Multidimensional Matching with a Potential Handicap: Smoking in the Marriage Market*

Pierre-André Chiappori	Sonia Oreffice	Climent Quintana-Domeque
Columbia University	Universitat d'Alacant and IZA	Universitat d'Alacant and IZA

August 2012

Abstract

We develop a two-dimensional matching model on the marriage market, where individuals are characterized by a continuous trait (e.g., socioeconomic status) and a discrete attribute (e.g., smoking status), and gains from marriage may be diminished by the discrete characteristic. We show that a stable match always exists but may fail to be pure, and we derive some general properties. We then further specify the model by assuming a quadratic surplus function. In that case, the stable match can be fully characterized in closed form. The model generates clear-cut conditions regarding matching patterns. Using CPS data and its Tobacco Use Supplements for the years 1996 to 2007, and proxing socioeconomic status by educational attainment, we find that these conditions are satisfied. There are fewer "mixed" couples where the wife smokes than vice-versa, and matching is assortative on education within smoking types of couples. Among non-smoking wives those with smoking husbands have on average 0.14 fewer years of completed education than those with non-smoking husbands. Finally, we find that among smoking husbands those who marry smoking wives have on average 0.16 more years of completed education than those with non-smoking wives.

Keywords: Marriage market, multidimensional matching, smoking, education.

JEL Codes: D1, J1.

*We thank Bernard Salanié and Marco Francesconi, as well as participants at the Barcelona MOVE Family Conference 2011, Royal Economic Society Meetings 2012, Columbia University, University of Aarhus, and Universidad Carlos III, for helpful comments and suggestions. Financial support from the NSF (award # SES-1124277) is gratefully acknowledged. Oreffice and Quintana-Domeque acknowledge financial support from the Spanish Ministry of Science and Innovation (ECO 2011-29751). The usual disclaimers apply.

1 Introduction

Multidimensional matching The analysis of matching models under transferable utility (from now on TU) has recently attracted considerable attention, from both theoretical and empirical perspectives. Most models focus on exactly one characteristic on which the matching process is assumed to be exclusively based. Various studies have thus investigated the features of assortative matching on income, wages or education (e.g., Becker, 1991; Lam, 1988; Choo and Siow, 2006; Flinn and Del Boca, 2005; Pencavel, 1998; Wong, 2003), but also on such preference-based notions as risk aversion (e.g., Chiappori and Reny, 2004; Legros and Newman, 2007) or desire to have a child (Chiappori and Oreffice, 2008).

One-dimensional matching models offer several advantages. Their formal properties are by now well established. In a TU context, they provide a simple and elegant way to explain the type of assortative matching patterns that are currently observed; namely, the stable match is positive (negative) assortative if and only if the surplus function is super (sub) modular. Moreover, it is possible, from the shape of the surplus function, to recover the equilibrium allocation of resources within each match, a feature that proves especially useful in many theoretical approaches. Arguments of this type have been applied, for instance, to explain why female demand for university education may outpace that of men (Chiappori, Iyigun and Weiss, 2009; Chiappori, Salanié and Weiss, 2011), or how women unwilling to resort to abortion still benefited from its legalization (Chiappori and Oreffice, 2008).

These advantages, however, come at a cost, notably in terms of realism. Empirical evidence strongly suggests that, in real life, matching processes are actually multidimensional; spouses tend to be similar in a variety of characteristics, including age, education, race, religion, and anthropometric characteristics such as weight or height (e.g., Becker, 1991; Hitsch, Hortaçsu, and Ariely, 2010; Oreffice and Quintana-Domeque, 2010; Qian, 1998; Silventoinen et al., 2003;

Weiss and Willis, 1997). On the theoretical front, the basic mathematical structure underlying matching models under TU, namely optimal transportation theory, does not require one-dimensional traits; the general results (e.g., dual equivalence between cost minimization on the set of stable matches and surplus maximization, existence of a stable matching, etc.) hold true under very general assumptions, and in particular in multidimensional contexts. Clearly, some properties (such as supermodularity of the surplus function or assortativeness of the stable matching) are intrisincally one-dimensional; however, they can be generalized to a multidimensional setting (e.g., Chiappori, McCann and Nesheim, 2010, for a recent survey). In principle, thus, we do have the conceptual tools needed to study multidimensional matching. In practice, however, multidimensional models raise a series of specific and often difficult problems. For that reason, situations involving multidimensional matching have very rarely been analyzed in the applied theory literature.[1]

To the best of our knowledge, the present paper is the first to propose and solve in closed form a model of 'truly' multidimensional matching under TU. The framework can be summarized as follows. We consider a matching game in which individuals differ in two characteristics. One characteristic is continuous, and can be interpreted as an index of socioeconomic status (from now on SES), reflecting differences in education, income, social prestige and others. The other characteristic is discrete, and more precisely dichotomic; in our empirical application, we use individuals' smoking status (which we conveniently assume to be exogenous). The surplus function we consider is differentiable and supermodular in the continuous indices, and is impacted by the discrete characteristic in a multiplicative way, i.e., it is diminished by the presence of a smoker. We first prove existence and generic uniqueness of the stable matching, and describe a general resolution strategy. Our approach relies on

[1] Among the few exceptions, one can mention Coles and Francesconi (2011), who study a bidimensional search model, and Hitsch, Hortaçsu, Ariely (2010), and Banerjee, Duflo, Gathak, Lafortune (2012), who consider dating patterns over multiple traits using a Gale-Shapley, non transferable utility (NTU) approach. In the TU framework, Chiappori, Oreffice and Quintana-Domeque (2012) consider situations in which several dimensions can be summarized by a single index, while Galichon and Salanié (2010) analyze the case of a product-separable surplus function from an empirical perspective.

the duality between stability and surplus maximization; specifically, we show that the stable matching can be derived as the solution of an optimal control problem. Moreover, we show how the surplus allocation between spouses can be recovered from the stability conditions.

An interesting aspect is that our surplus function, although supermodular in the continuous indices, does *not* satisfy the 'twisted' condition of Chiappori, McCann and Nesheim (2010), which generalizes the standard Spence-Mirrlees property to a multidimensional framework. In particular, when the respective smoking prevalences differ in the male and the female populations, the stable matching, while assortative in the continuous indices, need not be *pure* (in the sense that the corresponding measure is borne by the graph of a function). In practice, a violation of purity implies that (an open set of) identical agents may be matched with different mates.[2]

We then specialize our approach to a simple setting that assumes independence between SES and smoking status and uses a symmetric quadratic surplus function; however, and quite crucially, we allow for smoking prevalence to differ across genders, since in the US, and several other countries, males are more likely to smoke. In this simple framework, we show that the (unique) stable matching can be fully characterized in closed form solutions. This matching is not pure: an open set of male smokers are matched with either a smoking or a non-smoking wife with positive probability. We also show that the properties characterizing the stable matching generate testable implications on observed matching patterns. For instance, theory prohibits mixed couples in which the wife smokes while the husband does not; while empirical patterns are not so clear-cut, we should expect such couples to be *less* frequent than those in which he smokes and she does not. Among couples with identical smoking status matching should be assortative on SES. More interestingly, non-smoking wives married to a smoking husband should have a lower SES than those married to a non-smoking husband, while the opposite regularity should prevail among males (smoking husbands married to non-smoking

[2] An alternative but equivalent interpretation is that an open set of agents may at equilibrium be indifferent between several possible matches and randomize between them.

wives are of lower 'quality', i.e., lower SES). Lastly, while there is a well-known negative correlation between smoking and education (our measure of SES), the correlation should be less negative for men married to non-smoking women, when controlling for wives' education.

These predictions can readily be tested in 'reduced form' on actual data; the last part of this paper is devoted to such empirical tests. We use March CPS data combined with the CPS Tobacco Use Supplements for the period 1996 to 2007. The TUS supplements are widely used in medical research on smoking, provide a large sample size representative of the US population, and, crucially, it is possible to retrieve information on *both* spouses.[3] Focusing on young couples[4], we show that there is strong sorting by smoking status: there are 71.78% of couples were both spouses are non-smokers, and 10.01% where both smoke. Our data also reveal that there are fewer "mixed" couples where the wife smokes than vice versa, 6.50% versus 11.71% ; the corresponding ratio is 0.55, which is significantly lower than the 0.71 implied by the sole difference in relative smoking prevalence. Using education as our measure of SES, our regression analysis confirms that, among couples with identical smoking status, matching is assortative on education. We also find strong support for our third prediction: among non-smoking wives those with smoking husbands have on average 0.14 fewer years of completed education than those with non-smoking husbands; conversely, and in a more counterintuitive way from an applied point of view, among smoking husbands those who marry smoking wives exhibit on average 0.16 *more* years of completed education than those with non-smoking wives. Finally, the well known negative correlation between education and smoking is confirmed in our data. For instance, an additional year of education is associated with a reduction in the probability of smoking of around 3.6 percentage points in the case of married men. Considering only men married to non-smoking women, the magnitude of this relationship decreases to 2.8 percentage points. However, if one further controls for the wife's education, this correlation becomes even less negative, in line with the fourth and last

[3] In the sensitivity analysis subsection, we will provide estimates using alternative data sets.

[4] Couples whose husbands are between 24 and 34 years old, and wives between 22 and 32 years old.

prediction of the model.

Related literature The theoretical analysis of matching under TU dates back to Koopmans and Beckmann (1957), Shapley and Shubik (1971), and Becker (1973). In particular, the last two contributions show that the stable matching maximizes aggregate surplus, and that the associated individual surpluses solve the dual imputation problem. In turn, the surplus maximization problem belongs to the class of optimal transportation problems, which date back to Monge (1781) and Kantorovich (1942); see Villani (2003) and McCann and Guillen (2010) for recent illustrations. The precise connection between matching models and optimal transportation has been analyzed by Gretsky, Ostroy, and Zame (1999) in the discrete case, and by Ekeland (2010) and Chiappori, McCann, and Nesheim (2010) in the continuous one; the 'twisted buyer condition' used here, which generalizes Spence-Mirrlees to a multidimensional setting, is in Chiappori, McCann, and Nesheim (2010). Less work has been devoted to multidimensional matching, with the notable exception of Galichon and Salanié (2010), Chiappori, Oreffice and Quintana-Domeque (2012), and more recently McCann, Shi, Siow, Wolthoff (2012).[5] However, there exists a natural relationship between matching and hedonic models, on which a recent literature exists (e.g., Ekeland, Heckman and Nesheim, 2004, Heckman, Matzkin and Nesheim, 2010, and Nesheim, 2012, for a multidimensional approach). Finally, there exists a close connection between optimal transportation theory and principal-agents models under asymmetric information (e.g., Figalli, Kim and McCann, 2011). In particular, multidimensional models of adverse selection have been studied by a large literature, starting with Rochet and Choné (1998); see Rochet and Stole (2003) for a survey. The precise links between the latter models and our approach remain to be investigated.

From a more applied perspective, assortative marriage by smoking habits has been previously and extensively documented in the medical and biological literatures (e.g., Sutton,

[5] In addition, several recent models analyze multidimensional matching in a NTU framework from an empirical perspective (e.g., Banerjee, Duflo, Gathak, Lafortune, 2012; Hitch, Hortaçsu and Ariely, 2010).

1980; Venters et al., 1984; Sutton, 1993), although there has been very little economic focus. In the UK, Clark and Etilé (2006) use data from the British Household Panel Survey 1991–1999 to document positive sorting by smoking status. Recently, Maralani (2009) shows the existence of assortative mating by smoking across *old* cohorts of Americans born between the 1920s and the 1950s, using data from the Health and Retirement Study. However, there is no matching model investigating the formation of couples by smoking status, no focus on recent years, and even less so any consideration regarding the different surplus generated by smoker, non-smoker, and "mixed" couples, or the gender gap in smoking prevalence.

The next Section discusses the general model. Section 3 analyzes the symmetric quadratic case, and fully characterizes the stable matching. Section 4 is devoted to the empirical application. We first describe the data; we next test our predictions; sensitivity analysis are presented in the last subsection. Section 5 concludes.

2 The Model

2.1 The basic framework

2.1.1 Populations

We consider two populations (men and women) of equal size, normalized to one. Agents differ in two respects. First, they are characterized by a continuous index; one may, without loss of generality, assume that this index is uniformly distributed over the interval $[0,1]$. A possible interpretation is in terms of SES; then the index depends on the agent's income, education, prestige, or any combination of those. Second, agents are also characterized by some dichotomous indicator taking values in the set $\{N,S\}$; in our empirical application, S stands for 'smoker' and N for 'non-smoker', although alternative interpretations are obviously possible. An agent is thus formally characterized by a pair (x,X) if female and (y,Y) if male, where x or $y\in[0,1]$ is the agent's continuous index, and $X,Y\in\{N,S\}$ defines the agent's discrete characteristic. Let F (resp. G) denote the distribution of female (male) characteristics (x, X) $((y, Y))$ over the set $[0,1] \times \{N, S\}$, and F_X (G_Y) the distribution of female (male) status conditional on smoking status X (Y, with $X, Y \in \{N, S\}$).

From now on, and for expositional sake, we will refer to the continuous index as socioeconomic status (SES) and to the discrete characteristic as smoking status.

2.1.2 Surplus

We consider a frictionless matching model with TU à la Becker-Shapley-Shubik, in line with recent contributions on similar topics (e.g., Chiappori and Oreffice, 2008; Chiappori, Iyigun and Weiss, 2009; Galichon and Salanié, 2010). In any married couple, the sum of individual utilities is given by some function of the partner's characteristics; as it is customary in this literature, we define the surplus generated by marriage as the difference between this function and the sum of utility levels that each spouse would reach as single. In our framework, the surplus depends on both the discrete and the continuous characteristics of each partner. In order to keep the model tractable, we adopt a simplifying assumption motivated by the smoking interpretation described below. Specifically, we assume that the surplus Σ generated by a match between (x, X) and (y, Y) has the form:

$$\Sigma\left((x, X), (y, Y)\right) = f(x, y) \text{ if } X = Y = N$$
$$\Sigma\left((x, X), (y, Y)\right) = \lambda f(x, y) \text{ otherwise}$$

The function f is strictly increasing, continuously differentiable and supermodular, and satisfies $f(0,0) = 0$; moreover, $\lambda < 1$. In our interpretation, a person who smokes does not mind a partner with the same habit; only a non-smoking individual will perceive smoking as a negative attribute of the potential partner. λ reflects both the distaste for spousal smoking by a non-smoker individual (indirect effect of smoking on the non-smoker) and the direct health effect of smoking in marriage. To be more specific, we can write $\lambda(\mathbf{1}[X \neq Y], X, Y)$, where $\mathbf{1}[X \neq Y] = 1$ if smoking status differs between spouses ($= 0$ otherwise); $X = S$ if the wife is a smoker, $Y = S$ if the husband is a smoker. Without loss of generality, we assume that there are no gender asymmetries in these effects, so $\lambda(1, N, S) = \lambda(1, S, N)$ and that the reduction in surplus due to the distaste effect plus a direct health effect (for one of the spouses) is equal to the reduction in surplus due to two direct health effects when both are smokers, so $\lambda(1, N, S) = \lambda(1, S, N) = \lambda(0, S, S) = \lambda$. Thus, λ represents the decrease in surplus generated by the presence of (at least) a smoker in the couple; note that the surplus of a mixed (smoker-non smoker) couple is the same as that of a couple of smokers, but strictly less than for a non-smoking pair.[6]

2.2 Stable matching

2.2.1 Definition

A matching is defined as a measure μ on the set $([0,1] \times \{N, S\})^2$ and four functions $u_N(x), u_S(x), v_N(y)$ and $v_S(y)$. Intuitively, for two sets $A, B \subset [0,1] \times \{N, S\}$, $\mu[A, B]$ denotes the probability that a woman belonging to A is married with a man belonging to B; and for any female (x, X) (male (y, Y), with $X, Y \in \{N, S\}$), $u_X(x)$ ($v_Y(y)$) is the utility she (he) receives at a stable matching. A constraint on μ is that its marginal should equal the initial distributions of individuals; i.e., the marginal on the set of females (males) is F (G). In addition, on the support of μ, individual utilities satisfy:

$$u_X(x) + v_Y(y) = \Sigma((x, X), (y, Y)), \forall ((x, X), (y, Y)) \in Supp(\mu),$$

[6] The medical literature clearly shows that exposure to secondhand smoke is an important cause of premature death and disease for individuals who do not smoke themselves. However, this secondary effect does not seem to exist for smokers (CDC, 2006; Glymour et al., 2008; Mannino et al., 1997). From a more subjective perspective, a large body of survey evidence around the world shows that the attitudes toward smoking are different between smokers and non-smokers: smokers are less likely than non-smokers to be bothered by secondhand smoke exposure (e.g., Pilkington et al., 2006). In a nutshell, there are significant differences between smokers and non-smokers in the health effects of and the attitudes toward being close to a smoker.

reflecting the fact that if two agents may marry with positive probability, their individual utilities must add up to the surplus they generate when married.

A matching is *stable* if no matched agent would be better off unmatched, and if no two individuals would prefer being matched together to their current situation. Normalizing singles' utility to zero, stability can be summarized by the following set of inequalities: for any $(x, X), (y, Y)$ we have that

$$\begin{aligned} u_X(x) &\geq 0, v_Y(y) \geq 0 \text{ and} \\ u_X(x) + v_Y(y) &\geq \Sigma((x, X), (y, Y)) \end{aligned} \tag{1}$$

therefore

$$\begin{aligned} u_X(x) + v_Y(y) &\geq f(x, y) \text{ if } X = Y = N \\ &\geq \lambda f(x, y) \text{ otherwise} \end{aligned} \tag{2}$$

where an equality obtains on the support of μ. The first constraints in (1) reflect the requirement that married people should prefer marriage to singlehood; the second constraint in (1) expresses that any two individuals cannot, by forming a new match, strictly increase their current utilities.

2.2.2 Existence

Existence of a stable match stems from general results, which states that in a TU context, the minimization of aggregate utility over the set of stable matches is equivalent to the maximization of aggregate surplus over all possible assignments.[7] Formally, if $(\mu, u_N(x), u_S(x), v_N(y), v_S(y))$

[7] See Chiappori, McCann and Nesheim (2010) for a complete presentation.

is a stable matching, then the measure μ solves

$$\max_{\nu\in\mathcal{M}}\int\Sigma\left(\left(x,X\right),\left(y,Y\right)\right)d\nu\left(\left(x,X\right),\left(y,Y\right)\right) \tag{3}$$

where $\mathcal{M}$ denotes the set of measures on the set $\left([0,1]\times\{N,S\}\right)^2$ whose marginal distributions coincide with the initial measures F and G on the female and male populations, respectively. Since this set is compact and Σ is continuous in x and y, a solution exists. Conversely, for any solution $\bar{\mu}$ to the surplus maximization problem, consider the dual program:

$$\begin{aligned}&\min_{u_N,u_S,v_N,v_S}\int_{[0,1]\times\{N,S\}}\left(\mathbf{1}\left[X=S\right]u_S\left(x\right)+\mathbf{1}\left[X=N\right]u_N\left(x\right)\right)dF\left(x,X\right)\\&+\int_{[0,1]\times\{N,S\}}\left(\mathbf{1}\left[Y=S\right]v_S\left(y\right)+\mathbf{1}\left[Y=N\right]v_N\left(y\right)\right)dG\left(y,Y\right)\end{aligned}$$

under the constraints in (1). If $\left(\bar{u}_N,\bar{u}_S,\bar{v}_N,\bar{v}_S\right)$ denotes a solution, then $\left(\bar{\mu},\bar{u}_N,\bar{u}_S,\bar{v}_N,\bar{v}_S\right)$ defines a stable matching.

2.2.3 Purity

We now consider purity. The matching is *pure* when the support of the measure μ is borne by the graph of a function $\rho:[0,1]\times\{N,S\}\rightarrow[0,1]\times\{N,S\}$, so that almost all agents (x,X) are matched with probability one to exactly one agent $(y,Y)=\rho\left(x,X\right)$. In other words, purity forbids matches involving 'mixed strategies', whereby an open set of agents are each matched to several agents with positive probabilities. In a one-dimensional setting, the graph of the function ρ, which maps $[0,1]$ to itself, must be one to one; if it is continuous, it can only be monotonic, and we obtain the standard (positive or negative) assortativeness property. The notion of purity thus generalizes assortativeness to a general setting of multidimensional matching.

To prove purity (or assortativeness), the standard approach, in the one-dimensional case, relies on supermodularity. In a differentiable setting, supermodularity requires that the partial of the surplus function vis a vis one spouse's attribute be strictly injective (therefore monotonic) in the other person's attribute. In a multidimensional setting, the natural generalization of supermodularity is the 'twisted' condition[8], which is sufficient to prove purity of the stable match. The 'twisted' condition states that there exists a set S of measure zero such that for each distinct pair $((y_1, Y_1), (y_2, Y_2))$, any critical points of the function $(x, X) \longrightarrow \Sigma((x, X), (y_1, Y_1)) - \Sigma((x, X), (y_2, Y_2))$ lie in S. In our specific context, this would require that for almost all x_0, the partials of the surplus with respect to the SES x, computed at two points $((x_0, X), (y_1, Y_1))$ and $((x_0, X), (y_2, Y_2))$, cannot be equal unless $(y_1, Y_1) = (y_2, Y_2)$. One can easily check that this property may not hold for the model just described. If a woman with SES x_0 is a non-smoker, the partial of the surplus with respect to x is $\partial f(x_0, y_1)/\partial x$ if she marries a non-smoker with SES y_1, and $\lambda \partial f(x_0, y_2)/\partial x$ if she is matched with a smoker with SES y_2. While f is strictly supermodular, we may still have that:

$$\frac{\partial f}{\partial x}(x_0, y_1) = \lambda \frac{\partial f}{\partial x}(x_0, y_2)$$

with $y_2 > y_1$ since $\lambda < 1$. It follows that the stable matching may not be pure in our setting; indeed, we will show below that, in the quadratic case, it is not.

2.3 The symmetric case

We first start with a particular case in which genders are exactly symmetric; that is, we assume that (i) the surplus function is symmetric ($f(x, y) = f(y, x)$), and (ii) the distributions of characteristics are identical for men and women ($F = G$). Then the stable matching can easily be characterized.

Proposition 1 *Under the symmetry assumptions (i) and (ii) above, there exists a unique stable matching, which is completely assortative. Namely:*

[8] See for instance Chiappori, McCann and Nesheim (2010).

- *Smokers (almost) always marry smokers, and non-smokers (almost) always marry non-smokers.*
- *In each couple, agents (almost) always have the same SES. In particular, within each subpopulation, matching is assortative on SES.*

Proof. *Positive assortativeness within each smoking category directly follows from supermodularity. We simply need to show that it cannot be the case that an open set of non-smokers of one gender (say males) marry smokers. Assume it is; then an equal measure open set of non-smokers of the opposite gender also marry smokers. But since $\lambda < 1$, the total surplus is then less than in the completely assortative matching, a contradiction.* ■

While the symmetric case is obviously very specific, it constitutes an interesting benchmark. A first lesson that can be drawn from it is that two-dimensional matching of the type under consideration naturally leads to discriminated outcomes. In the symmetric context, even if the loss incurred when a non-smoker marries a smoker is very small (i.e., λ is very close to one), the marriage patterns exhibit complete discrimination, in the sense that smokers exclusively marry smokers and non-smokers exclusively marry non-smokers. In other words, minor differences in preferences may have a spectacular impact on marital patterns.

This conclusion, however, heavily relies on the very specific features of the framework under consideration. In particular, because of the assumed symmetry of the SES distributions across genders, no trade-off exists between the two characteristics at the stable match: there is no point, for a non-smoker, in accepting a smoker as a spouse, since a non-smoker with exactly the same SES is always available at the equilibrium.

Lastly, we can compute the corresponding utilities. Assume Ms. x (a non-smoker) marries Mr. y (also a non-smoker) at the stable match; note that $x = y$ by the previous Proposition. Let $u_N(x)$ (resp. $v_N(y)$) denote her (his) utility. Then, by stability

$$u_N(x) = \max_s f(x, s) - v_N(s)$$

where the maximum is reached for $s = y$. By the envelope theorem:

$$u'_N(x) = \frac{\partial}{\partial x} f(x, y)$$

where the right-hand side derivative is taken at the point (x, x). It follows that

$$u_N(x) = \int_0^x \frac{\partial}{\partial x} f(s, s)\, ds + K$$

where K is a constant. Symmetrically,

$$v_N(y) = \int_0^y \frac{\partial}{\partial y} f(s, s)\, ds + K'$$

and

$$\begin{aligned} u_N(x) + v_N(x) &= \int_0^x \frac{\partial}{\partial x} f(s, s)\, ds + K + \int_0^x \frac{\partial}{\partial y} f(s, s)\, ds + K' \\ &= 2\int_0^x \frac{\partial}{\partial x} f(s, s)\, ds + K + K' \\ &= f(x, x) \end{aligned}$$

which gives

$$K + K' = f(x, x) - 2\int_0^x \frac{\partial}{\partial x} f(s, s)\, ds = 0$$

By symmetry, one can assume $K = K' = 0$ and

$$u_N(x) = \frac{f(x, x)}{2}, v_N(y) = \frac{f(y, y)}{2}$$

By the same token,

$$u_S(x) = \lambda \frac{f(x, x)}{2}, v_S(y) = \lambda \frac{f(y, y)}{2}$$

and the utility loss due to smoking is proportional to the size of the surplus. Note, however, that this result is highly specific to the symmetric case, as it will become clear in the quadratic example below.

2.4 The general case: preliminary results

We now come back to the general case. Since the stable matching needs not be pure, we introduce additional notations. In what follows, let $p_N(x)$ be the probability that a non-smoking woman with SES x marries a smoker (then $1 - p_N(x)$ is the probability she marries a non-smoker). We define similarly $p_S(x')$, $q_N(y)$ and $q_S(y')$ as the probability of marrying a smoker for a female smoker, a male non-smoker and a male smoker, respectively.

2.4.1 Assortative matching within smoking categories

We first provide some qualitative properties of the equilibrium, which hold true irrespective of the exact distribution of smokers and non-smokers in the population and the form of the (supermodular) surplus. A first result expresses the fact that, at a stable matching, matching is positive assortative on SES within each smoking category:

Proposition 2 *Consider two matched couples,* $(x, X), (y, Y)$ *and* $(x', X), (y', Y)$ *with identical smoking status. For almost all such couples,* $x \geq x'$ *if and only if* $y \geq y'$*.*

Proof. *Assume, for instance, that* $x \geq x'$ *but* $y < y'$ *on a subset of positive measure. The surplus generated by any two such couples is*

$$\Sigma_1 = \Sigma((x, X), (y, Y)) + \Sigma((x', X), (y', Y))$$

while the matching $(x, X), (y', Y)$ *and* $(x', X), (y, Y)$ *would generate a surplus*

$$\Sigma_2 = \Sigma((x, X), (y', Y)) + \Sigma((x', X), (y, Y)) > \Sigma_1$$

by strict supermodularity of Σ in (x, y). Should this situation exist for a non-null set of couples, the matching would not maximize total surplus, a contradiction. ■

2.4.2 No randomization for high SES

The second result states that among individuals with high SES, matching is assortative on both SES and smoking status. Specifically:

Proposition 3 *Assume that the upper bound of the support of the measures F_S, F_N, G_S and G_N is 1. Then there exist thresholds x_N, x_S, y_N and y_S in $[0, 1)$ such that for almost all $x_N \leq x \leq 1, x_S \leq x' \leq 1, y_N \leq y \leq 1$ and $y_S \leq y' \leq 1$,*

$$p_N(x) = q_N(y) = 0 \text{ and } p_S(x') = q_S(y') = 1$$

Proof. *See the Appendix* ■

Proposition 3 states that non-smokers with high SES marry non-smokers with probability 1: marrying a smoker would decrease the surplus by a factor λ, which for high SES can only decrease total surplus.

2.4.3 Why randomization is possible for low SES

However, the previous result is only true at the top of the SES distribution; further down, randomization may appear at a stable matching. To see why, assume that the distributions of the male and female populations are such that men are much more likely to smoke than women. The assortative pattern described in Proposition 2, together with the measure restrictions, imply that both non-smoking wives and smoking husbands, being on the long side of the market, will have to marry 'down' (i.e., a spouse with relatively low SES). At some point, the marginal non-smoking wife may become indifferent between marrying a non-smoker or the marginal smoking husband, because the resulting loss in total surplus (due to $\lambda < 1$) is exactly offset by the high SES of the latter.

This logic suggests, however, that in a given neighborhood, all forms of randomization are not simultaneously possible. In our example, for instance, while female non-smokers may want to marry a smoking spouse, male non-smokers would not, since they would lose a share λ of the surplus *and* the opportunity to marry 'up'.

This feature is indeed general, as expressed by the following result.

Proposition 4 *Assume there exists an open set O such that, for all $x \in O$, $0 < p_N(x) < 1$ – so that x marries either a non-smoker y or a smoker y' with positive probability. Assume moreover that $q_S(y') > 0$ – so that y' also marries a smoker x' with positive probability. Then $q_N(y) = 0$ and $p_S(x') = 1$ almost surely. Moreover, $x' = x$ and $y' > y$.*

Similarly, if for all y in some open set O', $0 < q_N(y) < 1$ – so that y marries either a non-smoker x or a smoker x' with positive probability – and $p_S(x') > 0$ – so that x' also marries a smoker y' with positive probability, then $p_N(x) = 0$ and $q_S(y') = 1$ almost surely. Moreover, $x' > x$ and $y' = y$.

Proof. *See the Appendix* ■

Proposition 4 states that the various types of randomizations are mutually exclusive. In the neighborhood of some given SES, it may be the case that female non-smokers and male smokers intermarry with positive probability; but then, in this same neighborhood, female smokers and male non-smokers only marry their own type. Of course, the pattern may be opposite in a different neighborhood;[9] ultimately, the matching patterns depend on the distributions F and G.

[9] Examples can readily be constructed using disconnected subpopulations among both men and women.

2.5 The general case: surplus maximization

We now describe the mathematical form of the problem, and indicate a resolution strategy that works for arbitrary distributions.

2.5.1 Conditions on the measure

Start with the constraint that the marginals of the stable measure μ must coincide with the female and male population measures. Proposition 2 greatly simplifies the expression of these constraints. Indeed:

- Consider a non-smoking female with SES x, matched with a non-smoking male with SES $y = \phi_N(x)$. Then the number of non-smoking females with SES higher than x, married with a non-smoker, must equal the number of non-smoking males with SES higher than y, married with a non-smoker:

$$\int_x^1 (1 - p_N(t))\, dF_N(t) = \int_{\phi_N(x)}^1 (1 - q_N(s))\, dG_N(s) \tag{4}$$

  which defines the function ϕ_N.

- Similarly, for a non-smoking female with SES x marrying a smoker with SES $y' = \psi_N(x)$:

$$\int_x^1 p_N(t)\, dF_N(t) = \int_{\psi_N(x)}^1 (1 - q_S(s))\, dG_S(s) \tag{5}$$

  which defines ψ_N.

- For the other combinations of smoking status:

$$\int_x^1 (1 - p_S(t))\, dF_S(t) = \int_{\phi_S(x)}^1 q_N(s)\, dG_N(s) \tag{6}$$

$$\int_x^1 p_S(t)\, dF_S(t) = \int_{\psi_S(x)}^1 q_S(s)\, dG_S(s) \tag{7}$$

In particular, there exists a one-to-one relationship between the four probability functions (p_N, p_S, q_N, q_S) and the four matching functions $(\phi_N, \phi_S, \psi_N, \psi_S)$.

2.5.2 The surplus maximization program

Clearly, the exact marital patterns characterizing the stable matching depend on the joint distribution of SES and smoking status of the two populations. A specific case will be explicitly studied in the next section. Here, we just provide the general tool that can be used to solve the problem with arbitrary distributions. The idea, again, is to exploit the duality between stability and surplus maximization. With the previous notations, aggregate surplus is:

$$S = \int_0^1 \left[(1 - p_N(t)) f(t, \phi_N(t)) + \lambda p_N(t) f(t, \psi_N(t)) \right] dF_N(t) + \lambda \int_0^1 \left[(1 - p_S(t)) f(t, \phi_S(t)) + p_S(t) f(t, \psi_S(t)) \right] dF_S(t) \tag{8}$$

The first integral considers the contribution of the female non-smoking population. An individual with SES t may (with probability $p_N(t)$) be matched with a smoker with SES $\psi_N(t)$, in which case the couple generates a surplus $\lambda f(t, \psi_N(t))$; alternatively, she may (with probability $1 - p_N(t)$) be matched with a non-smoker with SES $\phi_N(t)$, generating a surplus $f(t, \phi_N(t))$. Similarly, the second integral represents the contribution of the female smoking population to total surplus.

A stable matching, defined by the functions p_N, p_S, q_N, q_S and $\phi_N, \phi_S, \psi_N, \psi_S$, linked by (4) to (7), maximizes aggregate surplus under the constraints $0 \leq p_A(t) \leq 1, 0 \leq q_A(t) \leq 1, 0 \leq \phi_A(t) \leq 1, 0 \leq \psi_A(t) \leq 1$, where $A = N, S$. The stable matching can therefore be derived as a solution to a maximization (optimal control) problem.

2.5.3 Computing individual utilities

Finally, once the functions p_N, p_S, q_N, q_S (or equivalently $\phi_N, \phi_S, \psi_N, \psi_S$), which define the stable matching, have been computed, one can readily recover the intracouple allocation of the surplus. Indeed, consider for instance a non-smoking wife with SES x. If her husband is a non-smoker with SES $\phi_N(x)$, then stability implies that:

$$u_N(x) = \max_y \left(f(x,y) - v_N(y)\right)$$

the maximum being reached for $y = \phi(x)$. It follows, from the envelope theorem, that

$$u'_N(x) = \frac{\partial}{\partial x} f(x,y)$$

where the right-hand side derivative is taken at the point $(x, \phi_N(x))$. Then:

$$u_N(x) = \int_0^x \frac{\partial}{\partial x} f(s, \phi_N(s))\, ds + K$$

where K is a constant; and a similar expression obtains for the other utilities. The various utilities are therefore defined up to an additive constant each; the constants, in turn, are pinned down by the adding up property on the support of μ and the indifference conditions. The next section provides an explicit example of such computations.

3 A quadratic example

3.1 Specification

As stated above, the exact form of the stable matching depends on the two distributions and on the surplus function. We now consider a simple example based on two assumptions. One is independence: for both males and females, SES and smoking status are independently distributed. Therefore, for each gender, the proportion of smokers is the same at each SES level. In what follows, let s_M (resp. s_W) denote the percentage of smokers in the male (female) population. Then, for any SES x (resp. y), a female (male) has a probability s_W (s_M) of being a smoker. If $s_W = s_M$, we are in the symmetric case discussed above. We concentrate in what follows on the more interesting situation in which $s_M \neq s_W$ (then we may, without loss of theoretical generality, consider the empirically relevance case $s_M > s_W$); otherwise we maintain a perfect symmetry between genders in terms of SES.

The second assumption posits that the gain generated by marriage are quadratic. Specifically, we assume that:

$$f(x, y) = xy$$

While this form can be given micro foundations[10], we use it mostly for practical purposes: it allows the derivation of a closed form solution. Needless to say, the techniques described in the Appendix apply to any functional form, although more complex functions may lead to tedious calculations.

3.2 Formal characterization of the stable matching

We now show that it is possible, in this simple example, to completely solve the matching model in closed form. Regarding smoking status, four categories of couples (at most) may appear: N-N, S-S, N-S, and S-N. Within these categories, per Proposition 2, supermodularity implies that matching will be assortative; i.e., men with higher SES will marry wives with higher SES.

[10] This form is used for instance in Browning, Chiappori and Weiss (2012). In their framework, x and y are individual incomes, and agents utilities are Cobb-Douglas with one private and one public good − i.e., $u_i(q_i, Q) = q_i Q$, where q_i is the private consumption of member i, and Q is the public consumption of the household. These functions are of the Generalized Quasi Linear (GQL) form of Bergstrom and Cornes (1983); they satisfy the transferable utility property, and any efficient allocation satisfies $Q = q_m + q_w = (x + y)/2$, leading to a surplus equal to xy.

3.2.1 Marital patterns

Our main result is the following:

Proposition 5 *There exists a unique stable matching, and there exists two numbers $\widetilde{x}$ and p, both between 0 and 1, such that the unique stable matching has the following features:*

- *All agents marry*
- *For all $x \geq \widetilde{x}$, a non-smoking woman with SES x is matched with probability 1 to a non-smoking husband with SES:*

$$y = \frac{1 - s_W}{1 - s_M} x - \frac{s_M - s_W}{1 - s_M}$$

  and a smoking woman with SES x is matched with probability 1 to a smoking husband with SES:

$$y' = \frac{s_W}{s_M} x + \frac{s_M - s_W}{s_M}$$

  In particular, $y < x < y'$: smoking men and non-smoking women marry 'down', whereas non-smoking men and smoking women marry 'up'.

- *For $x < \widetilde{x}$, a non-smoking woman with SES x is matched:*
  - *with probability p, to a smoking husband with SES:*

$$y' = \frac{1}{\lambda + s_M - \lambda s_M} x$$

  - *with probability $1 - p$, to a non-smoking husband with SES:*

$$y = \frac{\lambda}{\lambda + s_M - \lambda s_M} x$$

In particular, conditional on the SES x of the wife, smoking husbands have a higher SES than non-smoking ones – i.e., $y' > y$.

- *For $x < \widetilde{x}$, a smoking woman with SES x is matched with probability* 1 *to a smoking husband with SES:*

$$y' = \frac{1}{\lambda + s_M - \lambda s_M} x$$

- *Finally, there are no couples in which the wife smokes and the husband does not.*

In the Appendix, two proofs of this result are provided, one using the surplus optimization program and the other relying on a direct approach.

3.2.2 Interpretation

The stable matching is summarized in Figure 1. It can be interpreted as follows. First, in the line of Proposition 3, high SES non-smoking women tend to marry high SES non-smoking men, and high SES smoking women tend to marry high SES smoking men. Such a matching is stable because, for a given SES, a non-smoking person views a smoking potential partner as an inferior substitute for a non-smoking one, whereas a smoking person would view them as equivalent. Among these couples, assortative matching requires that, for any couple (x, y), the number of women with a SES above x be equal to the number of men with a SES larger than y. Since non-smoking women outnumber non-smoking men, non-smoking men and smoking women marry 'up', whereas conversely smoking men and non-smoking women marry 'down'. Below the threshold $\widetilde{x}$, however, the stable match involves randomization. Non-smoking women may be married with either a smoker or a non-smoker; the respective SES, y' and y, of these potential spouses, depends on the probability p (through the measure condition), and are such that Ms. x is indifferent between them – which pins down p. Smoking women, on the other hand, only marry smokers, as expected from Proposition 4. Lastly, and still following Proposition 4, if two non-smoking women with equal SES are married with a smoker and a non-smoker respectively, the smoker has a higher SES than the non-smoker; however, if two male smokers with equal SES are married with a smoker and a non-smoker, respectively, the two wives also have the same SES.

Insert Figure 1

The same patterns can equivalently be described using the husband's perspective. They can then be summarized as follows:

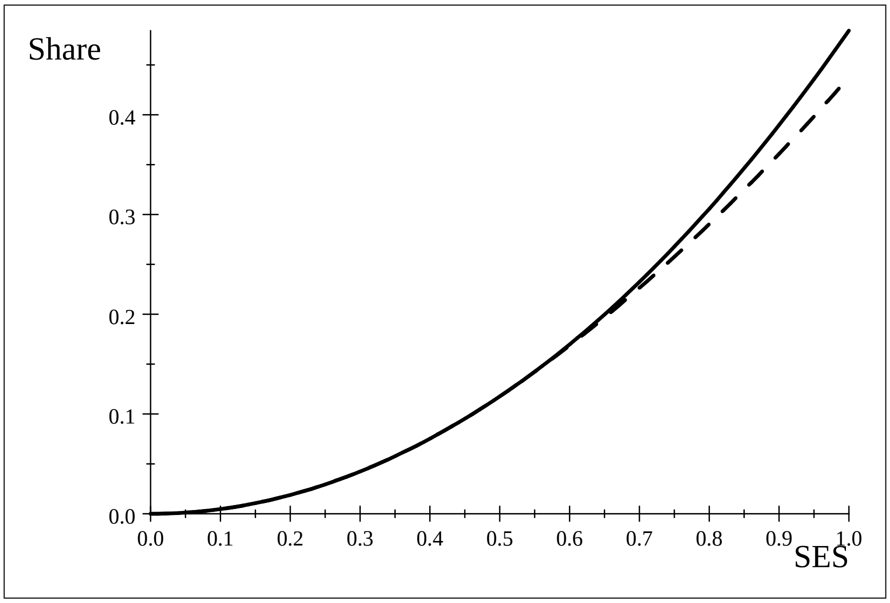
- Non-smoking husbands always marry a non-smoking wife with probability 1
- Smoking husbands with a high SES marry a high SES smoking wife with probability 1
- Smoking husbands with a low SES marry either a smoking or a non-smoking wife with positive probability; moreover, both potential wives have the same SES.

3.2.3 Surplus allocation

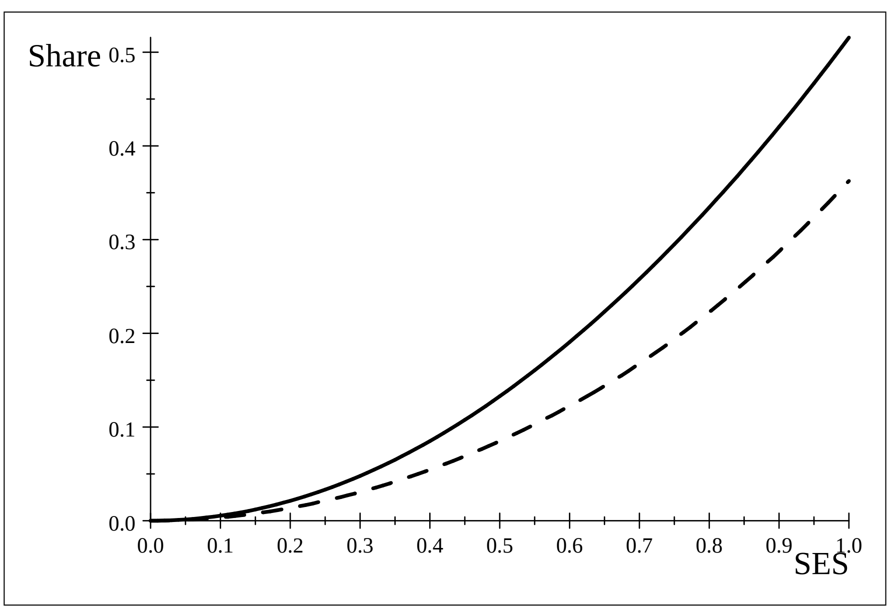
A by-product of the result is the derivation of the intrahousehold allocation of resources implied by equilibrium conditions. In our case, this allocation is exactly pinned down.[11] A precise characterization is given in the Appendix. Figures 2a and 2b represent these allocations as a function of the SES for females and males respectively, for $s_M = .25, s_W = .2$ and $\lambda = .8$; in both graphs, dashed lines correspond to smokers. As expected, both shares are increasing with SES. Comparing individuals with different smoking status, we find that the share of a male smoker is always smaller than for a non-smoker. For wives, however, the respective shares of smokers and non-smokers are identical below the threshold $\widetilde{x}$. Non-smoking women with SES lower than $\widetilde{x}$ marry smoking husbands with positive probability; since, from a smoking

[11] See Browning, Chiappori and Weiss (2012, ch. 8) for a detailed presentation of the technique used below.

husband's perspective, a smoking wife with identical SES is a perfect substitute, smoking and non-smoking women must have the same utility. Also, note that, despite the strict symmetry assumed for the male and female distribution of SES, the cost of being a smoker (in terms of surplus shares) is always much smaller for women than for men, reflecting the advantageous position of smoking women on the short side of the market.



Female share of the surplus (dashed for smokers)



Male share of the surplus (dashed for smokers)

3.3 Comparative statics

One advantage of closed form solutions is the possibility to derive comparative statics properties. In our case, these can be summarized as follows:

- A larger λ, by reducing the welfare cost of smoking, increases the threshold $\widetilde{x}$, and benefits male and female smokers, but also female non-smokers (who may marry male smokers and experience a smaller welfare loss); however, it hurts male non-smokers by reducing their comparative advantage.

- Increasing the proportion of male smokers s_M also increases the threshold. It only hurts male smokers by further deteriorating their position on the long side, but benefits all other agents.

- Finally, an increase in s_W has the opposite impact: it benefits male smokers and hurts all other agents.

3.4 Practical implementation

In practice, the frictionless process described in the model is never observed. Marriage markets are not frictionless; moreover, actual matching involves multidimensional characteristics, some of which may actually be unobserved by the econometrician (a direction followed for instance by Chiappori, Oreffice and Quintana-Domeque, 2012, and Galichon and Salanié, 2010), and may furthermore be affected by random shocks à la Shimer and Smith (2000). For all these reasons, observed matching patterns are largely stochastic. Still, the previous analysis suggests that these stochastic patterns should exhibit specific features due to the underlying competitive structure.[12] Specifically, we expect the following regularities to hold:

Prediction 1: Mixed couples in which the wife smokes while the husband does not (denoted S-N) should be less frequent than those in which he smokes and she does not (denoted

[12] One possible justification would involve a 'rank order' property à la Fox (2010).

N-S); more precisely, the ratio of S-N to N-S couples should be smaller than implied by the sole difference in relative smoking prevalence, i.e., than the ratio

$$r = \frac{s_W (1 - s_M)}{s_M (1 - s_W)}$$

In practice, in our sample, s_M is .22 and s_W is .17, so r is around .71; we expect the observed ratio to be significantly smaller than this threshold.

Prediction 2: Among couples with identical smoking habits (i.e., both smokers, denoted S-S, and both non-smokers, denoted N-N), matching should be assortative on SES.

Prediction 3: Non-smoking wives married with a smoking husband should have a *lower* SES than those married with a non-smoking husband; the same should hold for a smoking husband married with a non-smoking wife. That is, a smoking spouse is negatively correlated with SES for non-smoking women. For men, however, the opposite logic prevails; i.e., it is now a *non-smoking* spouse that is negatively correlated with SES for smoking men.

Prediction 4: When two non-smoking women with the same (low) SES marry respectively a smoker and a non-smoker, the non-smoker should on average be of lower SES than the smoker. That is, controlling for the wife's SES, the smoking habit of the husband should be *positively* correlated with his SES. This prediction is obviously specific to our simplified framework in which smoking prevalence is orthogonal to education. In practice, low SES is a powerful determinant of smoking behavior (e.g., CDC, 2010). We restate then the prediction as follows: the *conditional* correlation between male SES and smoking status, *given the SES of non-smoking wives,* should be less negative than the unconditional one. In addition, this pattern should not hold true for women.

The remaining part of the paper is devoted to testing these predictions.

4 Empirical Application

4.1 Smoking and education: basic facts

Cigarette smoking remains the Nation's leading cause of premature, preventable death; during 2000–2004, approximately 443,000 premature deaths in the United States each year were attributed to cigarette smoking (CDC, 2008). Smoking causes deaths from heart disease, stroke, lung and other types of cancer, and chronic lung diseases. In addition, exposure to secondhand smoke is considered an important cause of premature death and disease for individuals who do not smoke themselves, while this secondary effect does not seem to exist for smokers (CDC, 2006; Glymour et al., 2008; Mannino et al., 1997). From a more subjective perspective, a large body of survey evidence around the world shows that the attitudes toward smoking are different between smokers and non-smokers: smokers are less likely than non-smokers to be bothered by secondhand smoke exposure (e.g., Pilkington et al., 2006). In a nutshell, there are significant differences between smokers and non-smokers in the health effects of and the attitudes toward being close to a smoker.

A second feature that will be important in our approach is the asymmetry in the smoking prevalence across genders. In the US, and actually in many countries, male smokers largely outnumber female smokers, a discrepancy that has remained stable over the last decades. This gender asymmetry has been emphasized by the Surgeon General (e.g., Surgeon General Report, 2001), as well as by several studies in various fields, e.g., Gruber (2001) in economics and Öberg et al. (2011) in medicine. In 2007, in the United States, 26.5% of white men 18–24 years of age and 21.6% of white women 18–24 years of age were current cigarette smokers (NCHS, 2010); the prevalence of smokers among white men 25–34 years of age was 29.0% while it was 21.4% among white women of the same age (NCHS, 2010). This situation creates an interesting case for our matching model, since some men (the non-smokers) and some women (the smokers) are on the short side of the market.

Finally, a key role is also played by educational attainment, which is not only an important component of the 'quality' relevant in marital sorting and a standard measure of SES, but is also closely related to cigarette use. This *cigarette connection* is acknowledged by economists at least since the seminal work by Farrell and Fuchs (1982), who document a negative smoking gradient by this SES indicator (see also Gruber, 2001). Recently, De Walque (2010), using retrospective smoking histories constructed from the smoking supplements of the National Health Interview Surveys conducted between 1978 and 2000, shows that at least among women, college education has a negative effect on smoking prevalence, and that more educated individuals responded faster to the diffusion of information on the dangers of smoking after 1950. Note, however, that the gender gap in smoking prevalence is maintained across all education levels. In 2007, the (age-adjusted) prevalence of smokers by educational level among white men and women 25–years of age and over were as follows: 30.8% vs. 23.9% for those with less than high-school; 29.9% vs. 25.2% for those with high-school; 21.8% vs. 19.6% for those with some college; and 10.5% vs. 8.2% for those with college or above (NCHS, 2010).

4.2 Data Description

Estimations are based on the US Current Population Survey (CPS) data, the annual March CPS supplements and the Tobacco Use Supplements (TUS), for the years 1996 to 2007, which provide the most recent and largest samples of married couples for whom information on tobacco use is available.[13] The standard demographic and education variables are extracted from the annual March CPS supplements, to which data on smoking status are merged from the TUS. The TUS are monthly CPS supplements available discontinuously over time and in different months. Specifically, the available TUS of interest are January and May 1996, 1999, 2000; June 2001; February 2002; February and June 2003; May 2006; January 2007.

[13] Although the March CPS data are available up to 2012, the latest TUS is in 2007.

The CPS is a series of monthly cross sections, with a short longitudinal component. Individuals in the sample are interviewed eight times–four times, followed by a break of eight months, and then interviewed for the same four months the following year. As such, it is possible to match observations of the same individuals across months, using the household and person identification codes, along with the month-in-sample information. However, several observations are dropped due to the specific design of the rotation samples by 4-month periods. In addition, we also check for age, gender and race, to ascertain that the merged observations consistently belong to the same individual.[14]

The TUS-CPS is a National Cancer Institute (NCI)-sponsored survey of tobacco use and policy information that has been administered as part of the Current Population Survey (CPS) since 1992.[15] It is considered a key and reliable source of national, state, and sub-state level data on smoking and other tobacco use in US households, which is widely used in medical research on cancer and other consequences of smoking (e.g., Delnevo and Bauer, 2009; Mills, Messer, Gilpin, Pierce, 2009). It provides data on a nationally representative sample of about 240,000 civilian, non-institutionalized individuals ages 15 years and older.

We are able to match individuals across months, merging all these TUS supplements back to the March supplement of the corresponding year, to build a series of repeated cross-sections for the years 1996, 1999, 2000, 2001, 2002, 2003, 2006, and 2007. Due to the CPS rotation sample design described above, the sample size of each match is at most $\frac{1}{4}$, $\frac{1}{2}$, or 3/4, of the original March sample size (when matched to June, January and May, or February, respectively). In general, the farther from March the TUS supplement month is, the fewer observations can be matched, with the strong restriction that the TUS months of September (1992, 1995, 1998), November (2001 and 2003), and August 2006 cannot be merged back to March, as they do not share any respondent (Madrian and Lefgren, 1999). Nevertheless, our

[14] Madrian and Lefgren (1999) illustrate and explain the matching procedures to longitudinally merge the CPS respondents.

[15] The U.S. Centers for Disease Control and Prevention (CDC) co-sponsored the TUS-CPS with NCI between 2001 and 2007.

sample represents the most recent and largest sample of spouses, with detailed socioeconomic and smoking information, and to the best of our knowledge, it is the first time it is used to study marriage and smoking.

We specifically extract husbands and wives from one-family households from our merged CPS files. Married individual records of the reference person and her spouse are then matched on the household identification code (and household number) to create a single observation for each couple, keeping only observations of couples who lived in households with only one family.

Our main sample of husbands and wives consists of white couples, where the wife is between 22 and 32 years old and the husband is between 24 and 34 years old. This demographic group allows us to focus on *recently* married couples, as the sorting by smoking status and education arise in the marriage market at the time of the match. In the US the median age at first marriage is 27 for men and 25 for women (US Census Bureau, 1999-2003). On the other hand, a lower bound of 22 and 24 years old also allows us to include college graduates after they have completed their schooling. The additional two years in the husbands' bounds are based on the standard median / mean age difference of two years between male and female spouse (Chiappori, Iyigun and Weiss, 2009). Note that the March CPS does not record the duration of marriage; in particular, the June Fertility Supplements that used to provide the age at (first) marriage, do not contain it any longer in the most recent years that our study is concerned about.

In addition to individual age, we use the state of residence, year of interview, sample household weight and education of the individual. From 1992, the CPS records education as degrees attained rather than years of schooling completed. We thus assign the number of years of schooling to the corresponding degrees.[16] March CPS household weights are used to make our sample of couples representative of the US population.

From the Tobacco Use Supplement, we retrieve information on the smoking status of

[16] In the sensitivity analysis subsection, we will recode the education variable following Jaeger (1997).

each individual. Specifically, the respondents are asked whether and how often they smoke, whether they have smoked at least 100 cigarettes in their lifetime, and the actual number of cigarettes they smoke.[17] From the first two questions, we construct a dummy variable of smoking status, defining a person as a smoker if she reports to smoke every day or some days, and has smoked at least 100 cigarettes in her lifetime, and as non-smokers those who say that they never smoke, or those who have smoked less than 100 cigarettes in their lifetime (CDC, NCHS, 2010). Finally, note that each spouse directly *reports* his/her information, and self-reporting of smoking status is considered a reliable source of information, as it is found to be validated by measured serum cotinine levels (Caraballo, Giovino, Pechacek, Mowery, 2001).

4.3 A first look at the data

The main characteristics of the data are described in Tables 1-4. We present the summary statistics of married and single individuals, and the corresponding ones by smoking type of couple, the observed matching patterns by smoking status, and the correlations of smoking status and education by gender and marital status. A preliminary look at the data confirms that the smoking prevalence is higher for men than for women, with 22% of husbands smoking versus 17 % of wives (25 and 21 % for never-married, respectively), consistently with the gender gap reported by the National Center for Health Statistics (NCHS, 2010). Tables 1 and 2 also show that women are more educated than men across all smoking categories. The health status is very similar across spouses, and higher in couples where none is smoking than in those where both spouses are smoking. The average number of children under six years old is about 0.80 for couples and 0.50 for never-married individuals.

Panel A in Table 3 reports the observed matching by smoking status for husbands and wives. There is strong assortative mating by smoking status: about 72% of couples have

[17] The 2000 TUS Supplements do not record the question about the number of cigarettes smoked.

non-smoking spouses, and 10% consists of smokers. This is in line with evidence on marital sorting by smoking status in the UK (Clark and Etilé, 2006). However, a new and particularly interesting insight is provided by looking at "mixed" couples where one spouse is a smoker and the other one is not. Our data reveal that there are fewer "mixed" couples where the wife smokes than vice-versa, 6.50% versus 11.71%, so the ratio is 0.55 (s.e.=0.029), which is statistically significantly lower than the 0.71 (s.e.=0.021) implied by the sole difference in relative smoking prevalence.[18] Indeed, these figures are very different from the prevalences arising from "random matching" (Panel B). This finding supports *prediction 1.* In our theoretical framework, the matching of a smoking man to a non-smoking woman happens because of the shortage of smoking women relatively to smoking men, given that a smoking man would prefer a smoking spouse. At the same time, the opposite match of a smoking woman to a non-smoking man would be far less frequent, given that all smoking women, who are in short supply, would end up marrying a smoking man.

Regarding the correlation of education and smoking, Table 4 summarizes some clear patterns. We first note that both men and women exhibit a negative significant correlation between their smoking status and education. A second conclusion is that these correlations appear different by gender, with the male gradient being significantly larger than the female one. These findings are in line with the literature on smoking and education (see De Walque, 2010). In addition, Table 4 shows that these patterns are present for both married and single individuals.

All in all, these tables are consistent with the basic story presented above. Assortative matching takes place by smoking status, with strong positive sorting and an interesting pattern among the "mixed" couples, given the higher smoking prevalence of men than women. The following Section will test the remaining predictions.

[18] Standard errors are computed using the delta method. The difference in the ratios is statistically significant at the 1% (p-value=0.0000).

4.4 Main results

Table 5 presents evidence of sorting by education within smoking types of young couples. In this Table we regress own education on spouse's education controlling for own age, year- and state-fixed effects. For each group of couples, there is assortative mating by education. Although assortative mating by education has been extensively documented in the literature (Lam, 1988; Pencavel, 1998; Qian, 1998; Mare, 2008), here we show that it holds true within each spouses' smoking category, and with magnitudes comparable to the estimated educational sorting in the US (Pencavel, 1998), supporting *prediction 2*. Perhaps more interesting is the fact that our estimates also suggest that there is a gradient in assortative mating: stronger for couples were none of the spouses smoke (0.65), and weaker for those were both spouses smoke (0.45).

Prediction 3 is investigated in Table 6. It contains a series of regressions of young couples in which either wife's or husband's education is the dependent variable and spouse's education and smoking status are the explanatory variables, controlling for own age, year- and state-fixed effects, broken down by own smoking status.

Column (1) shows that among non-smoking wives those with smoking husbands have on average 0.14 fewer years of completed education than those with non-smoking husbands. In other words, a smoking husband marries, on average, a worse non-smoking spouse in terms of education than if he were to be a non-smoker. These male smokers are penalized for their handicap with a lower 'quality' spouse. This suggests that spousal smoking is a bad characteristic for non-smokers, and that there is a marriage market penalty associated to it, in terms of lower socioeconomic standards. However, among smoking wives, column (2) indicates that there is no statistically significant difference in the average years of completed education between those who marry smoking husbands and those who marry non-smoker ones and the coefficient has a much lower magnitude than in column (1). Smoking women mainly marry smoking men, and there is no reason why they should exhibit penalties or compensations.

Columns (3) and (4) display the same type of regressions as in columns (1) and (2), but for husbands. Now, among non-smoking husbands those with non-smoking wives have on average 0.21 more years of completed education than those with smoking wives, column (3). Thus, a non-smoking wife marries, on average, a better non-smoker spouse in terms of education. Another, more interesting confirmation of our model predictions is the finding in column (4): among smoking husbands those who marry smoking wives have on average 0.16 *more* years of completed education than those with non-smoking wives. Indeed, the theoretical analysis shows that, given the shortage of smoking women, smoking men who marry smoking women should be more educated, whereas no such effect should be observed for women. Smoking men prefer to marry smokers, to avoid being penalized by their non-smoking partners, and have to compete for smoking partner with their own education. Note that the sign of the coefficient in column (4) is opposite to the well-known negative gradient between own smoking and own education.

Overall, Table 6 provides support for *prediction 3*, namely: (i) among non-smoking women those who marry smoking men are less educated, and (ii) among smoking men those who marry smoking women are more educated, because there is a shortage of smoking women. Equivalently, since we are controling for spousal education, columns (1) and (4) provide evidence in favor of the additional implications on marrying 'up' or 'down' of our theoretical framework. Specifically, column (1) shows that non-smoking men, marrying non-smoking women, would marry 'up' in socioeconomic terms, given that non-smoking men are in short supply and that non-smoking women prefer to marry non-smokers. Conversely, column (4) shows that it is smoking women, marrying smoking men, who marry 'up' in socioeconomic terms, given that being a smoking wife of a smoking husband increase the husband's education relative to hers.

The evidence presented in Table 6 is also supportive of our assumptions on the surplus reduction due to smoking. If, for instance, smokers also prefered non-smokers then we would not observe the positive coefficient of column (4). By the same token, our empirical evidence also shows that a gender asymmetric perception of smoking cannot be the driving force behind the observed matching patterns. If men, regardless of their smoking status, perceived smoking in a woman as a defect, we would not observe the positive coefficient of column (4) either.

Table 7 focuses on husbands of non-smoking women, and shows that controlling for their wives' education significantly decreases the magnitude of the correlation of male education and smoking status. This supports our last and *fourth prediction*: among non-smoking low-quality women, their smoking husbands are "more" educated. Columns (3) and (4) reinforce our evidence, showing that this pattern does not hold for the wives of smoking men.

All in all, our reduced form analysis supports our predictions and, in particular, our counterintuitive contention that smoking husbands married to smoking women are on average more educated than those smoking men whose wife does not smoke. In addition, these findings offer new insights on the intial step of couple formation, and the role that smoking can have in shaping new families.

4.5 Sensitivity Analysis

4.5.1 Additional controls, alternative definitions, and different age groups

We proceed to a few robustness checks. First, we take into account "unobserved" heterogeneity by adding controls for individual and household characteristics, and including the interactions between state and year fixed effects. Our main results (signs, magnitudes, and significance) are robust to the inclusion of health status and number of children. Specifically, we construct a dummy variable for very healthy status (one if the status is excellent or very good, zero if good, fair or poor), and consider the number of own children in the family who are under age 6, as our analysis concerns young couples. The main regression specifications with these additional controls are reported in Table A1: controlling for number of children and health status does not change our estimates. Second, we relax the definition of smoker, by not considering the criterion based on the 100 cigarettes smoked in a lifetime. Although the number of smokers increases, the patterns of assortative mating by smoking status and the relationships between own education and spouse's smoking remain unchanged. Third, in terms of years of schooling, recoding the education variable following Jaeger (1997) confirms our findings, as shown in Table A2. Finally, we slightly modify the age group under analysis, including younger women whose age is between 20 and 30, and younger men whose age is between 22 and 32, to add younger married couples who are likely to be newly-weds. This sample yields the same patterns of results as our main estimates, as shown in Table A3, which reinforces our claim that the young couples in our sample represent recent marriages and the actual matching in the marriage market.[19]

4.5.2 Alternative data sets

To further explore the issue of recent marriages, we alternatively examine a very different data source which provides the information on duration of marriage, i.e. the Panel Study of Income Dynamics, and use the waves from 1999 to 2007. The PSID, recently used by Chiappori, Oreffice and Quintana-Domeque (2012) and Oreffice and Quintana-Domeque (2010) to study matching patterns of couples along socioeconomic and anthropometric characteristics, is a longitudinal household survey collecting a wide range of individual and household demographic, income, and labor-market variables. In addition, in all the most recent waves, from 1999 to 2007, the PSID provides detailed information on the smoking behavior of both heads and wives, specifically on smoking status and number of cigarettes, which we use to construct the corresponding dummy variable of whether an individual is a smoker. We then rely on the "Marital History File: 1985-2007" Supplement of the PSID to obtain the year of marriage and the number of marriages. Merging these data to the main files by the unique

[19] The information on duration of marriage or age at marriage is not available in the CPS in any of the years under consideration. However, our choice of very young couples along with the very large sample size of this data set allows us to focus on recently married couples, that is on the matches formed on the marriage market, with which our analysis is concerned.

household and person identifiers provides the information on how recently a couple formed.

It is important to acknowledge that the PSID is a very different dataset than the CPS. First, the PSID is a panel, not a cross-section. Second, its sample size is much smaller. Moreover, its effective sample size is, given its panel structure, even smaller. Hence, the PSID and its availability of the relevant information on marriage duration comes with the price of a huge reduction in sample size compared to the CPS. This dramatic reduction is exacerbated in our analysis, which is characterized by sub-dividing the sample according to spouse's smoking status.[20] Additionally, in the PSID all the variables are reported by the head of the household, including the information on the wife. The wives' smoking behavior is therefore proxy-reported by their husbands, while in the CPS it is self-reported.

Nevertheless, we replicate our main results on the positive sorting by smoking status, with the asymmetric prevalence of "mixed" couples (Table A4), and on the relationship between husbands' education and spouse's smoking status, and between non-smoking wives' education and their husbands' smoking status (Table A5). The estimates are much noisier and not statistically significant. However, the signs of the coefficients are the same as in our main CPS estimates, and their magnitudes are similar or higher. Reassuringly, the observed patterns when using PSID data are again consistent with our predictions.[21]

[20]This reduction is present although the age group has been widened to 24-36 (husbands) and 22-34 (wives), with or without the recently married provision.

[21]We could not find other data sets suitable for our study, which could compare to the reliable sample size, and the availability of both spouses' information and of young individuals characterizing the CPS. In fact, few nationally-representative data sets provide the information on smoking behavior, and even fewer provide it for both spouses. For instance, although the National Interview Survey has very detailed information on smoking behavior and health, any information concerning the spouse is absent by data set design. On the other hand, data sets such as the PSID, or its European counterparts, e.g. the BHPS and the GSOEP, provide the information on spouses' smoking but the sample size is relatively small, as they are panel surveys, a feature that does not concern our marriage market analysis. Finally, the Health and Retirement Survey (HRS) allows to construct retrospective data on couples' smoking status but only for older cohorts (Maralani, 2009), given that the HRS sample includes individuals who are 50 years old and above.

5 Conclusions

We devise a matching model where individuals are characterized by multidimensional attributes, one of which discrete. We show that a stable match exists, is unique, and may fail to be pure (i.e., may involve randomization patterns, whereby identical individuals may be matched with several different partners with positive probability). We provide several properties of the stable matching, and describe a systematic way of deriving the solution, based on an optimal control approach. We then specialize further our model by assuming that the surplus function is quadratic. In this context, we explicit derive a closed form characterization of the stable matching. The derivation is performed using two alternative approaches – one relying on optimal control and the other on a direct test of the stability conditions. Deriving closed form solutions has several advantages. One is that comparative statics exercises can be performed in a systematic way. Moreover, and perhaps more interestingly, the resulting allocation of resources within couples can be derived. This, in particular, opens the door to a more general model in which individual traits are *acquired* in some initial period: the corresponding investment decisions depend on the expected outcomes, which our matching model fully characterizes. This will be a topic of future research. Finally, we derive several testable predictions regarding the marital patterns stemming from our framework.

As an empirical illustration, we apply our framework to explore the interaction between smoking status and education at the time of marriage. Specifically, we study the matching between smokers and non-smokers, where smoking decreases the marital surplus in line with the medical and sociological literatures. Given the gender asymmetric smoking prevalence (with more smoking men than smoking women for all education levels), smoking women and non-smoking men are in short supply. We show that at the top of the 'quality' distribution, matching is pure and assortative by SES *and* smoking status; that is, educated non-smoking men marry educated non-smoking women, and educated smoking women marry educated smoking men. Below some 'quality' threshold, however, matching patterns become more complex. While non-smoking men still marry a non-smoking spouse, smoking men may be matched with either a smoker or a non-smoker. Equivalently, the husband of a smoking woman is still a smoker spouse; but a non-smoking wife may be married to either a smoker or a non-smoker. In that case, the smoker is typically of better 'quality' than the non-smoker.

Using March and TUS CPS data on young couples for the period 1996–2007, we show that there is strong sorting by smoking status: there are 71.78% of couples were both spouses are non-smokers, and 10.01% were both smoke. Our data also reveal that there are fewer "mixed" couples where the wife smokes than vice-versa, 6.50% versus 11.71%, that this difference is statistically significant, and that the ratio is 0.55, which is lower than the 0.71 implied by the sole difference in relative smoking prevalence. We also find that among non-smoking wives those with smoking husbands have on average 0.14 fewer years of completed education than those with non-smoking husbands. Moreover, we find that among smoking husbands those who marry smoking wives have on average 0.16 more years of completed education than those with non-smoking wives.

All in all, the main message of our paper is twofold. One is that multidimensional models of matching, although intrinsically more complex than unidimensional ones, are by no means intractable; we actually describe a general strategy for tackling problems of this type. Of particular interest is the fact that, unlike previous research (e.g., Chiappori, Oreffice, and Quintana-Domeque, 2012), individual traits need not be continuous, nor homogenously assessed in the population; in our framework, one trait is discrete and it impacts the marital surplus in a multiplicative way. While we concentrate on smoking in our empirical application, other aspects could readily be considered. Race is an obvious (and important) example, to which more work will be devoted. Our second message is that, despite their highly stylized nature, matching models of this kind can generate strong testable predictions, that can be tested either by estimating a structural model (as in Chiappori, Salanié and Weiss, 2011, or Chiappori, Dias and Meghir, 2011) or by directly testing a reduced form – an approach adopted here. Encouraging is the fact that our predictions, including the less expected ones, seem to be quite well supported by the data.

A Proof of Proposition 3

Take some small $\varepsilon > 0$ such that

$$f(1-\varepsilon, 1-\varepsilon) > \lambda f(1,1)$$

Define $\eta(\varepsilon) > 0$ by

$$\int_{1-\varepsilon}^{1} dF_N(s) = \int_{1-\eta(\varepsilon)}^{1} dG_N(s). \tag{9}$$

so that there are exactly as many non-smoking men with SES above $1-\eta(\varepsilon)$ as non-smoking women with SES above $1-\varepsilon$. We claim that almost all female non-smokers with SES at least $1-\varepsilon$ are married with a male non-smoker with SES at least $1-\eta(\varepsilon)$ (note that (9) then implies that, conversely, almost all male non-smokers with SES at least $1-\eta(\varepsilon)$ are married with a female non-smoker with SES at least $1-\varepsilon$). Assume not, then there exists a positive measure set O of female non-smokers with SES at least $1-\varepsilon$ married with a smoker. By (9), there must exist a set O' of identical measure gathering male non-smokers with SES at least $1-\eta(\varepsilon)$, who are *not* married with female non-smokers with SES at least $1-\varepsilon$. Then either almost all males in O' are married with non-smokers with SES less than $1-\varepsilon$, or a non-null subset of males in O' is matched with smokers.

We start with the second case. Let $x \in O, y$ her (smoking) match, and $y' \in O'$ matched with a smoker x'. Surplus is

$$S = \lambda f(x,y) + \lambda f(x',y')$$

while matching x and y' would generate a surplus

$$S_1 = f(x, y') + \lambda f(x', y)$$

By definition of ε, $S_1 > S$, a contradiction.

Assume now that almost all males in O' are married with non-smokers with SES less than $1 - \varepsilon$. Let $x \in O$ be matched with y_S, smoker, while $y' \in O'$ is matched with a non-smoking wife $x' < 1 - \varepsilon$. The surplus generated is thus

$$\Sigma = f(x', y') + \lambda f(x, y_S)$$

whereas mixing matches would generate

$$\Sigma_1 = f(x, y') + \lambda f(x', y_S)$$

Note that $y_S > 1 - \eta(\varepsilon)$, for otherwise

$$\begin{aligned}\Sigma_1 - \Sigma &= f(x, y') - f(x', y') - \lambda (f(x, y_S) - f(x', y_S)) \\ &> f(x, y') - f(x', y') - (f(x, y_S) - f(x', y_S)) > 0\end{aligned}$$

by supermodularity, which contradicts surplus maximization. Define

$$\phi(s) = f(x, s) - f(x', s)$$

then ϕ is differentiable and strictly positive on $[0, 1]$. We have that

$$|\phi(y') - \phi(y_S)| \leq |y' - y_S| M$$

where $M = \sup_{[0,1]} |\phi'|$, and where $|y' - y_S| \leq \eta(\varepsilon)$. It follows that

$$\phi(y_S) \leq \phi(y') + \eta(\varepsilon) M$$

therefore

$$\Sigma_1 - \Sigma = \phi(y') - \lambda\phi(y_S) \geq (1-\lambda)\phi(y') - \lambda\eta(\varepsilon) M$$

which is positive for ε small enough, a contradiction.

B Proof of Proposition 4

The proof relies on the following Lemma:

Lemma 1 *If an open set of non-smoking women are indifferent between marrying a smoker and a non-smoker, and marry any with positive probability at a stable match, then the smoker has a higher SES than the non-smoker. However, if a smoker is indifferent between marrying a smoker and a non-smoker, and marries any with positive probability at a stable match, then the two potential spouses have the same SES.*

Proof. *Assume Ms. x marries either Mr. y (a non-smoker) or Mr. y' (a smoker) at the stable match; let $u(x)$ denote her utility. Then by stability*

$$\begin{aligned} u(x) &= \max_s f(x,s) - v_N(s) \\ &= \max_{s'} \lambda f(x,s') - v_S(s') \end{aligned}$$

where $v_N(s)$ (resp. $v_S(s')$) is the utility of a non-smoker (smoker) with SES s (s'); note that the max is reached for $s = y$ and $s' = y'$ respectively. By the envelope theorem:

$$u'(x) = \frac{\partial}{\partial x} f(x,y) = \lambda \frac{\partial}{\partial x} f(x,y')$$

Since $\partial f / \partial x$ is strictly increasing in x, $\lambda < 1$ requires $y' > y$. The proof of the second claim is similar. ■

We can now prove the proposition. By Lemma 1, $y' > y$ and $x = x'$. Assume that $q_N(y) > 0$, i.e. that y marries a smoker x'' with positive probability. Then $x'' > x = x'$ by Lemma 1. But the couples (x'', y) and (x', y') generate a surplus

$$S = \lambda f(x'', y) + \lambda f(x', y')$$

while the mixed couples (x', y) and (x'', y') would generate a surplus

$$S_1 = \lambda f(x', y) + \lambda f(x'', y')$$

and $S_1 > S$ by supermodularity of f; an open set of marriages satisfying this pattern would violate surplus maximization.

Similarly, assume that $p_S(x') < 1$, i.e., that x' marries a non-smoker $\bar{y}$ with positive probability. Then $\bar{y} = y' > y$ by Lemma 1. The couples (x, y) and $(x', \bar{y})$ generate a surplus

$$S = f(x, y) + \lambda f(x', \bar{y})$$

while the mixed couples $(x, \bar{y})$ and (x', y) would generate a surplus

$$S_1 = f(x, \bar{y}) + \lambda f(x', y)$$

Since

$$f(x, \bar{y}) - f(x, y) = f(x', \bar{y}) - f(x', y) > \lambda (f(x', \bar{y}) - f(x', y))$$

we have that $S_1 > S$; again, an open set of marriages satisfying this pattern would violate surplus maximization. The proof of the last statement is identical.

C Proof of Proposition 5: surplus maximization

We now provide two proofs of Proposition 5, aimed at illustrating the various ways the problem can be approached. We start with surplus maximization. From Proposition 4, we know that $q_N(y) = 0$ and $p_S(x) = 1$ for almost all x and y, and that moreover $\psi_N(x) = \psi_S(x)$. Constraints (4) to (7) become:

$$(1 - s_W) \int_x^1 (1 - p_N(t))\, dt = (1 - s_M)(1 - \phi_N(x)) \tag{10}$$

for non-smoking couples, and

$$s_W(1 - x) = s_M \int_{\psi_S(x)}^1 q_S(s)\, ds = s_M \int_{\psi_N(x)}^1 q_S(s)\, ds \tag{11}$$

for smokers. Finally, regarding mixed marriages:

$$(1 - s_W) \int_x^1 p_N(t)\, dt = s_M \int_{\psi_N(x)}^1 (1 - q_S(s))\, ds \tag{12}$$

which, using (11), becomes

$$(1 - s_W) \int_x^1 p_N(t)\, dt = s_M (1 - \psi_N(x)) - s_W(1 - x) \tag{13}$$

and in particular, adding up (10) and (13):

$$x = (1 - s_M)\, \phi_N(x) + s_M \psi_N(x) \tag{14}$$

Define $P(x) = \int_x^1 p_N(t)\, dt$, then $p_N(x) = -P'(x) := p(x)$ and $P(1) = 0$; (10) gives

$$\phi_N(x) = 1 - \frac{1 - s_W}{1 - s_M}(1 - x - P(x))$$

and (13) gives

$$\psi_N(x) = \psi_S(x) = 1 - \frac{1-s_W}{s_M} P(x) - \frac{s_W}{s_M}(1-x)$$

Aggregate surplus is:

$$\begin{aligned} S = & \int_0^1 \left[(1-p_N(t)) f(t, \phi_N(t)) + \lambda p_N(t) f(t, \psi_N(t)) \right] dF_N(t) \\ & + \lambda \int_0^1 f(t, \psi_S(t)) \, dF_S(t) \end{aligned}$$

which becomes

$$S = (1-s_W) \int_0^1 \left[(1-p_N(t)) t\phi_N(t) + \lambda p_N(t) t\psi_N(t) \right] dt + \lambda s_W \int_0^1 t\psi_S(t) \, dt$$

or

$$\begin{aligned} S = & (1-s_W) \int_0^1 (1-p_N(t)) t \left(1 - \frac{1-s_W}{1-s_M} (1-t-P(t)) \right) dt \\ & + (1-s_W) \lambda \int_0^1 p_N(t) t \left(1 - \frac{1-s_W}{s_M} P(t) - \frac{s_W}{s_M}(1-t) \right) dt \\ & + \lambda s_W \int_0^1 t \left(1 - \frac{1-s_W}{s_M} P(t) - \frac{s_W}{s_M}(1-t) \right) dt \end{aligned} \tag{15}$$

which must be maximized in p_N under the constraints

$$\dot{P} = -p_N, \quad P(1) = 0$$

and

$$0 \leq p_N(x) \leq 1.$$

This optimal control problem can be solved using standard techniques. Let H be the hamiltonian:

$$\begin{aligned} H(t, p_N, P) = {} & (1 - s_W)(1 - p_N)\, t \left(1 - \frac{1 - s_W}{1 - s_M}(1 - t - P)\right) \\ & + (1 - s_W)\lambda p_N t \left(1 - \frac{1 - s_W}{s_M} P - \frac{s_W}{s_M}(1 - t)\right) \\ & + \lambda s_W t \left(1 - \frac{1 - s_W}{s_M} P - \frac{s_W}{s_M}(1 - t)\right) + \zeta p_N \end{aligned} \tag{16}$$

where ζ denotes the dual function.

The crucial remark, here, is that H is linear in p_N. Therefore:

- if $\partial H / \partial p_N > 0$, the optimum is reached for $p_N = 1$
- if $\partial H / \partial p_N < 0$, the optimum is reached for $p_N = 0$
- lastly, if $\partial H / \partial p_N = 0$, the optimum does not depend on p_N

Then one can show the following (computations available from the authors):

- if $t \geq \widetilde{x}$, where

  $$\widetilde{x} = \frac{(s_M + \lambda(1 - s_M))(s_M - s_W)}{s_M(1 - s_W) - \lambda s_W(1 - s_M)} \leq 1$$

  then $\partial H / \partial p_N < 0$, and $p_N(t) = 0$ at the optimum

- if $t < \widetilde{x}$, then the optimum requires

  $$p_N(t) = p = \frac{s_M(1 - s_W) - \lambda s_W(1 - s_M)}{(s_M + \lambda(1 - s_M))(1 - s_W)} \leq 1$$

  in which case ζ is constant, P is linear and $\partial H / \partial p_N = 0$ for all $t < \widetilde{x}$.

The final determination of the solution is straightforward.

D Proof of Proposition 5: stability conditions

A more direct approach consists in *assuming* that the equilibrium is as described in the Proposition, and deriving a complete characterization, including the resulting allocation of surplus between members; one can then check that the latter satisfy the stability conditions.

D.1 Population constraints

We first characterize the marital patterns, using the fact that since the surplus is supermodular in (x, y), couples within a given category must marry assortatively. Therefore:

- For $x \geq \widetilde{x}$, a non-smoking woman with SES $x \geq \widetilde{x}$ is matched with a non-smoking man with SES y such that the number of non-smoking women above x equals that of non-smoking men above y:

  $$(1 - s_W)(1 - x) = (1 - s_M)(1 - y)$$

  or equivalently:

  $$\begin{aligned} y &= \phi_N(x) = \frac{1 - s_W}{1 - s_M} x - \frac{s_M - s_W}{1 - s_M} \text{ and} \\ x &= \frac{1 - s_M}{1 - s_W} y + \frac{s_M - s_W}{1 - s_W} \end{aligned}$$

  In particular, the spouse of a non-smoking Ms. $\widetilde{x}$ has a SES:

  $$\widetilde{Y} = \frac{1 - s_W}{1 - s_M} \widetilde{X} - \frac{s_M - s_W}{1 - s_M}$$

- Similarly, a smoking woman with SES $x \geq \widetilde{x}$ is matched with a smoking man with SES

y such that

$$x = \frac{s_M}{s_W} y - \frac{s_M - s_W}{s_W} \text{ or}$$
$$y = \phi_S(x) = \frac{s_W}{s_M} x + \frac{s_M - s_W}{s_M}$$

In particular, if Ms. $\widetilde{x}$ is a smoker, her husband has a SES:

$$\widetilde{y}' = \frac{s_W}{s_M} \widetilde{x} + \frac{s_M - s_W}{s_M}$$

- For $x < \widetilde{x}$, a non-smoking woman with SES x marries a smoker with probability p, a non-smoker with probability $(1-p)$. Assortative matching implies that:
  - the number of non-smoking men above y equals the number of non-smoking women above x who marry a non-smoker:

$$(1 - s_W)(1 - p)(\widetilde{x} - x) = (1 - s_M)(\widetilde{y} - y)$$

or

$$y = \phi_N(x) = (1 - p)\frac{(1 - s_W)}{(1 - s_M)} x + p \frac{(1 - s_W)}{(1 - s_M)} \widetilde{x} - \frac{s_M - s_W}{1 - s_M}$$

In particular, since $x = 0$ marries $y = 0$, it must be the case that

$$p\widetilde{x} = \frac{s_M - s_W}{1 - s_W} \tag{17}$$

  - the number of smoking men above y who marry a non-smoker (which happens with probability q) equals the number of non-smoking women above x who marry a smoker:

$$(1 - s_W) p (\widetilde{x} - x) = q s_M (\widetilde{y}' - y)$$

therefore

$$y = \phi_S(x) = \frac{p}{q}\frac{(1 - s_W)}{s_M}x + \frac{qs_W - p(1 - s_W)}{qs_M}\widetilde{x} + \frac{s_M - s_W}{s_M}$$

which, as above, implies that

$$\widetilde{x} = \frac{q(s_M - s_W)}{p(1 - s_W) - qs_W} \tag{18}$$

- Finally, female smokers marry the fraction q of male smokers who marry their own:

$$s_W(\widetilde{x} - x) = (1 - q)s_M(\widetilde{y}' - y)$$

which gives

$$x = (1 - q)\frac{s_M}{s_W}y + q\widetilde{x} - \left(\frac{s_M}{s_W} - 1\right)(1 - q)$$

and

$$\widetilde{x} = \left(\frac{s_M}{s_W} - 1\right)\frac{(1 - q)}{q} \tag{19}$$

One can readily check that (17), (18) and (19) are redundant: if some numbers $(p, q, \widetilde{x})$ satisfy any two of them, they also satisfy the third.

D.2 Utilities

We next derive the allocation of intrahousehold welfare in each couple that supports the equilibrium. Let $u_N(x)$ (resp. $u_S(x), v_N(y), v_S(y)$) denote the utility of a female non-smoker (resp. female smoker, male non-smoker, male smoker) with SES x (resp. y). Start with high SES non-smoking women. Stability requires that:

$$u_N(x) + v_N(y) \geq xy$$

equality obtaining when x and y are matched at the stable equilibrium. It follows that:

$$u_N(x) = \max_y (xy - v_N(y))$$

and from the envelope theorem:

$$u_N'(x) = \frac{\partial}{\partial x}(xy)$$

the partial being taken at the point $y = \phi_N(x)$. Therefore for $x \geq \widetilde{x}$:

$$u_N'(x) = \phi_N(x) = \frac{1-s_W}{1-s_M}x - \frac{s_M - s_W}{1-s_M}$$

which gives

$$u_N(x) = \frac{1-s_W}{1-s_M}\frac{x^2}{2} - \frac{s_M - s_W}{1-s_M}x + K$$

and similarly

$$v_N(y) = \frac{1-s_M}{1-s_W}\frac{y^2}{2} + \frac{s_M - s_W}{1-s_W}y + K'$$

where K, K' are integration constants.

Similar computations give:

- for high SES female smokers:

$$u_S(x) = \lambda\left(1 - \frac{s_W}{s_M}\right)x + \lambda\frac{s_W}{s_M}\frac{x^2}{2} + L$$

$$v_S(y) = \lambda\left(1 - \frac{s_M}{s_W}\right)y + \lambda\frac{s_M}{s_W}\frac{y^2}{2} + L'$$

- for lower SES female non-smokers ($x < \widetilde{x}$) marrying a non-smoker:

$$u_N(x) = (1-p)\frac{(1-s_W)}{(1-s_M)}\frac{x^2}{2}$$
$$v_N(y) = \frac{1}{(1-p)}\frac{(1-s_M)}{(1-s_W)}\frac{y^2}{2}$$

- while if she marries a smoker:

$$\bar{u}_N(x) = \lambda\frac{p}{q}\frac{(1-s_W)}{s_M}\frac{x^2}{2}$$
$$\bar{v}_S(y) = \lambda\frac{q}{p}\frac{s_M}{(1-s_W)}\frac{y^2}{2}$$

- Finally, for lower SES female smokers:

$$u_S(x) = \lambda\frac{1}{1-q}\frac{s_W}{s_M}\frac{x^2}{2}$$
$$v_S(y) = \lambda(1-q)\frac{s_M}{s_W}\frac{y^2}{2}$$

D.3 Indifference conditions

The crucial remark, at this point, is that any agent who may marry either of two spouses with positive probability must be indifferent. This applies to non-smoking women: for any $x < \widetilde{x}$, it must be the case that:

$$u_N(x) = \bar{u}_N(x)$$

which requires that:

$$\lambda\frac{p}{q}\frac{(1-s_W)}{s_M} = (1-p)\frac{(1-s_W)}{(1-s_M)}$$
$$\lambda\left(\frac{qs_W - p(1-s_W)}{qs_M}\widetilde{x} + \frac{s_M - s_W}{s_M}\right) = \left(p\frac{(1-s_W)}{(1-s_M)}\widetilde{x} - \frac{s_M - s_W}{1-s_M}\right)$$

Together with (17), (18) and (19), these conditions imply that:

$$\begin{aligned}
\widetilde{x} &= \frac{(s_M + \lambda (1 - s_M)) (s_M - s_W)}{s_M (1 - s_W) - \lambda s_W (1 - s_M)} \leq 1 \\
p &= \frac{s_M (1 - s_W) - \lambda s_W (1 - s_M)}{(s_M + \lambda (1 - s_M)) (1 - s_W)} \leq 1 \\
q &= \frac{s_M (1 - s_W) - \lambda s_W (1 - s_M)}{s_M} \leq 1
\end{aligned} \tag{20}$$

The indifference condition also applies to male smokers; again, it leads to conditions (20).

D.4 Pinning down the constants

The constants can readily be recovered from continuity of utilities at $\widetilde{x}$ (resp. at $\widetilde{y}, \widetilde{y}'$). One gets:

$$\begin{aligned}
K &= \frac{(s_M - s_W)^2 (\lambda + s_M - \lambda s_M)}{2 (s_M - 1) (s_M (1 - s_W) - \lambda s_W (1 - s_M))} \\
K' &= -\frac{\lambda (s_M - s_W)^2}{2 (1 - s_W) (s_M (1 - s_W) - \lambda s_W (1 - s_M))} \\
L &= -\frac{1}{2} \frac{\lambda}{s_M} \frac{(s_M - s_W)^2 (\lambda + s_M - \lambda s_M)}{(s_M (1 - s_W) - \lambda s_W (1 - s_M))} \\
L' &= \frac{1}{2} \frac{\lambda}{s_W} \frac{(s_M - s_W)^2}{(s_M (1 - s_W) - \lambda s_W (1 - s_M))}
\end{aligned}$$

D.5 Comparative statics

The comparative statics predictions can directly be derived from these expressions. For instance:

$$\frac{\partial \widetilde{x}}{\partial \lambda} = \frac{s_M (s_M - s_W)(1 - s_M)}{(s_M - \lambda s_W - s_M s_W + \lambda s_M s_W)^2} > 0$$
$$\frac{\partial \widetilde{x}}{\partial s_M} = \frac{(1-\lambda)\left(s_M^2 (1 - s_W) + \lambda s_W (1 - s_M)^2\right)}{(s_M - \lambda s_W - s_M s_W + \lambda s_M s_W)^2} > 0$$
$$\frac{\partial \widetilde{x}}{\partial s_W} = -\frac{s_M (1-\lambda)(1 - s_M)(\lambda + s_M - \lambda s_M)}{(s_M - \lambda s_W - s_M s_W + \lambda s_M s_W)^2} < 0$$

D.6 Testing stability

Finally, we need to check the stability conditions for each possible couple. When the husband and the wife belong to one of the five category pairs that appear with positive probability in the stable match, these conditions are satisfied, since they stem directly from supermodularity. We therefore need to check them for the remaining $16 - 5 = 11$ pairs.

Starting with a non-smoker high SES wife who could marry a high SES smoker, we must check that for $x \geq \widetilde{x}, y' \geq \widetilde{y}'$:

$$P(x, y) = u_N(x) + v_S(y') - \lambda x y \geq 0$$

Here, P is convex in (x, y), and its minimum satisfies

$$\frac{\partial P(x,y)}{\partial x} = \frac{\partial P(x,y)}{\partial y} = 0$$

which gives

$$x = (s_M - s_W) \frac{\lambda + s_M - \lambda s_M}{s_M (2 - s_M - s_W) - \lambda (1 - s_M) (s_M + s_W)} = \widetilde{x}$$
$$y = (s_M - s_W) \frac{2 - \lambda - s_M + \lambda s_M}{s_M (2 - s_M - s_W) - \lambda (1 - s_M) (s_M + s_W)} = \widetilde{y}'$$

Since $P(\widetilde{x}, \widetilde{y}') = 0$ by definition, the condition is satisfied.

In the remaining ten cases, one can show, using similar computations, that the difference $u + v - S$ between the sum of individual utilities and the potential surplus is minimum either at the boundary of the interval over which the expression is valid and vanishes at these points, or at some interior point at which it is non-negative. The explicit calculations are available on demand from the authors.

D.7 Numerical simulations

In the numerical simulations, we take:

$$s_M = .25, s_W = .2, \lambda = .8$$

Then the functions take the following form:

$$u_N(x) = \begin{cases} 0.470\,59x^2 & \text{if} \quad x < 0.531\,25 \\ 0.533\,33x^2 - 6.666\,7 \times 10^{-2}x + 0.017\,71 & \text{if} \quad 0.531\,25 \leq x \end{cases}$$

$$u_S(x) = \begin{cases} 0.470\,59x^2 & \text{if} \quad x < 0.531\,25 \\ 0.32x^2 + 0.16x - 4.250\,0 \times 10^{-2} & \text{if} \quad 0.531\,25 \leq x \end{cases}$$

$$v_N(y) = \begin{cases} 0.531\,25y^2 & \text{if} \quad y < .5 \\ 0.468\,75y^2 + 0.062\,5y - 0.015\,63 & \text{if} \quad 0.5 \leq y \end{cases}$$

$$v_S(y) = \begin{cases} 0.34y^2 & \text{if } y < .625 \\ 0.5y^2 - 0.2y + 0.0625 & \text{if } 0.625 \leq y \end{cases}$$

References

[1] Banerjee, A., Duflo, E., Ghatak, M., Lafortune, J. (2012) "Marry for What? Caste and Mate Selection in Modern India," *American Economic Journal: Microeconomics*, forthcoming.

[2] Becker, G. (1973) "A Theory of Marriage: Part I," *Journal of Political Economy*, 81(4):813–846.

[3] Becker, G. (1991) *A treatise on the family*, Harvard University Press.

[4] Bergstrom, T., Cornes, R. (1983) "Independence of Allocative Efficiency from Distribution in the Theory of Public Goods," *Econometrica*, 51(6):1753–1765.

[5] Browning, M., Chiappori, P.-A., Weiss, Y. (2012) *Family Economics*, Cambridge University Press.

[6] Caraballo, R.S., Giovino, G.A., Pechacek, T.F., Mowery, P.D. (2001) "Factors associated with discrepancies between self-reports on cigarette smoking and measured serum cotinine levels among person aged 17 years or older: third national health and nutrition examination survey, 1988-1994," *American Journal of Epidemiology*, 153(8):807–814.

[7] CDC (2006) "The health consequences of involuntary exposure to tobacco smoke: A report of the Surgeon General," Washington, DC: U.S. Government Printing Office; 2006. Available from: `http://www.cdc.gov/tobacco/data_statistics/sgr/2006/index.htm`

[8] CDC (2008) "Smoking-attributable mortality, years of potential life lost, and productivity losses–United States, 2000–2004," *Morbidity and Mortality Weekly Report*, CDC, 57(45):1226–1228. Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5745a3.htm

[9] CDC (2010) "Vital Signs: Current Cigarette Smoking Among Adults Aged ≥18 Years — United States, 2009," *Morbidity and Mortality Weekly Report*, CDC, 59(35):1135–1140. Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5935a3.htm?s_cid=mm5935a3_w

[10] Chiappori, P.-A., Dias, M., Meghir, C. (2011) "The Marriage Market, Labor Supply and Education Choice," mimeo, Yale University.

[11] Chiappori, P.-A., Iyigun, M., Weiss, Y. (2009) "Investment in Schooling and the Marriage Market," *American Economic Review*, 99(5):1689–1713.

[12] Chiappori, P.-A., McCann, R., Nesheim, L. (2010) "Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness," *Economic Theory*, 42(2):317–354.

[13] Chiappori, P.-A., Oreffice, S. (2008) "Birth Control and Female Empowerment: An Equilibrium Analysis," *Journal of Political Economy*, 116(1):113–140.

[14] Chiappori, P.-A., Oreffice, S., Quintana-Domeque, C. (2012) "Fatter attraction: anthropometric and socioeconomic matching on the marriage market," *Journal of Political Economy*, forthcoming.

[15] Chiappori, P.-A., Reny, P. (2004) "Matching to share risk," mimeo, University of Chicago.

[16] Chiappori, P.-A., Salanié, B., Weiss, Y. (2011) "Partner Choice and the Marital College Premium," mimeo, Columbia University.

[17] Choo, E., Siow, A. (2006) "Who marries whom and why," *Journal of Political Economy*, 114(1):172–201.

[18] Clark, A., Etilé, F. (2006) "Don't give up on me baby: spousal correlation in smoking behaviour," *Journal of Health Economics*, 25(5):958–978.

[19] Coles, M., Francesconi, M. (2011) "On the Emergence of Toyboys: The Timing of Marriage with Aging and Uncertain Careers," *International Economic Review*, 52(3):825–853.

[20] Delnevo, C.D., Bauer, U.E. (2009) "Monitoring the tobacco use epidemic III: The host: data sources and methodological challenges," *Preventive Medicine*, 48(1):S16-S23.

[21] De Walque, D. (2010) "Education, information, and smoking decisions: evidence from smoking histories in the United States, 1940–2000," *Journal of Human Resources*, 45(3):682–717.

[22] Ekeland, I. (2010) "Existence, uniqueness and efficiency of equilibrium in hedonic markets with multidimensional types," *Economic Theory*, 42(2):275–315.

[23] Ekeland, I., Heckman, J., Nesheim, L. (2004) "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112(1):S60–S109.

[24] Farell, P., Fuchs, V. (1982) "Schooling and Health: The Cigarette Connection," *Journal of Health Economics*, 1(3):217–230.

[25] Figalli, A., Kim, Y-H., McCann, R. (2011) "When is multidimensional screening a convex program?," *Journal of Economic Theory*, 146(2):454–478.

[26] Flinn, C., Del Boca, D. (2005) "Household Time Allocation and Modes of Behavior: A Theory of Sorts," IZA Discussion Papers 1821.

[27] Fox, J. (2010) "Identification in matching games," *Quantitative Economics*, 1(2):203–254.

[28] Galichon, A., Salanié, B. (2010) "Matching with Trade-Offs: Revealed Preferences over Competing Characteristics," Columbia University, Discussion Paper: 0910-14.

[29] Glymour, M., DeFries, T., Kawachi, I., Avendano, M. (2008) "Spousal Smoking and Incidence of First Stroke The Health and Retirement Study," *American Journal of Preventive Medicine*, 35(3):245–248.

[30] Gretsky, N., Ostroy, J., Zame, W. (1999) "Perfect competition in the continuous assignment model," *Journal of Economic Theory*, 88(1):60–118.

[31] Gruber, J. (2001) "Tobacco at the crossroads: the past and future of smoking regulation in the United States," *Journal of Economic Perspectives*, 15(2):193–212.

[32] Heckman, J., Matzkin, R., Nesheim, L. (2010) "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 78(5):1569–1591.

[33] Hitsch, G., Hortaçsu, A., Ariely, D. (2010) "Matching and Sorting in Online Dating," *American Economic Review*, 100(1):130–163.

[34] Jaeger, D. (1997) "Reconciling the old and new Census Bureau education questions: recommendations for researchers," *Journal of Business and Economic Statistics*, 15(4):300–309.

[35] Kantorovich, L. (1942) "On the translocation of masses," *Dokl. Akad. Nauk SSSR*, 37(7–8):227–229.

[36] Koopmans, T.C., Beckmann, M. (1957) "Assignment problems and the location of economic activities," *Econometrica*, 25(1):53–76.

[37] Lam, D. (1988) "Marriage Markets and Assortative Mating with Household Public Goods: Theoretical results and empirical implications," *Journal of Human Resources*, 23(4):462–487.

[38] Legros, P., Newman, F. (2007) "Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities," *Econometrica*, 75(4):1073–1102.

[39] Madrian, B., Lefgren, L. (1999) "A note on longitudinally matching CPS respondents," NBER Technical Working Paper #247.

[40] Mannino, D., Siegel, M., Rose, D., Nkuchia, J., Etzel, R. (1997) "Environmental tobacco smoke exposure in the home and worksite and health effects in adults: results from the 1991 national health interview survey," *Tobacco Control*, 6(4):296–305.

[41] Maralani, V. (2009) "An Unequal Start: The Alignment of Education and Smoking in Families of Origin," mimeo, Yale University.

[42] Mare, R. (2008) "Educational Assortative Mating in Two Generations," mimeo, University of California at Los Angeles.

[43] McCann, R., Guillen, N. (2010) "Five lectures on optimal transportation: geometry, regularity and applications," University of Toronto.

[44] McCann, R., Shi, X., Siow, A., Wolthoff, R. (2012) "Becker Meets Ricardo: Multisector Matching with Social and Cognitive Skills," Working Paper 454, University of Toronto.

[45] Mills, A., Messer, K., Gilpin, L., Pierce, G. (2009) "The effect of smoke-free homes on adult smoking behavior: a review," *Nicotine Tobacco Research*, 11(10):1131–1141.

[46] Monge, G. (1781) "Mémoire sur la théorie des déblais et de remblais," *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, 666–704.

[47] NCHS (2010) National Center for Health Statistics. Health, United States, 2009: With Special Feature on Medical Technology. Hyattsville, MD. 2010.

[48] Nesheim, L. (2012) "Identification of Multidimensional Hedonic Models," mimeo, University College London.

[49] Öberg, M., Jaakkola, M., Woodward, A., Peruga, A., Prüss-Ustün, A. (2011) "Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries," *Lancet*; 377(9760):139–146.

[50] Oreffice, S., Quintana-Domeque, C. (2010) "Anthropometry and socioeconomics among couples: Evidence in the United States," *Economics and Human Biology*, 8(3):373–384.

[51] Pencavel, J. (1998) "Assortative Mating by Schooling and the Work Behavior of Wives and Husbands," *American Economic Review Papers and Proceedings*, 88(2):326–329.

[52] Pilkington, P., Gray, S., Gilmore, A., Daykin, N. (2006) "Attitudes towards second hand smoke amongst a highly exposed workforce: survey of London casino workers," *Journal of Public Health*, 28(2):104–110.

[53] Qian, Z. (1998) "Changes in Assortative Mating: The Impact of Age and Education, 1970–1990" *Demography*, 35(3):279–292.

[54] Rochet, J.-C., Choné, P. (1998) "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66(4):783–826.

[55] Rochet, J.-C., Stole, L. (2003) "The economics of multidimensional screening," in *Advances in Economics and Econometrics – Theory and Applications, Eighth World Congress*, Dewatripont, M., Hansen, L., Turnovsky, P. (Eds.), Econometric Society Monographs, Cambridge University Press, Cambridge, 2003, 150–176.

[56] Silventoinen, K., Kaprio, J., Lahelma, E., Viken, R.J., Rose, R.J. (2003) "Assortative mating by body height and BMI; Finnish twins and their spouses," *American Journal of Human Biology*, 15(5):620–627.

[57] Shapley, L.S., Shubik, M. (1971) "The Assignment Game I: The Core," *International Journal of Game Theory*, 1(1):111–130.

[58] Shimer, R., Smith, L. (2000) "Assortative matching and search," *Econometrica*, 68(2):343–370.

[59] Surgeon General (2001) "Women and Smoking: a report from the Surgeon General," US Department of Health and Human Services.

[60] Sutton, G. (1980) "Assortative marriage for smoking habits," *Annals of Human Biology*, 7(5):449–456.

[61] Sutton, G. (1993) "Do men grow to resemble their wives or vice versa?," *Journal of Biological Science,* 25(1):25–29.

[62] Venters, M., Jacobs, D., Luepker, R., Maimaw, L., Gillum, R. (1984) "Spouse concordance of smoking patterns: the Minnesota heart survey," *American Journal of Epidemiology*, 120(4):608–616.

[63] Villani, C. (2003) "Topics in Optimal Transportation," Volume 58 of Graduate Studies in Mathematics. American Mathematical Society, Providence.

[64] Weiss, Y., Willis, R. (1997) "Match quality, new information, and marital dissolution," *Journal of Labor Economics*, 15(1):293–329.

[65] Wong, L. (2003) "Structural Estimation of Marriage Models," *Journal of Labor Economics*, 21(3):699–728.

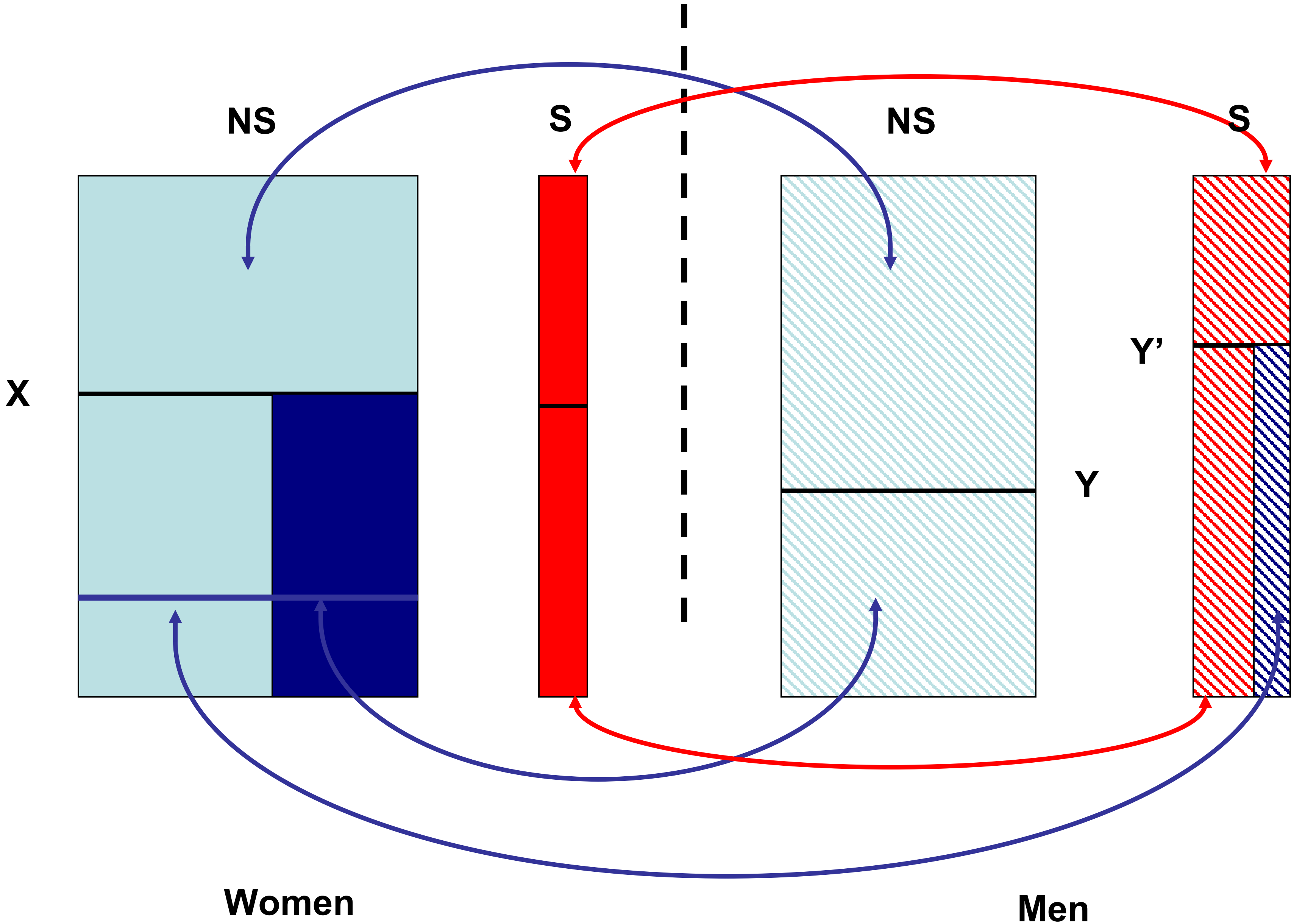

NS
S
NS
S
X
Y'
Y
Women
Men

Table 1:
Summary Statistics: Married versus Singles
Means (Standard Deviations)
Men aged 24-34, Women aged 22-32.
CPS 1996–2007.

	Married	
	Men	Women
Age	29.44	27.80
	(2.80)	(2.77)
Education	13.66	13.82
	(2.38)	(2.30)
Smoke	0.22	0.17
	(0.41)	(0.37)
# Children under age 6	0.82	
	(0.85)	
N	12,035	

	Singles (never married)	
	Men	Women
Age	29.22	27.17
	(3.12)	(3.14)
Education	13.63	13.78
	(2.41)	(2.33)
Smoke	0.25	0.21
	(0.43)	(0.41)
# Children under age 6	0.50	0.55
	(0.76)	(0.78)
N	28,086	29,102

Note: Sampling weights are used.

Table 2:
Summary Statistics: Married Couples by Smoking Status
Means (Standard Deviations)
Husband's age 24-34, Wife's age 22-32.
CPS 1996–2007.

	Both Non-Smokers		**Both Smokers**	
	Husband	Wife	Husband	Wife
Age	29.47	27.88	29.31	27.46
	(2.78)	(2.76)	(2.87)	(2.83)
Education	14.01	14.15	12.64	12.77
	(2.41)	(2.32)	(1.72)	(1.75)
Very Healthy	0.86	0.83	0.76	0.69
	(0.35)	(0.37)	(0.43)	(0.46)
# Children under age 6	0.81		0.85	
	(0.85)		(0.81)	
N	8,710		1,150	
	Smoking Husband & Non-Smoking Wife		**Non-Smoking Husband & Smoking Wife**	
	Husband	Wife	Husband	Wife
Age	29.28	27.66	29.59	27.68
	(2.83)	(2.76)	(2.91)	(2.81)
Education	12.70	13.18	13.10	13.06
	(2.33)	(2.29)	(1.88)	(1.90)
Very Healthy	0.75	0.76	0.76	0.70
	(0.44)	(0.43)	(0.43)	(0.46)
# Children under age 6	0.89		0.83	
	(0.86)		(0.84)	
N	1,408		767	

Note: Sampling weights are used.

Table 3:
Husband's age 24-34, Wife's age 22-32.
CPS 1996–2007.

A. Observed Matching

	Non-Smoking Wife	Smoking Wife
Non-Smoking Husband	**71.78%**	**6.50%**
	(8710)	(767)
Smoking Husband	**11.71%**	**10.01%**
	(1408)	(1150)

B. Random Matching

	Non-Smoking Wife	Smoking Wife
Non-Smoking Husband	**65.35%**	**12.92%**
Smoking Husband	**18.14%**	**3.59%**

Note: Sampling weights are used. Weighed % and (non-weighed number of observations).

Table 4:
Regression of Smoking Status on Education
Men aged 24-34, Women aged 22-32.
CPS 1996–2007.

I. Married	**SUR**		
	Men		Women
Education	−0.036*** (0.001)		−0.026*** (0.001)
Test of equality		$\chi^2(1) = 29.89$ p-value = 0.0000	
N		12,035	
II. Singles	**OLS**		
	Men		Women
Education	−0.037*** (0.001)		−0.030*** (0.001)
Test of equality [A]		*t*-test = 4.29 p-value = 0.0000	
N	28,086		29,102

Note: Sampling weights are used. All regressions include the following additional controls: age, year and state fixed effects. Standard errors in parentheses.

[A] Test of equality performed after estimating the following model: Smoking = a + b*education + c*female + d*education*female + additional controls.
The test of equality is Ho: d = 0.

*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table 5:
Sorting by education
Husband's age 24-34, Wife's age 22-32.
CPS 1996–2007.

	Both Non-Smokers		**Both Smokers**	
	Wife's Education	Husband's Education	Wife's Education	Husband's Education
Spouse's Education	0.633***	0.694***	0.458***	0.442***
	(0.013)	(0.014)	(0.035)	(0.034)
N	8710	8710	1150	1150
R^2	0.47	0.47	0.27	0.28

	Smoking Husband & Non-Smoking Wife		**Non-Smoking Husband & Smoking Wife**	
	Wife's Education	Husband's Education	Wife's Education	Husband's Education
Spouse's Education	0.600***	0.618***	0.495***	0.497***
	(0.031)	(0.036)	(0.044)	(0.041)
N	1408	1408	767	767
R^2	0.43	0.44	0.34	0.33

Note: All regressions include: own age, year and state fixed effects. Sampling weights are used. Robust standard errors in parentheses.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table 6:
Regression of Education by Smoking Status on Spouse's Education and Smoking Behavior
Husband's age 24-34, Wife's age 22-32.
CPS 1996–2007.

	Wife's Education		Husband's Education	
	Non-Smoker	Smoker	Non-Smoker	Smoker
Spouse's Education	0.630***	0.473***	0.684***	0.556***
	(0.012)	(0.027)	(0.013)	(0.026)
Spouse Smokes	−0.141**	−0.025	−0.209***	0.160**
	(0.060)	(0.086)	(0.074)	(0.076)
N	10118	1917	9477	ÿ558
R^2	0.47	0.29	0.46	0.36

Note: All regressions include: own age, year and state fixed effects. Sampling weights are used. Robust standard errors in parentheses.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table 7:
Regression of Smoking Status on Education
Husband's age 24-34, Wife's age 22-32.
CPS 1996–2007.

	Men with NS Women		Women with S Men	
	(1)	(2)	(3)	(4)
Own Education	−0.028***	−0.024***	−0.021***	−0.029***
	(0.002)	(0.002)	(0.006)	(0.007)
Test of Equality (*p-value*)	*0.0146*		*0.0341*	
Spouse's Education	--	−0.006**	--	0.014**
		(0.002)		(0.006)
N	10118	10118	2558	2558
R^2	0.05	0.05	0.06	0.06

Note: All regressions include: own age, year and state fixed effects. Sampling weights are used. Robust standard errors in parentheses.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table A1:
Regression of Education by Smoking Status on Spouse's Education and Smoking Behavior controlling for health status and number of children
Husband's age 24-34, Wife's age 22-32.
CPS 1996–2007.

	Wife's Education		Husband's Education	
	Non-Smoker	Smoker	Non-Smoker	Smoker
Spouse's Education	0.600***	0.460***	0.668***	0.541***
	(0.013)	(0.027)	(0.014)	(0.027)
Spouse Smokes	−0.108*	−0.029	−0.185**	0.164**
	(0.060)	(0.086)	(0.074)	(0.077)
Controlling for spouse's health status	YES	YES	YES	YES
Controlling for number of children under 6	YES	YES	YES	YES
N	10,118	1,917	9,477	2,558
R^2	0.49	0.29	0.46	0.37

Note: All regressions include: own age, year and state fixed effects. Spouse's health status is controlled for by a dummy variable: 1 if excellent or very good health, 0 if good, fair or poor. Sampling weights are used. Robust standard errors in parentheses.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table A2:
Regression of Education by Smoking Status on Spouse's Education and Smoking Behavior controlling for health status and number of children adjusting education following Jaeger (1997)
Husband's age 24-34, Wife's age 22-32.
CPS 1996–2007.

	Wife's Education		Husband's Education	
	Non-Smoker	Smoker	Non-Smoker	Smoker
Spouse's Education	0.598***	0.471***	0.667***	0.541***
	(0.011)	(0.028)	(0.012)	(0.025)
Spouse Smokes	−0.129**	−0.018	−0.210***	0.137*
	(0.058)	(0.085)	(0.073)	(0.075)
Controlling for spouse's health status	YES	YES	YES	YES
Controlling for number of children under 6	YES	YES	YES	YES
N	10,118	1,917	9,477	2,558
R^2	0.49	0.31	0.46	0.37

Note: All regressions include: own age, year and state fixed effects. Spouse's health status is controlled for by a dummy variable: 1 if excellent or very good health, 0 if good, fair or poor. Sampling weights are used. Robust standard errors in parentheses.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table A3:
Regression of Education by Smoking Status on Spouse's Education and Smoking Behavior
Husband's age 22-32, Wife's age 20-30.
CPS 1996–2007.

	Wife's Education		Husband's Education	
	Non-Smoker	Smoker	Non-Smoker	Smoker
Spouse's Education	0.617***	0.429***	0.685***	0.546***
	(0.014)	(0.030)	(0.017)	(0.030)
Spouse Smokes	−0.154**	−0.059	−0.172**	0.199**
	(0.066)	(0.097)	(0.081)	(0.085)
N	7861	1506	7252	2115
R^2	0.48	0.27	0.47	0.36

Note: All regressions include: own age, year and state fixed effects. Sampling weights are used. Robust standard errors in parentheses.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table A4:
Observed Matching

Husband's age 24-36, Wife's age 22-34.
PSID 1999-2007.

Weighed %

I. Full sample

	Non-Smoking Wife	Smoking Wife
Non-Smoking Husband	**74.74%**	**5.03%**
Smoking Husband	**11.43%**	**8.80%**

II. Recently married: marital duration ≤ 4 years

	Non-Smoking Wife	Smoking Wife
Non-Smoking Husband	**71.66%**	**6.88%**
Smoking Husband	**12.17%**	**9.29%**

Table A5:
Regression of Education by Smoking Status on Spouse's Education and Smoking Behavior

Husband's age 24-36, Wife's age 22-34.
PSID 1999-2007.

I. Full sample

	Wife's Education		Husband's Education	
	Non-Smoker	Smoker	Non-Smoker	Smoker
Spouse's Education	0.577***	0.683***	0.632***	0.641***
	(0.039)	(0.073)	(0.039)	(0.088)
Spouse Smokes	−0.213	0.052	−0.179	0.517*
	(0.181)	(0.302)	(0.279)	(0.268)
N	2035	350	1873	512
# Couples	945	213	881	293
R^2	0.48	0.60	0.47	0.57

II. Recently married: marital duration ≤ 4 years

	Wife's Education		Husband's Education	
	Non-Smoker	Smoker	Non-Smoker	Smoker
Spouse's Education	0.529***	0.792***	0.554***	0.548***
	(0.071)	(0.141)	(0.050)	(0.087)
Spouse Smokes	−0.137	0.095	−0.378	0.516
	(0.275)	(0.522)	(0.339)	(0.407)
N	941	188	868	261
# Couples	653	141	601	198
R^2	0.47	0.63	0.43	0.67

Note: All regressions include: own age, year and state fixed effects. Sampling weights are used. Standard errors clustered at the household id level in parentheses.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1