - 1.a) every weak P, N assortative matching distribution satisfies within-group P, N sorting;
 - 1.b) for every $Q \in \mathbb{C}_+(P, N)$, every solution to the planner's problem satisfies weak P, N sorting; and
 - 1.c) for every $Q \in \mathbb{C}(P, N)$, there exists a weak P, N assortative matching distribution that solves the planner's problem.
2. Suppose that there exists $M \in \mathcal{M}(F, G)$ that satisfies global P, N sorting. Then,
 - 2.a) for every $Q \in \mathbb{C}_+(P, N)$, every solution to the planner's problem satisfies global P, N sorting; and
 - 2.b) for every $Q \in \mathbb{C}(P, N)$, M solves the planner's problem.

Proposition 1 summarizes the relationship between the proposed sorting patterns and multidimensional complementarities. First, it states that within-group P, N sorting is the least informative sorting pattern among all. Secondly, it offers a partial identification of the set of optimal matching distributions: (i) the set of optimal matching distributions and the set of weak P, N assortative matching distributions intersect for P, N modular output functions; and (ii) the set of optimal matching distributions is a subset of the set of weak P, N assortative matching distributions for strictly P, N modular output functions. On the other hand, a weak P, N assortative matching distribution cannot be optimal for any strictly P, N modular output function when it is strictly dominated by another matching distribution. This observation immediately follows from the fact that the absence of N, P concordant pairs is necessary but not sufficient for undominance in the P, N modular order. Altogether, Proposition 1 suggests that P, N sorting is a suitable general sorting class to study the multidimensional matching markets.

Although it provides new insights into optimal matching with multidimensional agents, this sorting result is not very practical for empirical purposes. Proposition 1 states that one cannot observe N, P concordant pairs under strictly P, N modular functions when every outcome-relevant attribute is observed by econometricians. In practice, econometricians only observe sorting patterns between observed attributes that are aggregated over unobserved characteristics. Therefore, linking production complementarities between observed attributes to matching patterns between these attributes when outcome-relevant and unobserved characteristics are present is of theoretical and empirical interest. In the next section, I examine a matching model in which agents have outcome-relevant and unobserved characteristics in addition to their observed attributes.

2 Sorting with Unobserved Characteristics

Choo and Siow (2006) propose a matching model with multidimensional agents in which the agents have outcome-relevant characteristics that are not observed by econometricians. In this section, I examine a homoskedastic extension of their model.

Consider a two-sided matching model with equal numbers of firms and workers. Here, firm $f \in \mathbb{F}$ is described by full attribute vector \tilde{x}^f , and \tilde{y}^w describes worker $w \in \mathbb{W}$. Matching scheme $\tilde{\mathbf{m}} = \{\tilde{m}_{fw}\}$ is a matrix such that the cell value associated with firm f and worker w , i.e. \tilde{m}_{fw} , equals one if firm f and worker w are matched, and it equals zero otherwise. Let $x^f \in \mathbb{R}^K$ denote the observable attributes of firm f , and $y^w \in \mathbb{R}^L$ denote the observable attributes of worker w . The matching output produced by firm f and worker w , denoted by $\tilde{Q}(\tilde{x}^f, \tilde{y}^w)$, is determined by firm f 's and worker w 's full attributes. An optimal matching matrix maximizes the aggregate output.

Notice that an optimal matching matrix is determined by agents' full attributes. However, econometricians observe only observable attributes. Consequently, the empirical goal is to estimate the complementarities between observable attributes by using the optimal matching density function between observable attributes implied by an optimal matching matrix. Optimal matching density function $m : \mathbb{R}^K \times \mathbb{R}^L \rightarrow [0, 1]$, associated with matching matrix $\tilde{\mathbf{m}} = \{\tilde{m}_{fw}\}$, is defined through aggregation over unobserved characteristics:

$$m(x, y) = \frac{\sum_f \sum_w \tilde{m}_{fw} 1_{\{x^f = x \text{ and } y^w = y\}}}{\sum_f \sum_w \tilde{m}_{fw}}.$$

Notice that if the effect of unobserved characteristics on the optimal matching between observables cannot be controlled by observed attributes, then one cannot consistently estimate the preferences on observed attributes. Example 4 demonstrates two channels through which the unobserved characteristics affect optimal matching between observables, in a way that cannot be explained by observed attributes when the distributions of unobserved characteristics conditional on observed attributes are unknown.

Example 4. Consider a matching market with two firms: $\{(10, 10), (20, 20)\}$, and two workers: $\{(20, 10), (10, 20)\}$. Suppose that econometricians do not observe the first characteristics, and the second attributes are observable.

Complementarity between unobserved characteristics: $\tilde{Q}(\tilde{x}^f, \tilde{y}^w) = 5\tilde{x}_1^f \tilde{y}_1^w + 2\tilde{x}_2^f \tilde{y}_2^w$

Complementarity between unobserved and observed attributes: $\tilde{Q}(\tilde{x}^f, \tilde{y}^w) = 2\tilde{x}_2^f \tilde{y}_1^w + \tilde{x}_2^f \tilde{y}_2^w$

For these two matching output functions, the following optimal matching matrix and density are obtained.

		Worker	
		(20, 10)	(10, 20)
Firm	(10, 10)	0	1
	(20, 20)	1	0

Optimal matching matrix

		Worker	
		10	20
Firm	10	0	.5
	20	.5	0

Optimal matching density

Note that the optimal matching between firms' and workers' second attributes exhibits negative sorting. Negative sorting is consistent with negative complementarity and cannot be obtained under strictly positive complementarity according to Becker's (1973) sorting results. Based on this observation, econometricians may infer negative complementarity between firms' and workers' second attributes whereas the underlying process \tilde{Q} exhibits positive complementarity between these attributes.

In order to limit the outcome-relevance of unobserved characteristics, three key assumptions are adopted in empirical matching literature.

Assumption 1. *There is a large number of agents for each observed type.*

Under the large market assumption, one can focus on the asymptotic properties of optimal matching. In this context, the small sample properties of optimal sorting patterns remain an open question.

Assumption 2. *Matching output function \tilde{Q} can be expressed as follows:*

$$\tilde{Q}(\tilde{x}^f, \tilde{y}^w) = Q(x^f, y^w) + \varepsilon_f(\tilde{x}^f, y^w) + \eta_w(x^f, \tilde{y}^w).$$

Here, the first part of the matching output is determined by observed attributes. The deterministic matching complementarities are to be estimated by using empirical matching density. However, notice that the matching outcome is not determined solely by the deterministic matching complementarities. The last two parts of the equation represent idiosyncratic production shocks that affect the matching outcome. This separability assumption rules out complementarities between unobserved characteristics. By doing so, it allows for a tractable aggregation over unobserved characteristics.

Remark 1. Consider two firms f and f' , and two workers w and w' such that $x^f = x^{f'} = x$ and $y^w = y^{w'} = y$. For these four agents, Assumption 2 implies that

$$\tilde{Q}(\tilde{x}^f, \tilde{y}^w) + \tilde{Q}(\tilde{x}^{f'}, \tilde{y}^{w'}) = \tilde{Q}(\tilde{x}^{f'}, \tilde{y}^w) + \tilde{Q}(\tilde{x}^f, \tilde{y}^{w'}).$$

Remark 1 states that unobserved aggregate output generated by pairs of couples (f, w) and (f', w') that have the same observed attributes does not change with a partner-swap between these two pairs. Assumptions

1 and 2 make it is possible to link optimal matching between full attributes to the one between observed attributes when the distribution of ε_f conditional on x and the distribution of η_w conditional on y are known.

Assumption 3. *The idiosyncratic production shocks satisfy the following conditions:*

- (a) *for all $f \in \mathbb{F}$ such that $x^f = x$, $\varepsilon^f = \{\varepsilon_f(\tilde{x}^f, y)\}_y$ is drawn from probability distribution \mathcal{F}_x ; and*
- (b) *for all $w \in \mathbb{W}$ such that $y^w = y$, $\eta^w = \{\eta_w(x, \tilde{y}^w)\}_x$ is drawn from probability distribution \mathcal{G}_y .*

Under these assumptions, the full attribute vector of firm f can be represented by (x^f, ε^f) , where probability distribution of $\varepsilon_f(\tilde{x}^f, y)$ conditional on $x^f = x$ is \mathcal{F}_x . Similarly, the full attribute vector of worker w can be represented by (y^w, η^w) , where probability distribution of $\eta_w(x, \tilde{y}^w)$ conditional on $y^w = y$ is \mathcal{G}_y . For example, Choo and Siow (2006) assume that \mathcal{F}_x and \mathcal{G}_y are standard Gumbel distributions, i.e. $Pr\{\varepsilon_f(\tilde{x}^f, y) \leq \varepsilon | x^f = x\} = \exp\{-\exp\{-\varepsilon\}\}$. In this section, I consider a simple homoskedastic extension of this distributional assumption by relaxing Galichon and Salanie's (2010) assumption. Let the distribution of ε_f conditional on $x_w = x$ be a Gumbel distribution with location parameter $\alpha(x)$ and scale parameter σ . Similarly, let the distribution of η_w conditional on $y_w = y$ be a Gumbel distribution with location parameter $\beta(y)$ and scale parameter δ .

$$Pr\{\varepsilon_f(\tilde{x}^f, y) \leq \varepsilon | x^f = x\} = \exp\left\{-\exp\left\{-\left(\frac{\varepsilon - \alpha(x)}{\sigma}\right)\right\}\right\}$$

$$Pr\{\eta_w(x, \tilde{y}^w) \leq \eta | y^w = y\} = \exp\left\{-\exp\left\{-\left(\frac{\eta - \beta(y)}{\delta}\right)\right\}\right\}$$

Galichon and Salanie (2010) establish the uniqueness of optimal matching density. Furthermore, they present a relationship between optimal matching density function and double difference of the deterministic output function, $Q : \mathbb{R}^K \times \mathbb{R}^L \rightarrow \mathbb{R}$, when $\alpha(x) = 0$ and $\beta(y) = 0$ for all x and y . Remember that the difference between two independent random variables following a Gumbel distribution with the same location and scale parameters follows a logistic distribution with zero location parameter. Consequently, Galichon and Salanie's (2010) results hold with type-dependent location parameters as well.

Lemma 2. *The optimal matching density function of observable attributes $m : \mathbb{R}^K \times \mathbb{R}^L \rightarrow [0, 1]$ satisfies the following condition for all $x, x' \in \mathbb{R}^K$ and $y, y' \in \mathbb{R}^L$:*

$$\log\left\{\frac{m(x, y) m(x', y')}{m(x', y) m(x, y')}\right\} = (\sigma + \delta)^{-1} \{Q(x, y) + Q(x', y') - Q(x', y) - Q(x, y')\}. \quad (1)$$

Lemma 2 offers a simple relationship between the complementarity structure and optimal sorting pattern between observable attributes. Based on this relationship, Siow (2015) derives a semi-parametric identification strategy for matching markets in which each agent has only one observable attribute, i.e. $K = L = 1$.

He shows that the deterministic output function is supermodular(submodular) if and only if the left-hand side of Equation (1) is greater(less) than 1 for all $x < x'$ and $y < y'$. I establish a similar result by using P,N modularity.

Definition 13. Matching density function $m' : \mathbb{R}^K \times \mathbb{R}^L \rightarrow [0, 1]$ is log P,N modular if and only if $\log m'$ is P,N modular.

Proposition 2. *Optimal matching density between observable attributes is log P,N modular if and only if the deterministic output function is P,N modular.*

Corollary 1. *The probability of observing P,N concordant pairs relative to N,P concordant pairs is higher when the deterministic output function is P,N modular.*

In the previous section, Proposition 1 states that for strictly P,N modular deterministic output functions, the fraction of N,P concordant pairs equals zero when the scale parameters of the idiosyncratic output components are zero, i.e. every outcome relevant attribute is observed by econometricians. Proposition 2 and Corollary 1 assert that for positive values of scale parameters, positive fraction of N,P concordant pairs may be observed when the deterministic output function is strictly P,N modular. Furthermore, the fraction of P,N concordant pairs are higher than the fraction of N,P concordant pairs when the deterministic output function is P,N modular. Thus, the matching models in which the agents have unobserved characteristics offer milder sorting patterns between observed attributes.

Lemma 2 also allows me to obtain a comparative static result that links the changes in deterministic complementarities to the optimal matching density function. Bojilov and Galichon (2015) present comparative static results regarding the changes in the deterministic complementarities when (a) the deterministic output function is quadratic, i.e. $Q(x, y) = \sum_k \sum_l \theta_{k,l} x_k y_l$; and (b) the observable attributes follow Gaussian distributions. The quadratic functional form assumption offers a simple relationship between complementarities and model parameters. More specifically, complementarity between firms' k^{th} and their workers' l^{th} observable attributes is governed by parameter $\theta_{k,l}$ alone. I offer a similar comparative static result without imposing restrictions on the matching output and the distributions of the observable attributes.

Definition 14. Deterministic output function $Q : \mathbb{R}^K \times \mathbb{R}^L \rightarrow \mathbb{R}$ exhibits higher P,N modularity compared to $Q' : \mathbb{R}^K \times \mathbb{R}^L \rightarrow \mathbb{R}$ if and only if, for all P,N concordant (x, y) and (x', y') , the following condition holds:

$$Q(x, y) + Q(x', y') - Q(x', y) - Q(x, y') \geq Q'(x, y) + Q'(x', y') - Q'(x', y) - Q'(x, y').$$

Here, a P,N modular increase implies (a) an increase in complementarity between firms' i^{th} and their workers' j^{th} attributes for all $(i, j) \in P$, (b) an decrease in complementarity between firms' p^{th} and their

workers' q^{th} attributes for all $(p, q) \in N$. In this context, one can use an uneven P,N modular increase to formulate a skill-biased⁴ complementarity change, non-parametrically.

Proposition 3. *The fraction of P,N concordant pairs relative to the fraction of N,P concordant pairs under optimal matching rises with P,N modular increases in the deterministic output function.*

Proposition 3 strengthens the result presented in Corollary 1. For any deterministic output function (P,N modular or otherwise), a change toward P,N modularity increases the fraction of P,N concordant pairs and decreases the fraction of N,P concordant pairs. This result allows us to make inference regarding the dynamic changes in deterministic complementarities without making any functional form assumptions on the deterministic matching output. In particular, an increase in the fraction of all P,N concordant pairs relative to the fraction of N,P concordant pairs is consistent with a P,N modular increase in the deterministic output function. For parametric purposes, one can use a quadratic function to parameterize a P,N modular increase in the deterministic output function.

Corollary 2.

Function $Q(x, y) = \sum_k \sum_l \theta_{k,l} x_k y_l$ exhibits higher P,N modularity compared to $Q'(x, y) = \sum_k \sum_l \beta_{k,l} x_k y_l$ if and only if

- (a) $\theta_{i,j} \geq \beta_{i,j}$ for all $(i, j) \in P$;
- (b) $\theta_{i,j} \leq \beta_{i,j}$ for all $(i, j) \in N$; and
- (c) $\theta_{i,j} = \beta_{i,j}$ for all $(i, j) \notin P \cup N$.

In this framework, the relationship between bivariate complementarities and bivariate optimal sorting patterns is complicated. That makes the results presented in Proposition 3 and Corollary 2 practically useful.

Example 5. Consider a matching market in which firms and workers have two binary observable attributes.

Assume that $\sigma + \delta = 1$ and

$$\begin{bmatrix} \theta_{1,1} & \theta_{1,2} \\ \theta_{2,1} & \theta_{2,2} \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$$

for quadratic output function $Q(x, y) = \sum_k \sum_l \theta_{k,l} x_k y_l$. The density functions of firms and workers, p_F and p_G , are given below:

$$p_F(x) = \begin{cases} .1 & x = (0, 0) \text{ or } x = (1, 1) \\ .4 & x = (0, 1) \text{ or } x = (1, 0) \end{cases} \quad p_G(y) = \begin{cases} .4 & y = (0, 0) \text{ or } y = (1, 1) \\ .1 & y = (0, 1) \text{ or } y = (1, 0) \end{cases}$$

⁴See Lindenlaub (2017) for a parametric examination of skill-biased changes in the U.S. labor market.

For this parameterization, I numerically approximate optimal matching density $m : \{0, 1\}^2 \times \{0, 1\}^2 \rightarrow (0, 1)$ as follows by using iterative proportional fitting procedure⁵.

		Workers			
		(0, 0)	(0, 1)	(1, 0)	(1, 1)
Firms	(0, 0)	.0848	.0097	.0013	.0042
	(0, 1)	.2739	.0848	.0042	.0371
	(1, 0)	.0371	.0042	.0848	.2739
	(1, 1)	.0042	.0013	.0097	.0848

Optimal matching density function between agents

		Workers	
		0	1
Firms	0	.208	.292
	1	.292	.208

Optimal matching density function between the second attributes

Logarithm of the optimal matching density function exhibits P,N modularity for $P = \{(1, 1), (2, 2)\}$ and $N = \{\}$, which is consistent with Proposition 2. Based on Proposition 2, one can also predict that the deterministic output function is P,N modular for $P = \{(1, 1), (2, 2)\}$ and $N = \{\}$ given the optimal matching density. On the other hand, the optimal matching density exhibits negative correlation between the second attributes of firms and workers. Consequently, an econometrician may predict negative complementarity between the second observable attributes of firms and workers while the deterministic output function exhibits positive complementarity between these attributes, i.e. $\theta_{2,2} = 1 > 0$.

An analysis based on two univariate marginals of multidimensional objects is potentially deceptive not only for the binary case. Let $F : \mathbb{R}^K \rightarrow [0, 1]$ and $G : \mathbb{R}^L \rightarrow [0, 1]$ be the distribution functions of firms' and workers' observable attributes. Assume that the deterministic output function is quadratic, i.e. $Q(x, y; \theta) = \sum_k \sum_l \theta_{k,l} x_k y_l$. Due to Galichon and Salanie (2015), the logarithm of optimal matching density function is obtained as follows for some ϕ and φ :

$$\log m(x, y; \theta, F, G) = W + \frac{Q(x, y; \theta) + \phi(x; \theta, F, G) + \varphi(y; \theta, F, G)}{\sigma + \delta} \quad (2)$$

where

$$W = -\log \left\{ \sum_{x' \in X} \sum_{y' \in Y} \exp \left\{ \frac{Q(x', y'; \theta) + \phi(x'; \theta, F, G) + \varphi(y'; \theta, F, G)}{\sigma + \delta} \right\} \right\}.$$

Let $m_{k,l}(a, b)$ denote the fraction of couples for which (\clubsuit) firms' k^{th} attributes equal a and (\spadesuit) workers' l^{th} attributes equal b . By using Equation (2), the logarithm of bivariate k, l marginal matching density function can be formulated as follows:

$$\log m_{k,l}(a, b; \theta, F, G) = \frac{\theta_{k,l}}{\sigma + \delta} ab + W + \Lambda_{k,l}(a, b) \quad (3)$$

⁵For details, see Deming and Stephan (1940).

where

$$\Lambda_{k,l}(a,b) = \log \left\{ \sum_{x' \in \mathbb{R}^K} \sum_{y' \in \mathbb{R}^L} \mathbb{I}\{x_k = a, y_l = b\} \exp \left\{ \frac{\phi(x'; \boldsymbol{\theta}, F, G) + \varphi(y'; \boldsymbol{\theta}, F, G) + \sum_{(i,j) \neq (k,l)} \theta_{i,j} x_i y_j}{\sigma + \delta} \right\} \right\}.$$

Suppose that a firm's k^{th} attribute may take f_k different values: $\{a_1, \dots, a_{f_k}\}$, and a worker's l^{th} attribute may take w_l different values: $\{b_1, \dots, b_{w_l}\}$ such that $a_{i+1} > a_i$ for $i = 1, \dots, f_k - 1$ and $b_{j+1} > b_j$ for $j = 1 \dots w_l - 1$. Let $s_{k,l}$ denote the logarithm of the local odds ratio for the i^{th} value of a firm's k^{th} attribute and the j^{th} value of a worker's l^{th} attribute:

$$s_{k,l}(i,j) := \log m_{k,l}(a_i, b_j) + \log m_{k,l}(a_{i+1}, b_{j+1}) - \log m_{k,l}(a_{i+1}, b_j) - \log m_{k,l}(a_i, b_{j+1}).$$

Based on Equation (3), it is easy to see that

$$s_{k,l}(i,j) = \frac{\theta_{k,l}}{\sigma + \delta} (a_{i+1} - a_i) (b_{j+1} - b_j) + \Omega(i,j) \quad (4)$$

where $\Omega(i,j) = \Lambda_{k,l}(a_i, b_j) + \Lambda_{k,l}(a_{i+1}, b_{j+1}) - \Lambda_{k,l}(a_i, b_{j+1}) - \Lambda_{k,l}(a_{i+1}, b_j)$.

Notice that the right-hand side of Equation (4) is not only a function of $\theta_{k,l}$ but also all other model parameters, i.e. the distributions of observable attributes and other complementarity parameters. Consequently, an inference based on bivariate sorting patterns is potentially inconsistent and biased. I address this issue in a separate paper by proposing a multidimensional dependence class and devising two cardinal measures of multidimensional dependence.

3 Household Composition and Healthcare Insurance Market

The relationship between agents' health status and their spouses' education levels has been well-studied in medical literature. Jaffe et al. (2005) find the mortality risk among men with cardiovascular disease is higher for those who are married to less-educated women. In addition, they also document that the mortality risk among women with breast cancer is higher for those who are married to less-educated men. Jaffe et al. (2006) note similar findings for men with cardiovascular disease, and show that one's wife's education level is a stronger predictor of her husband's mortality than his own education level. Kravdal (2008) and Skalická and Kunst (2008) report that one's mortality risk decreases with former and current spouses' education levels in Norway. Nilsen et al. (2012) also document a strong association between spousal education and one's self-rated health in Norway. Brown et al. (2014) confirm that spousal education is positively associated with self-rated health in the U.S.

In literature, educational attainment has also been proven to be an important predictor of mortality differentials (see Kunst and Mackenbach, 1994; Elo and Preston, 1996; Borrell et al., 1999; Mackenbach et al., 1999; Manor et al., 2000; Krokstad et al., 2002; Manor et al., 2004). Since better-educated individuals are more likely to be healthy, an attraction between better-educated or healthier individuals may generate a spurious positive association between agents’ health status and their spouses’ education levels. This channel may disqualify one’s spouse’s education level as a robust predictor of one’s health status. Identifying robust predictors of one’s health status is essential for health insurance carriers. In this section, I examine whether or not one’s spouse’s health status and education level are robust predictors of his/her own health status by using the empirical matching framework described in the previous section. This empirical exercise demonstrates how one can apply the aforementioned sorting theory to address several policy-related questions.

3.1 Data

In this paper, I use the IPUMS-CPS data series for 2010-2017. The census data contains individual-level information regarding education and self-rated health levels for 266,569 couples.

2010	2011	2012	2013	2014	2015	2016	2017	TOTAL
35,643	34,664	33,875	34,023	33,498	33,093	30,754	31,019	266,569

Number of Households

For quantitative purposes, I represent each agent by two indices (education and health). For the i^{th} couple in the sample, I denote the woman’s attributes by $x_i = (x_{i,E}, x_{i,H})$, and the man’s attributes by $y_i = (y_{i,E}, y_{i,H})$.

	Index				
	1	2	3	4	5
Education	Less than high school	High school	Some college	College	Post-college
Health	Poor	Fair	Good	Very Good	Excellent

Table 2: Variables

3.2 Association Concepts and Measures

Since the variables of interest are ordinal, I examine the association between these variables by using a rank-correlation measure (Kruskal's gamma) for each year. Here, I define key concepts and measures to calculate association between spouses' education levels and health status.

Definition 15. The i^{th} and the j^{th} couples exhibit concordance(discordance) between

- (a) women's health status and education levels if and only if $(x_{i,H} - x_{j,H})(x_{i,E} - x_{j,E}) > (<) 0$;
- (b) men's health status and education levels if and only if $(y_{i,H} - y_{j,H})(y_{i,E} - y_{j,E}) > (<) 0$;
- (c) men's education levels and their wives' health status if and only if $(x_{i,H} - x_{j,H})(y_{i,E} - y_{j,E}) > (<) 0$;
- (d) men's health status and their wives' education levels if and only if $(y_{i,H} - y_{j,H})(x_{i,E} - x_{j,E}) > (<) 0$;
- and
- (e) men's and their wives' health status if and only if $(x_{i,H} - x_{j,H})(y_{i,H} - y_{j,H}) > (<) 0$.

To measure the association between agents' health status and education levels, I calculate the following Kruskal's gamma statistics for each survey year.

$$\begin{aligned} \Gamma_{H,E}^{W,W} &= \frac{C_{H,E}^{W,W} - D_{H,E}^{W,W}}{C_{H,E}^{W,W} + D_{H,E}^{W,W}} & \Gamma_{H,E}^{M,M} &= \frac{C_{H,E}^{M,M} - D_{H,E}^{M,M}}{C_{H,E}^{M,M} + D_{H,E}^{M,M}} \\ \Gamma_{H,E}^{W,M} &= \frac{C_{H,E}^{W,M} - D_{H,E}^{W,M}}{C_{H,E}^{W,M} + D_{H,E}^{W,M}} & \Gamma_{H,E}^{M,W} &= \frac{C_{H,E}^{M,W} - D_{H,E}^{M,W}}{C_{H,E}^{M,W} + D_{H,E}^{M,W}} & \Gamma_{H,H}^{W,M} &= \frac{C_{H,H}^{W,M} - D_{H,H}^{W,M}}{C_{H,H}^{W,M} + D_{H,H}^{W,M}} \end{aligned}$$

Here, $C_{k,l}^{a,b}$ denotes the fraction of pairs of couples that exhibits concordance between a 's attribute- k and b 's attribute- l for $a, b \in \{W (Women), M (Men)\}$ and $k, l \in \{E (Education), H (Health)\}$. Similarly, $D_{k,l}^{a,b}$ denotes the fraction of pairs of couples that exhibits discordance between a 's attribute- k and b 's attribute- l (see Definition 15). Kruskal's gamma takes values between -1 and 1, inclusively. Values close to -1 indicate strong negative association, and those close to 1 indicate strong positive association.

3.3 Descriptive Statistics

Agents' Own Health Status and Education Levels

As it is illustrated in Table 3, there is a weak positive association between agents' own health status and education levels. The association is slightly more pronounced among women. Tables 10 and 11 indicate that the lower tail distribution of self-rated health does not vary substantially by education level. On average, a level increase in education is associated with .2398 level increase in self-rated health among women, and .2104 level increase among men.

Table 3: Agents' Own Health Status and Education Levels

Statistic	Survey Year							
	2010	2011	2012	2013	2014	2015	2016	2017
$\Gamma_{H,E}^{W,W}$.3121	.2934	.3176	.301	.3044	.2807	.2614	.2707
$\Gamma_{H,E}^{M,M}$.2708	.263	.2748	.2741	.2563	.2475	.2297	.2486

Agents' Own Health Status and Their Spouses' Education Levels

Table 4 shows that there is a weak positive association between agents' own health status and their spouses' education levels. The association is slightly more pronounced for men. Tables 12 and 13 indicate that lower tail distribution of agents' own health status does not vary substantially by their spouses' education levels. On average, a level increase in one's spouse's education is associated with .2189 level increase in one's own health among men, and .2015 level increase among women. These results suggest that the association between men's health status and their wives' education levels is higher than the association between men's health status and their own education levels. On the other hand, the association between women's health status and their husbands' education levels is lower than the association between women's health status and their own education levels.

Table 4: Agents' Own Health Status and Their Spouses' Education Levels

Statistic	Survey Year							
	2010	2011	2012	2013	2014	2015	2016	2017
$\Gamma_{H,E}^{W,M}$.2756	.2565	.2753	.2602	.255	.238	.2331	.244
$\Gamma_{H,E}^{M,W}$.2771	.2669	.2791	.273	.2731	.2597	.2239	.2522

Agents' Own and Their Spouses' Health Status

Table 5 reports a strong positive association between agents' own and their spouses' health status. Individuals are most likely to be married to spouses with the same self-rated health (see Table 14). This strong positive association has an important actuarial implication: the risk associated with a two-person family plan is higher than the aggregate risk associated with two individual plans.

Table 5: One's Health and One's Spouse's Health

Statistic	Survey Year							
	2010	2011	2012	2013	2014	2015	2016	2017
$\Gamma_{H,H}^{W,M}$.7586	.7469	.7328	.7423	.7384	.7486	.7396	.7498

The nation's first and largest private online marketplace for health insurance, *eHealth, Inc.*, reports average insurance premiums and deductibles⁶ for individual and family plans every year. According to the latest⁷ report, a two-person family insurance costs \$717, whereas an individual plan costs \$310 to a single man, and \$332⁸ to a single woman. Although per capita insurance premium is higher for married individuals, insurance policies with lower insurance premiums also have higher deductibles. According to the same report, a two-person family plan has \$8,113 deductible. In addition, an individual plan has \$4,457 deductible for men and \$4,259 deductible for women. Based on these figures, it is not clear whether or not the insurers assess the extra risk associated with family plans adequately. The evaluation of the extra risk implied by the positive association between spouses' health status requires an in-depth analysis which is beyond the scope of this paper.

3.4 Parametric Estimation

Here, I estimate a parametric matching model to explain the association patterns presented above. The main goal is to decompose the mechanism that generates the association patterns into two channels: attraction and distribution. The empirical framework presented in Section 2 allows me to dissociate these two effects. Therefore, this decomposition reveals spurious association patterns. For the purposes of parametric estimation, I adopt two additional assumptions.

Assumption 4. *Deterministic output function $Q : \{1, 2, 3, 4, 5\}^2 \times \{1, 2, 3, 4, 5\}^2 \rightarrow \mathbb{R}$ is quadratic:*

$$Q(\mathbf{x}_r, \mathbf{y}_c) = \sum_{k \in \{H, E\}} \sum_{l \in \{H, E\}} \theta_{k,l} x_{r,k} y_{c,l}$$

where \mathbf{x}_r denotes the attribute vector of a type- r woman, and \mathbf{y}_c denotes the attribute vector of a type- c man for $r, c \in \{1, \dots, 25\}$.

⁶<https://www.healthcare.gov/glossary/deductible/>

⁷https://news.ehealthinsurance.com/_ir/68/20169/eHealth%20Health%20Insurance%20Price%20Index%20Report%20for%20the%202016%20October%202016.pdf

⁸The gender-premium gap exists despite the fact that gender discrimination in premium calculations is illegal.
<https://www.legalmatch.com/law-library/article/health-insurance-discrimination-laws.html>
<https://www.nwlc.org/wp-content/uploads/2015/08/Individual%20Insurance.pdf>

Education	Health				
	1	2	3	4	5
1	1	6	11	16	21
2	2	7	12	17	22
3	3	8	13	18	23
4	4	9	14	19	24
5	5	10	15	20	25

Table 6: Agent's Types

Assumption 5. *The scale parameters of idiosyncratic output components add up to 1, i.e. $\sigma + \delta = 1$.*

Under Assumptions 1-5, the probability of observing a marriage between a type-r woman and a type-c man among all marriages of type-r women, denoted by $p(\mathbf{x}_r, \mathbf{y}_c)$, can be formulated as follows:

$$p(\mathbf{x}_r, \mathbf{y}_c) := \frac{m(\mathbf{x}_r, \mathbf{y}_c)}{\sum_{j=1}^{25} m(\mathbf{x}_r, \mathbf{y}_j)} = \frac{\exp\{Q(\mathbf{x}_r, \mathbf{y}_c) + \varphi(\mathbf{y}_c)\}}{\sum_{j=1}^{25} \exp\{Q(\mathbf{x}_r, \mathbf{y}_j) + \varphi(\mathbf{y}_j)\}}. \quad (5)$$

An equivalent model can be obtained by choosing the first type of man as base category:

$$p(\mathbf{x}_r, \mathbf{y}_c | \boldsymbol{\theta}) = \begin{cases} \frac{1}{1 + \sum_{j \neq 1} \exp \left\{ \sum_{k \in \{H, E\}} \sum_{l \in \{H, E\}} \theta_{k,l} x_{r,k} (y_{j,l} - y_{j,1}) + (\varphi(\mathbf{y}_j) - \varphi(\mathbf{y}_1)) \right\}} & c = 1 \\ \frac{\exp \left\{ \sum_{k \in \{H, E\}} \sum_{l \in \{H, E\}} \theta_{k,l} x_{r,k} (y_{j,l} - y_{j,1}) + (\varphi(\mathbf{y}_c) - \varphi(\mathbf{y}_1)) \right\}}{1 + \sum_{j \neq 1} \exp \left\{ \sum_{k \in \{H, E\}} \sum_{l \in \{H, E\}} \theta_{k,l} x_{r,k} (y_{j,l} - y_{j,1}) + (\varphi(\mathbf{y}_j) - \varphi(\mathbf{y}_1)) \right\}} & c \neq 1 \end{cases}. \quad (6)$$

Let $f^m(\mathbf{x}_r)$ denote the fraction of type-r married women in a sample of N married couples. The likelihood function is formulated as follows:

$$L_N(\boldsymbol{\theta}) = \prod_{i=1}^N m(\mathbf{x}^{(i)}, \mathbf{y}^{(i)} | \boldsymbol{\theta}) = \prod_{i=1}^N \prod_{r=1}^{25} \prod_{c=1}^{25} \{m(\mathbf{x}_r, \mathbf{y}_c | \boldsymbol{\theta})\}^{d_{r,c}^{(i)}} = \prod_{r=1}^{25} \prod_{c=1}^{25} \{f^m(\mathbf{x}_r) p(\mathbf{x}_r, \mathbf{y}_c | \boldsymbol{\theta})\}^{n_{r,c}}$$

where

$m(\mathbf{x}^{(i)}, \mathbf{y}^{(i)} | \boldsymbol{\theta})$: probability of observing the i^{th} couple for parameter vector $\boldsymbol{\theta}$;

$m(\mathbf{x}_r, \mathbf{y}_c | \boldsymbol{\theta})$: probability of observing a marriage between a type-r woman and a type-c man for parameter vector $\boldsymbol{\theta}$;

$d_{r,c}^{(i)}$: binary identifier for the i^{th} couple that equals 1 if the i^{th} marriage is between a type-r woman and a type-c man; and

$n_{r,c}$: number of marriages between type-r women and type-c men.

I estimate the complementarities by maximizing the following log-likelihood function:

$$\mathcal{L}_N(\boldsymbol{\theta}) = \log L_N(\boldsymbol{\theta}) = \sum_{r=1}^{25} n_r \log \{f^m(\mathbf{x}_r)\} + \sum_{r=1}^{25} \sum_{c=1}^{25} n_{r,c} \log \{p(\mathbf{x}_r, \mathbf{y}_c | \boldsymbol{\theta})\} \quad (7)$$

where n_r is the total number of type- r women in the sample.

The log-likelihood function has 28 parameters: 4 attraction parameters, $\{\theta_{k,l}\}$, and 24 deterministic sympathy parameters, $\{\varphi(\mathbf{y}_j) - \varphi(\mathbf{y}_1)\}$. It is a well-known fact that the numerical methods which use a gradient ascent algorithm are less accurate for high-dimensional parameter spaces. In order to overcome this problem, I estimate the parameters^{9,10} by using simulated annealing. As it is illustrated in Tables 10 and 11, the distributions of the agents vary only slightly over time. For this reason, I use the entire sample and estimate only one set of parameters. The estimated parameters are reported in Table 7.

Table 7: Estimated Complementarities

		Men	
		Health	Education
Women	Health	.7625	-.0375
	Education	-.0226	.5572

These results indicate an attraction between individuals with the same education levels and health status. The interpretation of these parameters is a little bit complicated. To interpret $\theta_{H,H}$, consider four agents x, x', y , and y' such that (a) $(x_E - x'_E) = (y_E - y'_E) = 0$, and (b) $(x_H - x'_H)(y_H - y'_H) > 0$. For these agents, one is $e^{\{.7625(x_H - x'_H)(y_H - y'_H)\}}$ times more likely to observe concordance than discordance between spouses' health status (see Equation (1)). In other words, everything held constant, one is at least $2.1436 = e^{.7625}$ times more likely to observe concordance than discordance between spouses' health status. Furthermore, the likelihood of concordance rises with $(x_H - x'_H)(y_H - y'_H)$. The educational attraction parameter can also be interpreted in the same way: everything held constant, one is at least $1.7458 = e^{.5572}$ times more likely to observe concordance relative to discordance between spouses' education levels.

Contrary to weak positive association between agents' own health status and their spouses' education levels, the estimation results suggest a disaffection between healthier individuals and more educated individuals: everything held constant, one is slightly more likely to observe fewer educated individuals marrying healthier individuals. Altogether, the attraction analysis suggests that the weak positive association between agents' own health status and their spouses' education levels is a product of three factors: (a) an attraction between better-educated individuals, (b) an attraction between healthier individuals, and (c) a weak positive

⁹https://en.wikipedia.org/wiki/Simulated_annealing

¹⁰<https://www.mathworks.com/help/gads/simulannealbnd.html>

association between agents' health status and their own education levels. This channel provides a structural support for strong positive association between one's health and one's spouse's health, and an empirical justification for the aforementioned premium gap between family and individual health insurance plans.

Finally, Tables 8 and 9 show that these inferences are highly accurate. The structural model described in Section 3.4 is a close approximation of the limited household formation process. As it is illustrated in Table 8, the predicted association levels are not very different from the empirical association levels. Furthermore, Table 9 demonstrates that the statistical distance between the empirical and the estimated household distributions is small. Shannon entropy measure indicates that a code to generate the predicted household distribution has 7.837 average description length¹¹. If we use a code to generate the empirical household distribution, it will have $8.2322 = 7.837 + .3952$ average description length. Therefore, the efficiency loss associated with using a structural model is only $100 \times \frac{.3952}{8.2322} = 4.8$ percent.

Table 8: Empirical and Predicted Association Levels

Statistics	$\Gamma_{H,H}^{W,M}$	$\Gamma_{H,E}^{W,M}$	$\Gamma_{E,H}^{W,M}$	$\Gamma_{E,E}^{W,M}$
Empirical	.7439	.2546	.2638	.6468
Predicted	.6545	.2017	.2218	.6041

Table 9: Difference Between Realized and Estimated Distributions

Kullback-Leibler Divergence ($\sum_i \hat{m}_i \log_2 \{\hat{m}_i/m_i\}$)	.3952
Shannon Entropy of Predicted Distribution ($-\sum_i \hat{m}_i \log_2 \{\hat{m}_i\}$)	7.837
Efficiency Loss (%)	4.8

Appendix: Proofs

Proof. of Lemma 1

Without loss of generality, assume that $X := \text{supp}(F)$ and $Y := \text{supp}(G)$ have countably many elements. For given matching distribution M , define $\text{vec}M$ as the density vector defined by M . More specifically, the rows of $\text{vec}M$ represents the densities of $(x, y) \in X \times Y$ implied by M . For P,N concordance improving

¹¹https://www.princeton.edu/~cuff/ele201/kulkarni_text/information.pdf

transfer with unit mass $\tau_{P,N}(x, y, x', y'; 1)$, define P,N transfer vector $t_{P,N}(x, y, x', y')$ such that (a) the row associated with (x, y) and (x', y') is 1; (b) the row associated with (x', y) and (x, y') is -1 ; and (c) the rest of the rows are 0. Let $\mathcal{T}(P, N)$ denote the set of P,N transfer vectors.

It suffices to show that the following statement holds:

$$\text{vec}M - \text{vec}M' = \sum_{t \in \mathcal{T}(P, N)} \alpha_t t, \forall \alpha_t \geq 0 \Leftrightarrow M \succeq_{P, N} M'. \quad (8)$$

Define $Q \cdot t_{P,N}(x, y, x', y')$ as follows:

$$Q \cdot t_{P,N}(x, y, x', y') = Q(x, y) + Q(x', y') - Q(x', y) - Q(x, y').$$

Note that the aggregate output improves with P,N transfers:

$$\begin{aligned} & Q \cdot t_{P,N}(x, y, x', y') \\ &= Q(x_1, \dots, x_K, y_1, \dots, y_L) + Q(x'_1, \dots, x'_K, y'_1, \dots, y'_L) - Q(x_1, \dots, x_K, y'_1, \dots, y'_L) - Q(x'_1, \dots, x'_K, y_1, \dots, y_L) \\ &= \sum_{i=1}^K \sum_{j=1}^L \left\{ \begin{array}{l} Q(x_1, \dots, x_{i-1}, x_i, x'_{i+1}, \dots, x'_K, y_1, \dots, y_{j-1}, y_j, y'_{j+1}, \dots, y'_L) \\ + Q(x_1, \dots, x_{i-1}, x'_i, x'_{i+1}, \dots, x'_K, y_1, \dots, y_{j-1}, y'_j, y'_{j+1}, \dots, y'_L) \\ - Q(x_1, \dots, x_{i-1}, x_i, x'_{i+1}, \dots, x'_K, y_1, \dots, y_{j-1}, y'_j, y'_{j+1}, \dots, y'_L) \\ - Q(x_1, \dots, x_{i-1}, x'_i, x'_{i+1}, \dots, x'_K, y_1, \dots, y_{j-1}, y_j, y'_{j+1}, \dots, y'_L) \end{array} \right\} \\ &\geq 0. \end{aligned}$$

Consequently, it holds that

$$Q \in \mathbb{C}(P, N) \Leftrightarrow Q \cdot t \geq 0 \quad \forall t \in \mathcal{T}(P, N). \quad (9)$$

Equation (8) holds if and only if $\text{vec}M - \text{vec}M'$ belongs to the convex cone $\mathcal{C}(P, N)$ generated by $\mathcal{T}(P, N)$:

$$\mathcal{C}(P, N) = \left\{ \sum_{t \in \mathcal{T}(P, N)} \alpha_t t : \alpha_t \geq 0, \forall t \in \mathcal{T}(P, N) \right\}.$$

From Equation (9), it follows that $\mathbb{C}(P, N)$ is the dual cone of $\mathcal{C}(P, N)$. Since $\mathcal{C}(P, N)$ is convex and closed, $\mathcal{C}(P, N)$ is the dual cone of $\mathbb{C}(P, N)$ due to Luenberger (1969, p:215), i.e. for $\{\beta_t\} > 0$, it holds that

$$\sum_{t \in \mathcal{T}(P, N)} \beta_t t \in \mathcal{C}(P, N) \Leftrightarrow \sum_{t \in \mathcal{T}(P, N)} \beta_t Q \cdot t \geq 0 \quad \forall Q \in \mathbb{C}(P, N).$$

Therefore, $M \succeq_{P, N} M'$ if and only if $\text{vec}M - \text{vec}M' \in \mathcal{C}(P, N)$. □

Claim.

(a) The set of P,N undominated distributions is a subset of weak P,N assortative matching distributions.

(b) The set of P,N dominant distributions and global P,N assortative matching distributions coincide.

Proof. Kakutani Fixed Point Theorem: Let $A \subseteq \mathbb{R}^N$ be a non-empty, compact and convex set; $U : A \mapsto P(A)$ be a non-empty-valued, convex-valued correspondence with a closed graph. Correspondence $U : A \mapsto P(A)$ has a fixed point.

For a given $M \in \mathcal{M}(F, G)$ and P, N , define correspondence

$$U(\text{vec}M; P, N) = \begin{cases} \text{vec}M + \sum_{t \in \mathcal{T}(P, N) \setminus \mathcal{T}(N, P)} \alpha_t t + \sum_{t \in \mathcal{T}(P, N) \cap \mathcal{T}(N, P)} \beta_t t & , \forall \alpha_t, \beta_t \geq 0 \text{ and for some } \alpha_t > 0 \\ \text{vec}M & oth. \end{cases} \quad (10)$$

It is clear that $U(\cdot; P, N)$ is non-empty-valued, convex-valued, and has a closed graph. Therefore, $\exists M^* \in \mathcal{M}(F, G)$ such that $\text{vec}M^* \in U(\text{vec}M^*, P, N)$ by Kakutani fixed point theorem.

(a) Let $\text{vec}M^*$ be a fixed point of the correspondence described in Equation 10. Due to Lemma 1, every P,N undominated matching distribution corresponds to a fixed point of the correspondence above. This proves that the set of P,N undominated distributions is non-empty. By definition, any distribution violating weak P,N sorting cannot be a fixed point of the correspondence.

(b) (\Rightarrow) Suppose not. Let $M \in \mathcal{M}(F, G)$ be P,N dominant and not globally P,N assortative. Without loss of generality, assume that $(1, 1) \in P$ and M violates $(1, 1)$ positive sorting by assigning positive mass to $(x, y'), (x', y)$ such that $x_1 > x'_1$ and $y_1 > y'_1$. For matching distribution $M' \in \mathcal{M}(F, G)$ satisfying positive sorting between the first attributes, we have $\int Q dM' > \int Q dM$ when $Q(x, y) = x_1 y_1$. Consequently, M is not a P,N dominant distribution.

(\Leftarrow) Let $M \in \mathcal{M}(F, G)$ satisfy global P,N assortativeness. For any $(x, y), (x', y') \in \text{supp}(M)$, we have (\clubsuit) $(x_i - x'_i)(y_j - y'_j) \geq 0$ for all $(i, j) \in P$ and (\spadesuit) $(x_p - x'_p)(y_q - y'_q) \geq 0$ for all $(p, q) \in N$. Since every pair of matched couples under M is P,N weak concordant, any matching distribution $M' \in \mathcal{M}(F, G)$ satisfies the following condition:

$$\text{vec}M = \text{vec}M' + \sum_{t \in \mathcal{T}(P, N)} \alpha_t t, \text{ for } \alpha_t \geq 0.$$

Due to Lemma 1, it immediately follows that any matching distribution which satisfies global P,N sorting is P,N dominant. \square

Claim. If the set of P,N dominant distributions is non-empty, then it coincides with the set of P,N undominated distributions.

Proof. of Claim. (\Rightarrow) Trivial.

(\Leftarrow) Suppose not. Let $M \in \mathcal{M}(F, G)$ be P,N undominated but not P,N dominant. Consider a P,N dominant distribution: $M' \in \mathcal{M}(F, G)$. Since M is not P,N dominant and M' is, it holds that (a) $\int QdM \leq \int QdM'$ for all $Q \in \mathbb{C}(P, N)$; and (b) there exists output function $Q \in \mathbb{C}_+(P, N)$ such that $\int QdM < \int QdM'$ due to Equation (9). Consequently, M' strictly dominates M in P,N modular order. Therefore, M is not a P,N undominated distribution. \square

Proof. of Proposition 1.

1.a) If a matching distribution does not satisfy within-group P,N sorting, then there exist a P,N concordance improving transfer which is not N,P concordance improving. In other words, the matching distribution does not satisfy weak P,N sorting.

1.b) Suppose not. Let M be a solution to the planner's problem for $Q \in \mathbb{C}_+(P, N)$ which is not a weakly P,N assortative matching distribution. Since a swap between pair of couples which violates weak P,N sorting strictly improves the aggregate output, M cannot be a solution.

1.c) The existence of a weak P,N assortative distribution immediately follows from the existence of the fixed point of the correspondence given in Equation 10.

Case 1: $P \cup N = \emptyset$.

Trivial. Every matching distribution $M \in \mathcal{M}(F, G)$ is associated with the same level of aggregate output.

Case 2: $P \cup N \neq \emptyset$.

Let $M \in \mathcal{M}(F, G)$ be a solution to the planner's problem for $Q \in \mathbb{C}(P, N)$. Suppose M is not weak P,N assortative distribution. Choose $M' \in \mathcal{M}(F, G)$ such that (i) $\text{vec}M' \in U(\text{vec}M; P, N)$, and (ii) $\text{vec}M' \in U(\text{vec}M'; P, N)$. Since M' is obtained from M via a sequence of P,N concordance improving transfers, it cannot worsen the level of aggregate output for any $Q \in \mathbb{C}(P, N)$. Consequently, M' is at least as good as M for any $Q \in \mathbb{C}(P, N)$. Thus, M' also solves the planner's problem. By construction, M' is a P,N undominated distribution. Thus M' satisfies weak P,N sorting.

2.a) Let M be a solution to the planner's problem for $Q \in \mathbb{C}_+(P, N)$ that does not satisfy global P,N sorting. Since the set of P,N dominant distributions coincides with the set of globally P,N assortative matching distributions, and M is not a globally P,N assortative matching distribution. Thus, there exists an N,P concordant pair of matched couples under M . In other words, every globally P,N assortative matching distribution strictly dominates M due to Lemma 1. Therefore, M cannot solve the planner's problem.

2.b) The set of P,N dominant distributions coincides with the set of globally P,N assortative matching distributions. Thus, every globally P,N assortative matching distribution solves the planner's problem for

all $Q \in \mathbb{C}(P, N)$.

□

Proof. of Proposition 2. Immediate from Lemma 2.

□

Proof. of Corollary 1. Immediate from Proposition 2.

□

Proof. of Proposition 3. Immediate from Lemma 2.

□

Proof. of Corollary 2. Immediate from Definition 14.

□

Appendix: Tables

Table 10: Distributions of Women's Health Status and Education Levels by Survey Year

2010	Education				
Health	1	2	3	4	5
1	.0082	.0136	.008	.0026	.0015
2	.0178	.0374	.0237	.0094	.0042
3	.0336	.1035	.0772	.0423	.0216
4	.0227	.0968	.0964	.0754	.0411
5	.0115	.0541	.0747	.0788	.0442

2011	Education				
Health	1	2	3	4	5
1	.0073	.015	.009	.0036	.0015
2	.0165	.0364	.0241	.0109	.0049
3	.0333	.0971	.0745	.0432	.0224
4	.0214	.0975	.1017	.0781	.0422
5	.0123	.0538	.0701	.0774	.0458

2012	Education				
Health	1	2	3	4	5
1	.0074	.0154	.0083	.0037	.0015
2	.0161	.0386	.0252	.0098	.0047
3	.0342	.0947	.0754	.0419	.0215
4	.0214	.0945	.0987	.0806	.0434
5	.0099	.0530	.0702	.0827	.0471

2013	Education				
Health	1	2	3	4	5
1	.007	.0138	.0085	.0031	.0014
2	.0171	.0368	.0254	.0106	.0055
3	.0309	.0937	.0757	.0432	.0229
4	.0189	.0921	.0987	.0827	.0477
5	.0114	.0538	.0708	.0791	.0492

2014	Education				
Health	1	2	3	4	5
1	.0079	.0144	.0082	.0031	.0018
2	.0167	.0374	.0256	.0118	.0064
3	.0275	.0933	.0764	.0445	.024
4	.0185	.0899	.0984	.0836	.0492
5	.0121	.0471	.0709	.0805	.0509

2015	Education				
Health	1	2	3	4	5
1	.006	.0133	.0094	.0036	.002
2	.0151	.0336	.0247	.0122	.0053
3	.0296	.0915	.0758	.0459	.0255
4	.0212	.0914	.0946	.0835	.0485
5	.0125	.0519	.068	.0816	.0531

2016	Education				
Health	1	2	3	4	5
1	.0061	.0118	.0078	.0031	.0021
2	.0139	.0347	.0241	.013	.0067
3	.0271	.0894	.0792	.0488	.0291
4	.0204	.0855	.0970	.0871	.0517
5	.0122	.05	.0682	.0806	.0503

2017	Education				
Health	1	2	3	4	5
1	.0062	.0121	.0082	.0029	.0024
2	.0133	.0325	.0258	.0122	.0051
3	.0275	.0868	.0809	.0514	.0283
4	.0172	.0844	.0968	.0914	.053
5	.0116	.0501	.0661	.0811	.0526

Table 11: Distributions of Men's Health Status and Education Levels by Survey Year

2010	Education				
Health	1	2	3	4	5
1	.0127	.0146	.0084	.0046	.0021
2	.0191	.0376	.0223	.0113	.0064
3	.04	.0967	.069	.0441	.0256
4	.026	.0981	.0862	.0734	.0431
5	.0158	.061	.0637	.0693	.049

2011	Education				
Health	1	2	3	4	5
1	.0108	.015	.0098	.0041	.0027
2	.0185	.0347	.0227	.0125	.0069
3	.0386	.0942	.067	.0417	.0237
4	.0262	.0997	.0901	.0742	.0468
5	.0153	.058	.0621	.0707	.0503

2012	Education				
Health	1	2	3	4	5
1	.0108	.0168	.0081	.004	.0032
2	.0203	.0367	.0225	.0108	.0069
3	.0386	.0875	.068	.0434	.0275
4	.0256	.0955	.0878	.0776	.0484
5	.0148	.0568	.0644	.074	.0502

2013	Education				
Health	1	2	3	4	5
1	.0115	.0145	.0089	.0029	.0019
2	.0188	.0344	.0232	.0118	.0077
3	.0342	.0935	.068	.0446	.0288
4	.0246	.0932	.0907	.0811	.0484
5	.0143	.056	.0629	.0718	.0524

2014	Education				
Health	1	2	3	4	5
1	.0101	.0156	.0077	.0038	.0019
2	.0191	.0354	.0244	.0114	.0079
3	.0337	.0912	.0707	.0468	.0301
4	.0235	.0921	.0912	.0783	.0484
5	.015	.0588	.0613	.0701	.0514

2015	Education				
Health	1	2	3	4	5
1	.0088	.0135	.0093	.004	.0018
2	.0172	.0343	.0242	.0126	.008
3	.0337	.0914	.0702	.0462	.0293
4	.0257	.0914	.0879	.0777	.0507
5	.0159	.0577	.0615	.0752	.0517

2016	Education				
Health	1	2	3	4	5
1	.0087	.0132	.0081	.0038	.0029
2	.015	.032	.0241	.0124	.0078
3	.0336	.089	.0756	.0511	.0304
4	.0248	.0894	.0869	.0818	.0528
5	.0159	.0561	.0619	.0716	.0511

2017	Education				
Health	1	2	3	4	5
1	.0081	.0122	.0082	.0031	.0023
2	.0152	.0344	.0244	.0141	.0079
3	.0336	.0916	.0735	.0504	.0315
4	.0221	.0859	.0887	.085	.0537
5	.0126	.0555	.0596	.075	.0511

Table 12: Distributions of Women's Health Status and Their Husbands' Education Levels by Survey Year

2010	Education				
Health	1	2	3	4	5
1	.0095	.0125	.0071	.0031	.0017
2	.0196	.0346	.0211	.0106	.0064
3	.0415	.0993	.0692	.0435	.0246
4	.0286	.0996	.0861	.0737	.0444
5	.0146	.062	.0661	.0717	.049

2011	Education				
Health	1	2	3	4	5
1	.0078	.0133	.0085	.0045	.0023
2	.0188	.0349	.0215	.0109	.0068
3	.0399	.0929	.0664	.0429	.0284
4	.0283	.1014	.0909	.0743	.0461
5	.0147	.0591	.0644	.0707	.0505

2012	Education				
Health	1	2	3	4	5
1	.0089	.0137	.0073	.0037	.0029
2	.0195	.0352	.0216	.0114	.0066
3	.0395	.0908	.0677	.0418	.0278
4	.0279	.0973	.0891	.0772	.047
5	.0143	.0564	.065	.0754	.0519

2013	Education				
Health	1	2	3	4	5
1	.0081	.0124	.0072	.0036	.0024
2	.0199	.0332	.0229	.0113	.0081
3	.0363	.0922	.0666	.0428	.0284
4	.0248	.0948	.0903	.0816	.0487
5	.0142	.059	.0667	.0729	.0515

2014	Education				
Health	1	2	3	4	5
1	.0084	.0133	.0075	.0038	.0025
2	.0196	.035	.0244	.0108	.0081
3	.0338	.0908	.0677	.0453	.028
4	.0255	.0947	.0911	.0782	.0501
5	.014	.0594	.0646	.0724	.0511

2015	Education				
Health	1	2	3	4	5
1	.0068	.0129	.0084	.0037	.0024
2	.017	.0308	.0234	.0125	.0073
3	.0352	.09	.0685	.0453	.0294
4	.026	.0955	.0891	.079	.0496
5	.0163	.059	.0637	.0752	.0528

2016	Education				
Health	1	2	3	4	5
1	.0066	.0114	.0074	.0031	.0023
2	.016	.0326	.023	.0125	.0083
3	.0338	.0873	.0744	.048	.0301
4	.0252	.0928	.0892	.0827	.0519
5	.0164	.0557	.0625	.0744	.0524

2017	Education				
Health	1	2	3	4	5
1	.0061	.0126	.0074	.0034	.0022
2	.0155	.0321	.0219	.0121	.0072
3	.0322	.0899	.0722	.0497	.0313
4	.0231	.0886	.0926	.0856	.0528
5	.0146	.0564	.0606	.0769	.053

Table 13: Distributions of Men's Health Status and Their Wives' Education Levels by Survey Year

2010	Education				
Health	1	2	3	4	5
1	.0089	.0172	.0101	.004	.0021
2	.0174	.0386	.023	.0113	.0063
3	.0325	.0983	.0783	.0438	.0226
4	.0225	.0954	.0945	.0743	.0401
5	.0124	.0558	.074	.075	.0415

2011	Education				
Health	1	2	3	4	5
1	.0086	.0172	.0099	.0042	.0025
2	.0158	.0357	.0251	.0128	.006
3	.0316	.0958	.0748	.0446	.0221
4	.0218	.0966	.099	.0772	.0424
5	.0132	.0547	.0705	.0744	.0437

2012	Education				
Health	1	2	3	4	5
1	.008	.0181	.0104	.0036	.0027
2	.0157	.0382	.0253	.0121	.0058
3	.031	.0933	.0731	.0443	.0231
4	.0223	.0912	.0996	.0797	.0421
5	.0118	.0554	.0693	.0792	.0445

2013	Education				
Health	1	2	3	4	5
1	.0078	.0163	.0099	.0035	.0021
2	.0151	.0362	.0257	.0123	.0067
3	.0305	.0931	.0751	.045	.0253
4	.02	.0911	.1001	.0811	.0458
5	.012	.0536	.0683	.0767	.0468

2014	Education				
Health	1	2	3	4	5
1	.0078	.0154	.0096	.0039	.0026
2	.0148	.0392	.0244	.0126	.0072
3	.0281	.0925	.0778	.0478	.0264
4	.019	.086	.0979	.0826	.0479
5	.013	.049	.0698	.0766	.0482

2015	Education				
Health	1	2	3	4	5
1	.0068	.00151	.0091	.004	.0025
2	.0144	.0035	.0275	.0132	.0062
3	.029	.0905	.0756	.0478	.0279
4	.0218	.0886	.0914	.0829	.0477
5	.0124	.0526	.0679	.0791	.0501

2016	Education				
Health	1	2	3	4	5
1	.0061	.0129	.0103	.0047	.0026
2	.0121	.0352	.0247	.0135	.0085
3	.0284	.0885	.08	.0515	.0314
4	.0207	.0856	.0938	.086	.0496
5	.0123	.052	.0675	.077	.0477

2017	Education				
Health	1	2	3	4	5
1	.0055	.0133	.009	.0039	.0021
2	.0126	.0348	.0271	.0135	.0078
3	.0283	.0872	.0816	.0539	.0299
4	.0178	.0816	.0958	.0881	.0522
5	.0115	.0489	.0642	.0798	.0493

Table 14: Distributions of Women's and Their Husbands' Health Status by Survey Year

2010	Men				
Women	1	2	3	4	5
1	.0135	.0074	.0069	.0038	.0024
2	.0096	.041	.0242	.0111	.0064
3	.0116	.0287	.1863	.0338	.0177
4	.0049	.0125	.0402	.2493	.0254
5	.0027	.0071	.0179	.0287	.2068

2011	Men				
Women	1	2	3	4	5
1	.0129	.0072	.0083	.0053	.0027
2	.0105	.0415	.0217	.0119	.0072
3	.0111	.0277	.1794	.0343	.0179
4	.0051	.012	.0406	.2584	.0248
5	.0027	.0069	.0189	.0271	.2038

2012	Men				
Women	1	2	3	4	5
1	.013	.0065	.0084	.0053	.0032
2	.01	.0417	.0227	.0121	.0078
3	.0114	.0285	.1724	.0357	.0196
4	.0055	.0136	.0424	.252	.0251
5	.0031	.0068	.0189	.0297	.2046

2013	Men				
Women	1	2	3	4	5
1	.0116	.0064	.008	.0051	.0027
2	.0097	.0417	.0249	.0123	.0068
3	.0111	.0272	.1747	.0367	.0167
4	.0045	.0131	.0415	.2547	.0263
5	.0027	.0075	.02	.0293	.2047

2014	Men				
Women	1	2	3	4	5
1	.0114	.0082	.0084	.0043	.0032
2	.0098	.0443	.0245	.0126	.0066
3	.0089	.0269	.1772	.0335	.019
4	.006	.0122	.0418	.2523	.0272
5	.003	.0066	.0206	.0307	.2006

2015	Men				
Women	1	2	3	4	5
1	.0113	.0073	.0083	.0044	.003
2	.0091	.0405	.0228	.0119	.0067
3	.0093	.0279	.1805	.0326	.018
4	.0051	.0136	.0407	.2542	.0257
5	.0027	.007	.0185	.0303	.2087

2016	Men				
Women	1	2	3	4	5
1	.011	.0054	.0072	.0046	.0027
2	.0083	.0421	.024	.0112	.0068
3	.0102	.023	.1868	.0343	.0194
4	.005	.0129	.042	.2567	.0252
5	.0022	.008	.0197	.0289	.2024

2017	Men				
Women	1	2	3	4	5
1	.0106	.0067	.0072	.0045	.0027
2	.0074	.0426	.0021	.0105	.0063
3	.0085	.0267	.188	.0359	.016
4	.0051	.0126	.0421	.2576	.0253
5	.0022	.0073	.0215	.027	.2035

References

- [1] Becker, G. S. A theory of marriage: Part i. *Journal of Political Economy* **81(4)**, (1973).
- [2] Belasen, A. R. and Belasen, A. T. Dual effects of improving doctor-patient communication: patient satisfaction and hospital ratings. (2018).
- [3] Belot, M. and Francesconi, M. Dating preferences and meeting opportunities in mate choice decisions. *The Journal of Human Resources* **42(8)**, 474-508 (2012).
- [4] Bojilov, R. and Galichon, A. Matching in closed-form: equilibrium, identification, and comparative statics. *Economic Theory* **61(4)**, 587–609 (2016).
- [5] Borrell, C., Regidor, E., Arias, L., Navarro, P., Puigpinos R., Dominguez, V., and Plansencia, A. Inequalities in mortality according to educational level in two large Southern European cities. *International Journal of Epidemiology* **28(1)**, 58–63 (1999).
- [6] Brown, D. C., Hummer, R. A., and Hayward, M. D. The Importance of Spousal Education for the Self-Rated Health of Married Adults in the United States *Popul. Res. Policy Rev.* **33(1)**, 127–151 (2014).
- [7] Chade, H., Eeckhout J., and Smith, L. Sorting through search and matching models. *Journal of Economic Literature* **55(2)**, 1-52 (2017).
- [8] Chiappori, P. A., Oreffice, S., and Quintana-Domeque, C. Bidimensional matching with heterogeneous preferences: Smoking in the marriage market. *Journal of European Economic Association* **16(1)**, 161–198 (2017).
- [9] Chiappori, P. A., Oreffice, S., and Quintana-Domeque, C. Fatter attraction: anthropometric and socioeconomic matching on the marriage market. *Journal of Political Economy* **120(4)**, 659-695 (2012).
- [10] Chiappori, P. A., McCann, R., and Pass, B. Multidimensional matching. *Preprint*. (2016).
- [11] Choo, E., and Siow, A. Who marries whom and why. *Journal of Political Economy* **114(1)**, 175-201 (2006).
- [12] Colangelo, A., Scarsini, M., and Shaked, M. Some notions of multivariate positive dependence. *Insurance: Mathematics and Economics* **37(1)**, 13-26 (2005).
- [13] Deming, D. J. The growing importance of social skills in the labor market. *The Quarterly Journal of Economics* **132(4)**, 1593-1640 (2017).
- [14] Domingue, B. W., Fletcher, J., Conley, D., and Boardman, J. D. Genetic and educational assortative mating among US adults. *PNAS*, **111(22)**, 7996-8000 (2014).
- [15] Dupuy, A. and Galichon, A. Personality traits and the marriage market. *Journal of Political Economy* **122(6)**, 1271-1319 (2014).

- [16] Deming, W. E. and Stephan, F. F. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* **11(4)**, 427-444 (1940).
- [17] Elo, I. and Preston S. H. Educational differentials in mortality: United States, 1979–85. *Soc. Sci. Med.* **42(1)**, 47–57 (1996).
- [18] Fletcher, J. M. and Padron, N. A. Heterogeneity in spousal matching models. (2015).
- [19] Gabaix X. and Landier, A. Why has CEO pay increased so much?. *Quarterly Journal of Economics* **123(1)**, 49–100 (2008).
- [20] Gale D. and Shapley, L. S. College admissions and the stability of marriage. *The American Mathematical Monthly* **69(1)**, 9-15 (1962).
- [21] Galichon, A. and Salanié, B. Matching with trade-offs: revealed preferences over competing characteristics. (2010).
- [22] Galichon, A. and Salanié, B. Cupid’s invisible hand: social surplus and identification in matching models. *Columbia University Academic Commons*. (2015).
- [23] Gemici, A. and Laufer, S. Marriage and cohabitation. (2010).
- [24] Girsberger, E. M., Rinawi, M., and Krapf, M. Wages and employment: The role of occupational skills. *IZA Discussion Paper Series* **11586** (2018).
- [25] Greenwood, J., Guner, N., Kocharkov, G., and Santos, C. Marry your like: Assortative mating and income inequality. *American Economic Review* **104(5)**, 348-53 (2014).
- [26] Gretskey, N. E., Ostroy, J. M., and Zame W. R. The nonatomic assignment model. *Econ. Theory*, **2(1)**, 103–27. (1992).
- [27] Guvenen, F., Kuruscu, B., Tanaka, S., and Wiczer, D. Multidimensional skill mismatch (2018).
- [28] Günter, J., Hitsch, G. J., Hortaçsu, A., and Ariely, D. Matching and sorting in online dating. *American Economic Review* **100(1)**, 130-163 (2010).
- [29] Hugtenburg, J. G., Timmers, L., Elders, P. J., Vervloet, M., and Dijk, L. Definitions, variants, and causes of nonadherence with medication: a challenge for tailored interventions. *Patient Prefer Adherence* **7**, 675–682 (2013).
- [30] Jaffe, D. H., Eisenbach, Z., Neumark, Y. D., Manor, O. Does one’s own and one’s spouse’s education affect overall and cause-specific mortality in the elderly?. *International Journal of Epidemiology* **34(6)**, 1409–1416 (2005).
- [31] Jaffe, D. H., Eisenbach, Z., Neumark, Y. D., Manor, O. Effects of husbands’ and wives’ education on each other’s mortality. *Social Science & Medicine* **62(8)**, 2014–2023 (2006).

- [32] Jagosh, J., Boudreau, J. D., Steinert, Y., MacDonald, M. E., and Ingram, L. The importance of physician listening from the patients' perspective: Enhancing diagnosis, healing, and the doctor-patient relationship. *Patient Education and Counseling* **85(3)**, 369–374 (2011).
- [33] Klostad, C. A., McDermott, R., and Hatemi, P. K. The dating preferences of liberals and conservatives. *Political Behavior* **35(3)**, 519–538 (2013).
- [34] Kravdal, Ø. A broader perspective on education and mortality: Are we influenced by other people's education?. *Social Science & Medicine* **66(3)**, 620–636 (2008).
- [35] Kremer, M. The o-ring theory of economic development. *Quarterly Journal of Economics* **108(3)**, 551–75 (1993).
- [36] Krokstad, S., Johnsen, R., and Westin, S. Trends in health inequalities by educational level in a Norwegian total population study. *Journal of Epidemiology and Community Health* **56(5)**, 375–380 (2002).
- [37] Kunst, A. E. and Mackenbach, J. P. The size of mortality differences associated with educational level in nine industrialized countries. *American Journal of Public Health* **84(6)**, 932–937 (1994).
- [38] Lindenlaub, I. Sorting multidimensional types: theory and application. *The Review of Economic Studies* **84(2)**, 718–789 (2017).
- [39] Luenberger, D. G. Optimization by vector space methods. *Wiley* (1969).
- [40] Muller, A. and Scarsini, M. Fear of loss, inframodularity, and transfers. *Journal of Economic Theory* **147(4)**, 1490–1500 (2012).
- [41] Mackenbach, J. P. et al. Socioeconomic inequalities in mortality among women and among men: an international study. *American Journal of Public Health* **89(12)**, 1800–1806 (1999).
- [42] Mangasarian, O. L. Uniqueness of solution in linear programming. *Linear Algebra and its Applications* **(25)**, 151–162 (1979).
- [43] Manor, O. et al. Educational differentials in mortality from cardio-vascular disease among men and women: The Israel Longitudinal Mortality Study. *Annals of Epidemiology* **(14)**, 453–460 (2004).
- [44] Manor, O. et al. Mortality differentials among women: The Israel Longitudinal Mortality Study. *Social Science & Medicine* **51(8)**, 1175–1188 (2000).
- [45] Mazzi, M. A., Rimondini, M., Zee, E., Boerma, W., Zimmermann, C., and Bensing, J. Which patient and doctor behaviours make a medical consultation more effective from a patient point of view. Results from a European multicentre study in 31 countries. *Patient Education and Counseling* **101(10)** (2018).
- [46] Meyer, M. and Strulovici, B. The supermodular stochastic ordering (2013).
- [47] Nelsen, R. B. An introduction to copulas. *Springer Series in Statistics* (2006).

- [48] Nilsen, S. M. et al. Education-based health inequalities in 18,000 Norwegian couples: the Nord-Trøndelag Health Study (HUNT). *BMC Public Health* **(12)**, 998 (2012).
- [49] Scarsini, M. On measures of concordance. *Stochastica* **(8)**, 201–218 (1984).
- [50] Shapley, L. S. and Shubik, M. The assignment game I: the core. *International Journal of Game Theory* **1(1)**, 111–30 (1972).
- [51] Shaked, M and Shanthikumar, J. G. Supermodular stochastic orders and positive dependence of random vectors. *Journal of Multivariate Analysis* **61(1)**, 86-101 (1997).
- [52] Shaked, M. and Shanthikumar, J. G. Stochastic order. *Springer* (2007).
- [53] Siow, A. Testing Becker’s theory of positive assortative matching. *Journal of Labor Economics* **33(2)**, 409-44 (2015).
- [54] Skalicka, V. and Kunst, A. E. Effects of spouses’ socioeconomic characteristics on mortality among men and women in a Norwegian longitudinal study. *Soc. Sci. Med.* **66(9)**, 2035-47 (2008).
- [55] Stavropoulou, C. Non-adherence to medication and doctor-patient relationship: Evidence from a European survey. *Patient Education and Counseling* **83(1)**, 7-13 (2011).
- [56] Szekli, R., Disney, R. L., and Hur, S. MR/GI/1 queues with positively correlated arrival stream. *Journal of Applied Probability* **31(2)**, 497-514 (1994).
- [57] Villani, C. Optimal transport, old and new. *Springer* (2008).