ISSN 1471-0498



DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

BIDIMENSIONAL MATCHING WITH HETEROGENEOUS PREFERENCES: SMOKING IN THE MARRIAGE MARKET

Pierre-Andre Chiappori, Sonia Oreffice and Climent Quintana-Domeque

Number 696
January 2014

Manor Road Building, Manor Road, Oxford OX1 3UQ

Bidimensional Matching with Heterogeneous Preferences: Smoking in the Marriage Market*

Pierre-André Chiappori
Columbia University

Sonia Oreffice
University of Surrey and IZA

Climent Quintana-Domeque
University of Oxford and IZA

January 2014

Abstract

We develop a bidimensional matching model under transferable utility, where individuals are characterized by a continuous trait (e.g., socioeconomic status) and a binary attribute (e.g., smoking status). The model is "truly multidimensional", in the sense that the impact of the traits cannot be summarized by a one-dimensional index. We present a general resolution strategy based on optimal control theory, and characterize the stable matching. We derive testable predictions about equilibrium matching patterns. Using US data, we find that the observed marital sorting of smokers and non-smokers by education is consistent with our model.

Keywords: Marriage market, multidimensional matching, continuous and discrete characteristics, heterogeneous preferences.

JEL Codes: D1, J1.

*We thank Bernard Salanié and Marco Francesconi, as well as participants at the "The Econometrics of Matching Invited Session" of the EEA 2013, SaM Mainz 2013, Paris Matching Workshop 2013, Royal Economic Society Meetings 2012, Barcelona MOVE Family Conference 2011, and seminars at University of Oxford, University of Chicago, Columbia University, University of Aarhus, Universitat Pompeu Fabra 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 a theoretical and an empirical perspective. Most models focus on 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; Pencavel, 1998), 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 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).

However, empirical evidence strongly suggests that matching processes are actually multidimensional; spouses tend to be similar in a variety of characteristics, including age, education, race, religion, anthropometric characteristics and personality traits (e.g., Becker, 1991; Dupuy and Galichon, 2012; 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. Although some properties (such as supermodularity of the surplus function or assortativeness of the stable matching) are intrisincally one-dimensional, they can be generalized to a multidimensional setting (see Chiappori, McCann and Nesheim, 2010, for a recent survey). But even though the conceptual framework is known, multidimensional models raise a series of specific and often difficult problems. Not surprisingly, theoretical frameworks

involving multidimensional matching have very rarely been analyzed in the applied theory literature.

While a part of the literature on the marriage market use 'pure' (i.e. frictionless) matching models, other contributions refer to a search framework. Search models present several advantages. They are particularly adequate in situations where frictions are of utmost economic importance; moreover, they can be quite useful in dynamic contexts (a point recently made by Robert Shimer). On the other hand, they tend to be less tractable and more opaque than matching models, and typically rely on strong (and often untestable) assumptions on preferences or the nature of the matching technology. In the end, here as in most of economic theory, different models can be relevant for different purposes. A standard analysis of supply and demand on the market for some commodity will typically disregard, at least in a first stage, informational frictions or local horizontal differentiation. Similarly, while an analysis of the labor market that concentrates on unemployment can greatly benefit from a precise investigation of the frictions affecting this market, the latter may be less relevant if the issue at stake is which CEOs end up managing which firms. From this perspective, search and matching models are probably more complements than substitutes. The goal of the present contribution is to illustrate the feasibility and the potential benefits of the latter approach in a complex, multidimensional setting.

Our paper The present paper considers 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, or any combination of those. The other characteristic is discrete, and more precisely dichotomic; in our empirical application, we use individuals' smoking status. In most of the analysis, we assume smoking status to be exogenous. However, we then discuss how it can be endogeneized. Specifically, we describe how our matching model can then be considered at the second stage of a game, the first stage consisting of people choosing their status,. We show that the model's outputs (and especially the derivation of the intracouple allocation of the surplus, conditional on the spouses' smoking status, an equilibrium outcome of matching models) provide the crucial information needed to analyze the first stage.

The surplus function we consider is differentiable and supermodular in the continuous

indices, and is (multiplicatively) impacted by the discrete characteristic: it may (but needs not to) be diminished by the presence of a smoker. A crucial aspect of our setting is that this impact is heterogeneous, in the sense that smokers do not mind marrying a smoker, whereas non-smokers do mind. This assumption, which is in line with both the medical literature on the health consequences of smoking and the public health literature on smoking perception, has two immediate consequences. One is to rule out a single index representation: in our context, the trade-off between the two characteristics is perceived differently by different potential spouses, which is precisely what index models forbid. An equally important implication is that our surplus function, although supermodular in the continuous indices, does *not* satisfy the 'twist' condition of Levin (1993) and Chiappori, McCann and Nesheim (2010), which generalizes the standard Spence-Mirrlees property to a multidimensional framework. It follows that the stable matching needs not be pure: (an open set of) identical agents may be matched with different mates.[1]

We first analyze, as a benchmark, a fully symmetric version of the model, in which male and female characteristics play the same role in the surplus function and are identically distributed. We show that the stable matching exhibits perfect segregation: only individuals with the *same* discrete attribute marry each other (and matching within each class is perfectly assortative on the continuous trait). We then move to the general case. We prove existence and generic uniqueness of the stable matching and present a general resolution strategy. Our approach relies on 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. This strategy is not specific to our particular model (in which only two characteristics exist, one of which is discrete and the other continuous); it can be used in any discrete or continuous multidimensional framework.

A closed form characterization of the stable matching can be derived for specific settings. For instance, we provide it (in supplementary material) when the surplus is quadratic and distributions are uniform. Closed-form characterizations typically do not obtain in the general case. Still, we derive general predictions on the form that the stable matching may take. For individuals at the top of the continuous distribution, matching must be assortative on both

[1] 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.

smoking and SES. However, this pattern does not hold for lower status individuals. While, *conditional on both spouses' smoking status*, assortative matching on SES still prevails, people with different smoking habits may match; moreover, such 'cross-matches' typically take the form of randomization, whereby some individuals are indifferent between a smoking and a nonsmoking mate, and may marry either with positive probability. We show, however, that the two possible types of randomization are locally mutually exclusive: if, at some given SES level, smoking men and nonsmoking women marry with positive probability, then in the same neighbourhood nonsmoking men cannot marry smoking women (although the opposite pattern may prevail in other regions of the SES distribution). Lastly, we show how the surplus allocation between spouses can be recovered from the stability conditions. This is crucial in the perspective of endogeneizing smoking status, because incentives to quit smoking are altered by the consequences of quitting, including in terms of marital prospects; in particular, a matching model allows to exactly compute the 'marital cost' of smoking for each gender and each SES level.

Finally, we further specify the model by assuming that, for any SES level, there are more male smokers than female smokers among individuals above that status (what we call the 'prevalence constrained case'). This assumption, which is strongly supported by the data, generates additional predictions. One is that randomization can take only one form; specifically, some female nonsmokers and some male smokers randomize (in the sense that they may be matched with either a smoking or a nonsmoking partner with positive probability), but neither female smokers nor male nonsmokers do. Moreover, non-smoking wives married to a smoking husband should have on average 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 wives are on average of lower 'quality', i.e., lower SES); and 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.

Taking the model to the data in a structural way is far beyond the scope of this paper. There is little doubt that in real life the set of characteristics on which people match is large, and that many (if not most) of these characteristics are unobservable to the econometricians. An existing, econometric approach to matching models, pioneered by Choo and Siow (2006)

and Chiappori, Salanié and Weiss (2011), is based on a formalization of this intuition. However, extending this methodology to multidimensional setting with discrete and continuous characteristics is still an open question (particularly in the presence of randomization), despite recent and promising advances (Dupuy and Galichon 2012). More modestly, we test the predictions just described in reduced form; 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, which 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.[2] Focusing on young couples, we first document that both the absolute (and the cumulative) number of smokers at each year of education (our proxy for SES) is always higher for men than for women, regardless of their marital status. We then 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. While 'mixed' couples where the wife smokes and the husband does not can actually be found, they are much less frequent than vice versa (6.50% versus 11.71%, i.e., a ratio of 0.55, significantly lower than what would be implied by the sole difference in relative smoking prevalence).

Our regression analysis confirms that, among couples with identical smoking habits, matching is assortative on education. 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, that is, smoking husbands who marry non-smoking wives 'marry down' on SES; conversely, among smoking husbands those who marry smoking wives exhibit on average 0.16 *more* years of completed education than those with non-smoking wives, that is, smoking wives who marry smoking husbands 'marry up' on SES. 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 one of our theoretical predictions.

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

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 presentations. The precise connection between matching models and optimal transportation has been analyzed by Gretsky, Ostroy, and Zame (1999) in the discrete case, and by Gretsky, Ostroy and Zame (1992), Ekeland (2010) and Chiappori, McCann, and Nesheim (2010) in the continuous one. The 'twist condition', which generalizes Spence-Mirrlees to a multidimensional setting, was initially introduced by Levin (1999); the version used here is in Chiappori, McCann, and Nesheim (2010). Regarding the theoretical analysis of multidimensional matching models, Chiappori and McCann (2013b) consider the case of continuous characteristics, and derive a system of Partial Differential Equations that fully characterize the stable match; note, however, that their approach does not apply to our case, in which one characteristic is discrete. Low (2013) studies a matching model in which one side of the market (women) is characterized by two characteristics, whereas men traits are one dimensional. Lastly, McCann, Shi, Siow and Wolthoff (2012) introduce a multisector analysis in a bidimensional framework with homogeneous preferences. Finally, there exists a close connection between optimal transportation theory and principal-agents models under asymmetric information (e.g., Carlier 2001, 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.

Of the few existing empirical studies using a multidimensional context, most use either a search model (Wong 2003a,b, Coles and Francesconi 2011) or a NTU matching framework on dating patterns (Hitsch, Hortaçsu and Ariely 2010, Banerjee, Duflo, Gathak and Lafortune 2013). While the framework in Chiappori, Oreffice and Quintana-Domeque (2012) is a priori multidimensional, they concentrate on settings in which the several dimensions can be summarized by a single index - so that, technically, the matching is again one-dimensional, although the (index) dimension is unknown and has to be estimated. Less work has been

devoted to 'truly' multidimensional matching, with some notable exceptions. Galichon and Salanié (2010) analyze the case of a product-separable surplus function in the TU framework from an empirical perspective. Recently, Dupuy and Galichon (2012) have extended Choo and Siow (2006) and Galichon and Salanié (2010) by considering continuous characteristics in a TU econometric framework. Moreover, there exists a natural relationship between matching and hedonic models, for which identification issues are studied by a recent literature (e.g., Ekeland, Heckman and Nesheim, 2004; Heckman, Matzkin and Nesheim, 2010; and Nesheim 2012 for a multidimensional approach).

Regarding our empirical application, assortative marriage by smoking habits has been previously and extensively documented in the medical and biological literatures (e.g., Sutton, 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. 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. In addition, and regarding the potential concern of endogenous smoking behavior to marital status, Banks, Kelly and Smith (2013) use retrospective evidence from the Health and Retirement Study to show that most smoking behavior is initiated before marriage.

The next Section contains the model. Section 3 analyzes the implications of endogeneizing the binary attribute (i.e., smoking). Section 4 is devoted to the empirical application. Finally, 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 socioeconomic status (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 cumulative distribution of female (male) characteristics (x,X) $((y,Y))$ over the set $[0,1]\times\{N,S\}$, and let $F_X(x)$ $(G_Y(y))$ denote the number of females (males) with smoking status X (Y, with $X,Y \in \{N,S\}$) and SES no larger than x (y). In particular, $F_X(1)$ and $G_Y(1)$ respectively denote the total number of females and males with smoking habit X and Y.

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 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:

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

The function S is strictly increasing, continuously differentiable and supermodular, and satisfies $S(0, 0) = 0$; moreover, $\lambda < 1$. A simplifying assumption is that the impact on marital surplus of the spouses' smoking habits is fully summarized by a single parameter λ, which represents the decrease in surplus due to the presence of (at least) a smoker in the couple. In our framework, thus, the surplus of a mixed (smoker-nonsmoker) couple is the same as that of a couple of smokers, but strictly less than for a non-smoking pair. While this assumption is not necessary (a model with more complex impacts could still be solvable using the techniques described below, albeit at the cost of significant additional complexity), we believe it is appropriate in view of the empirical application under consideration. Indeed, the main effect of smoking is to reduce life expectancy − a feature that is reflected by the multiplicative nature of the impact (intuitively, smoking reduces the present discounted value of future welfare proportionally). This is true of the *direct* effects of own smoking, but also of the *indirect* (second-hand) effects of having a smoking spouse; indeed, 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). Moreover, 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 much less likely than non-smokers to be bothered by secondhand smoke exposure (e.g., Pilkington et al., 2006).

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:

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

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

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

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 maximiza-

tion of aggregate surplus over all possible assignments.[3] Formally, if $(\mu, u_N(x), u_S(x), v_N(y), v_S(y))$ is a stable matching, then the measure μ solves

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

where $\mathcal{M}$ denotes the set of measures on the set $([0,1] \times \{N,S\})^2$ 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:

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

under the constraints in (1). If $(\bar{u}_N, \bar{u}_S, \bar{v}_N, \bar{v}_S)$ denotes a solution, then $(\bar{\mu}, \bar{u}_N, \bar{u}_S, \bar{v}_N, \bar{v}_S)$ 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 $\rho : [0,1] \times \{N,S\} \to [0,1] \times \{N,S\}$, so that almost all agents (x,X) are matched with probability one to exactly one agent $(y,Y) = \rho(x,X)$. 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 gen-

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

eralization of supermodularity is the 'twist' condition, which is sufficient to prove purity of the stable match.[4] 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 $\partial S(x_0, y_1)/\partial x$ if she marries a non-smoker with SES y_1, and $\lambda \partial S(x_0, y_2)/\partial x$ if she is matched with a smoker with SES y_2. While S is strictly supermodular, we may still have that:

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

with $y_2 > y_1$ since $\lambda < 1$. It follows that the stable matching may not be pure in our setting.

2.3 The symmetric case

2.3.1 Characterization

We first consider, as a benchmark, a setting that is totally symmetric between genders. This implies that (i) male and female characteristics play the same role in the surplus function ($S(x, y) = S(y, x)$) 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. Smokers only marry smokers, and non-smokers only marry non-smokers; moreover, in each couple, spouses have the same 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.* ■

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

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 segregated 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 segregation, 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, particularly in terms of segregation.

In addition, one can easily compute each spouse's utility at a stable matching. 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

$$u_N(x) + v_N(x) = S(x, x)$$

By symmetry:

$$u_N(x) = \frac{S(x,x)}{2}, v_N(y) = \frac{S(y,y)}{2} \text{ and similarly}$$
$$u_S(x) = \lambda\frac{S(x,x)}{2}, v_S(y) = \lambda\frac{S(y,y)}{2}$$

The preceding conclusions, however, heavily rely on the very specific features of the symmetric framework. In particular, because of the assumed symmetry of the SES distributions across genders, no trade-off exists between the two characteristics at the stable matching: there is no point, for a non-smoker, in considering a smoker as a spouse, since a non-smoker with exactly the same SES is always available at equilibrium.

2.4 The general case: qualitative results

We now consider 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. In this subsection, we

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 exact form of the (supermodular) surplus.

2.4.1 Assortative matching among couples with identical smoking habits

A first result expresses the fact that, at a stable matching, matching is positive assortative on SES among couples with identical smoking habits:

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*

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

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

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

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* $y_S \leq y' \leq 1$,

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

Proof. *See the Appendix* ■

Proposition 3 states that non-smokers with high enough 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. Similarly, smokers with high enough SES only marry smokers.

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 lower 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 higher 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;[5] ultimately, the matching patterns depend on the distributions F and G.

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

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

  which defines the function ϕ_N.

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

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

  which defines ψ_N.

[5] Examples can readily be constructed, for instance by using disconnected subpopulations among both men and women.

- For the other combinations of smoking status:

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

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

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. Here, we 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:

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

The first integral considers the contribution of the female non-smoker 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 $\lambda S(t, \psi_N(t))$; alternatively, she may (with probability $1 - p_N(t)$) be matched with a non-smoker with SES $\phi_N(t)$, generating a surplus $S(t, \phi_N(t))$. Similarly, the second integral represents the contribution of the female smoker 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 Recovering 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-smoker wife with SES x. If her husband is a non-smoker with SES $\phi_N(x)$, then stability implies that:

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

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

$$u'_N(x) = \frac{\partial S}{\partial x}$$

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

$$u_N(x) = \int_0^x \frac{\partial}{\partial x} S(t, \phi_N(t))\, dt + K$$

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.

In particular, it becomes possible to compute the difference $u_N(x) - u_S(x)$, which can be interpreted as the cost of smoking on the marriage market − or equivalently as the gain that would result from quitting − for a woman with SES x.

2.6 The constrained general case

We now introduce an additional assumption, namely that "men smoke more than women," in the following sense:

Assumption MP (Male Prevalence) *For any SES level x,*

$$F_S(x) \geq G_S(x)$$

In words, for any SES x, there are more male than female smokers with SES x or larger. This may describe a situation in which men are more likely to smoke, but also one in

which men and women have similar smoking habits but the male SES distribution first degree dominates the female one, as well as many others. As we shall see in the empirical part, the MP assumption is satisfied in the data. Under that assumption it is possible to characterize the stable matching as follows:

Proposition 5 *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 $y_S \leq y' \leq 1$,*

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

- *for almost all $x < x_N, x' < x_S, y < y_N$ and $y' < y_S$,*

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

- *for any x such that $p_N(x) > 0$, let $\phi_N(x)$ and $\psi_N(x)$ respectively denote the nonsmoker and smoker mate of (x, N). Then*

$$\psi_N(x) > \phi_N(x)$$

- *for any y' such that $q_S(y') < 1$, let $\phi_S(y')$ and $\psi_S(y')$ respectively denote the nonsmoker and smoker mate of (y', S). Then*

$$\phi_S(y') = \psi_S(y')$$

The proof is in the Appendix. In practice, Proposition 5 first states that high SES people are exclusively matched to partners with identical smoking habits –a result already derived by Proposition 3. For lower SES, randomization may happen, as Proposition 4 indicates, *but only in one direction*: non-smoking women may marry smokers, but smoking women and non-smoking men always marry their own. Lastly, when a female non-smoker may marry both a smoker and a non-smoker with positive probability, then the smoker has a higher SES.

A remark coming directly from the optimal control formulation sheds some light on the form of the randomization. The Hamiltonian corresponding to the program (8) is linear in $p_N(x)$, the coefficient being equal to $\lambda S(x, \psi_N(x)) - S(x, \phi_N(x))$, which is always non positive. For large values of x, $\psi_N(x)$ and $\phi_N(x)$ are both close to 1, and since $\lambda < 1$ the coefficient is negative, implying $p_N(x) = 0$. Randomization requires the coefficient to be zero; therefore, whenever randomization takes place, it must be the case that

$$S(x, \phi_N(x)) = \lambda S(x, \psi_N(x))$$

(which, incidentally, immediately implies $\psi_N(x) > \phi_N(x)$). Since both $\phi_N(x)$ and $\psi_N(x)$ depend on the probability $p_N(x)$ (by 4 and 5), $p_N(x)$ must be such that this equality holds over the relevant domain.

Finally, we note that *where exactly* randomization occurs, and *with which probability*, depends on the exact form of the surplus function and the two distributions. As an illustration, we provide (in the supplementary material) an example where (i) SES are uniformly distributed on $[0, 1]$, (ii) smoking status and SES are independent, (iii) men are more likely to be smokers than women at all SES levels and (iv) the surplus S is quadratic and symmetric. In that case, the program can be solved in closed form solution. There is a threshold below which non-smoking women randomize their mate; moreover, the probability does not depend on the SES. Obviously, more complex patterns can be constructed by varying the distributions of characteristics.

2.6.1 Recovering individual utilities

We now compute individual utilities at the stable matching, as a function of SES, gender and smoking habit. Specifically:

Proposition 6 *Under the MP assumption, individual utilities* $u_N(x), u_S(x), v_N(y), v_S(y)$ *are all increasing. For all* x *and all* y*,*

$$u_N(x) \geq u_S(x) \text{ and } v_N(y) \geq v_S(y)$$

Moreover, whenever $p_N(x) > 0$*, then*

$$u_N(x) = u_S(x)$$

Lastly, for x *and* y *large enough, the differences* $u_N(x) - u_S(x)$ *and* $v_N(y) - v_S(y)$ *are increasing (in* x *and* y *respectively).*

The proof is in the Appendix. In short, utility increases with the person's SES, and smokers are never better off than non-smokers. However, whenever randomization takes place, a woman's welfare does not depend on her smoking status: both smokers and non-smokers marry a smoking husband with positive probability, with whom they generate the same surplus. For men, on the other hand, welfare is always smaller for smokers. Finally, at the top of the SES distribution, the cost of being a smoker increases with social status. In our uniform-quadratic example, things are even simpler: the cost, for a woman, of being a smoker (as measured by the difference $u_N(x) - u_S(x)$) is zero below the randomization threshold and increases with SES above it; for men, it is always positive and always increasing.

3 Endogenous smoking

So far, we have assumed that the smoking status was exogenous. While this assumption may be an acceptable approximation (Banks, Kelly and Smith (2013) show that most smoking behavior is initiated before marriage), an interesting implication of our approach is that it naturally leads to endogenizing smoking behavior. Suppose that agents are ex ante characterized by their propensity to smoke, which is exogenous and randomly distributed over the population (possibly with a positive mass at zero); whether they decide to smoke or not depends both on their propensity and on the expected benefits that quitting (or not starting) smoking would generate on the marriage market. Technically, the situation may be modeled as a two stage game: agents (non cooperatively) decide whether to smoke in the first stage, which generate an SES distribution by gender and smoking habit, and play the matching game described above in the second stage.

A two stage game of this type may be quite difficult to solve, if only because it in principle requires a fixed point argument in a space of distributions. Agents must, in stage one, anticipate the distributions F and G that will prevail at the beginning of stage two, in order to compute their equilibrium utility as a smoker and as a nonsmoker; the difference (the equilibrium cost of smoking) motivates their first stage decisions, which in turn generate the actual distributions F and G, and an equilibrium is reached if the initial expectations are self-fulfilling. In our context, however, a simpler approach exists. Consider a matching game in which agents are characterized by two continuous traits, namely their SES and their propensity to smoke; once matched, they jointly choose their smoking habits so as to maximize surplus. That these two settings lead to the same outcome was initially remarked by McCann et al. (2012); see Chiappori and McCann (2013a) for a detailed analysis.

Even with this simplification, a full model of this type is quite complex, and typically does not have an analytical solution (although numerical resolution using linear programming is always possible, cf McCann et al. 2012); moreover, the detailed properties of the stable match depend on the form of both the surplus function and the joint distribution of SES and smoking propensity. However, some common characteristics can be described. First, even if the (ex ante) distribution of smoking propensities is independent of SES, smoking will be ex post correlated with SES; indeed, the benefits of quitting (as measured by the

difference in expected utility, $u_N(x) - u_S(x)$ for women and $v_N(y) - u_S(y)$ for men) vary with the SES. Second, we expect this difference to be increasing in SES (for higher SES in general, and always in the uniform quadratic example); therefore, independence between SES and propensity would lead to a negative correlation between smoking and SES. Third, the difference is positive for high SES women (who do not randomize) but may be zero for lower SES ones (who may, and do in the uniform quadratic case, randomize). Therefore, we expect that the proportion of smokers among women should not vary much with SES at the bottom of the SES distribution, but should drop significantly at the top, while patterns for men should be more gradual.

4 Empirical Application

Our empirical application will use years of education as a measure of SES. The discrete (and dichotomic) attribute we consider is smoking status.

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. As we have already mentioned, 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).

A second intriguing aspect 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).

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 ret-

rospective 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). Although many mechanisms can be invoked to explain these patterns, they certainly fit the predictions of the previous endogenous smoking section.

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 largest samples of married couples for whom information on tobacco use is available. 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.

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.[6]

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

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.[7] 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 aged 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 sample is large, with detailed socioeconomic and smoking information on both spouses, 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

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

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.[8] 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 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.[9] 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 definition, 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 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

[8] In the sensitivity analysis subsection, we will recode the education variable following Jaeger (1997).
[9] The 2000 TUS Supplements do not record the question about the number of cigarettes smoked.

structure.[10] More concretely, if the MP (Male Prevalence) assumption is satisfied, then 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 N-S).

Actually, the MP assumption by itself implies an asymmetry between the proportions of 'mixed couples': if there are more male smokers than female, then one expects more N-S couples. To take an easy benchmark, assume that the proportion of smokers in the male (resp. female) population is equal to s_M (resp. s_W) at all SES levels; then if matching was random with respect to smoking habits, the ratio of S-N to N-S couples should be

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

In our sample, s_M is .22 and s_W is .17, so the ratio r implied by the sole difference in relative smoking prevalence 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, or both non-smokers, denoted N-N), matching should be assortative on SES.

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

Prediction 4: When two non-smoker women with the same (low) SES marry respectively a smoker and a non-smoker, the non-smoker should be on average of lower SES than the smoker. That is, controlling for the (nonsmoking) wife's SES, the smoking habit of the husband should be *positively* correlated with his SES. This result seems counter intuitive, since a well-established empirical fact is that the (unconditional) correlation

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

between smoking and education is *negative*: low SES is an important predictor of smoking behavior (e.g., CDC, 2010). In the theoretical framework we use, however, (perfect) assortative matching on SES implies that, for given smoking habits, the husband's SES is fully determined by his wife's. In reality, agents match on several characteristics (many of them unobservable), so that the wife's SES is not a perfect control for the husband's, meaning that the (unconditional) negative correlation between smoking and SES is likely to persist. We therefore restate 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. Note that this pattern should not hold true for women.

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

4.4 Empirical Results

Descriptive statistics Tables 1 and 2 display the summary statistics of married and single individuals, and the corresponding ones by smoking type of couple. Table 1 is consistent with the gender gap reported by the National Center for Health Statistics (NCHS, 2010): 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). Table 2 shows that women are more educated than men in almost all types of couples, and that we are actually indeed focusing on young couples (and individuals), since the average number of children under six years old is about 0.80 for couples and 0.50 for never-married individuals. The health status is very similar across spouses, and higher in couples where none is smoking than in those where both spouses are smoking.

[Table 1 about here]

[Table 2 about here]

Testing the male prevalence assumption We compute the number of smokers by gender and educational attainment (years of education) using two samples: one of married individuals (our main sample), and an augmented sample which includes never-married inviduals as well. Figures 1 and 2 show that the number of male smokers is always higher than the number of

female smokers for each educational attainment. Analyzing the cumulative number of smokers by education and gender for both samples yields a spectacular pattern: the plotted cumulative curve of the number of male smokers is *always* above the cumulative curve of the number of female smokers, for each educational attainment (years of education), and for both types of samples, as shown in Figures 3 and 4. The patterns displayed in all these figures support the MP assumption.

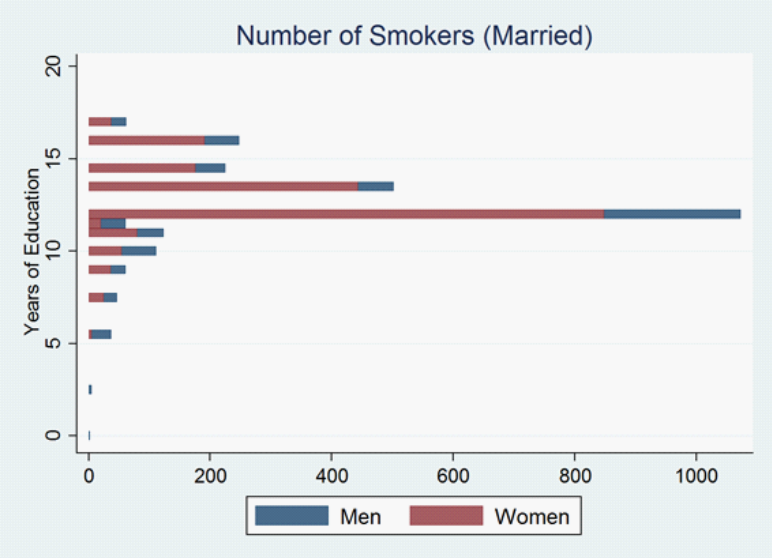


Figure 1: Number of smokers by years of education. Married men and women.

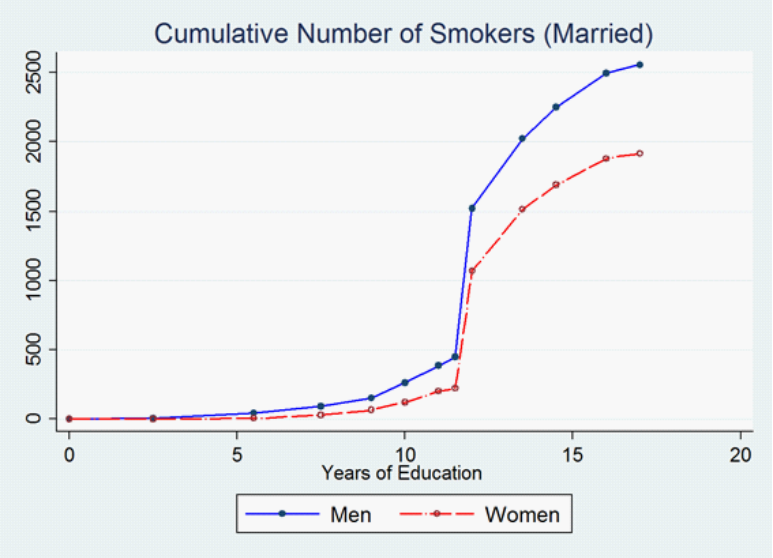


Figure 2: Cumulative number of smokers by years of education. Married men and women.

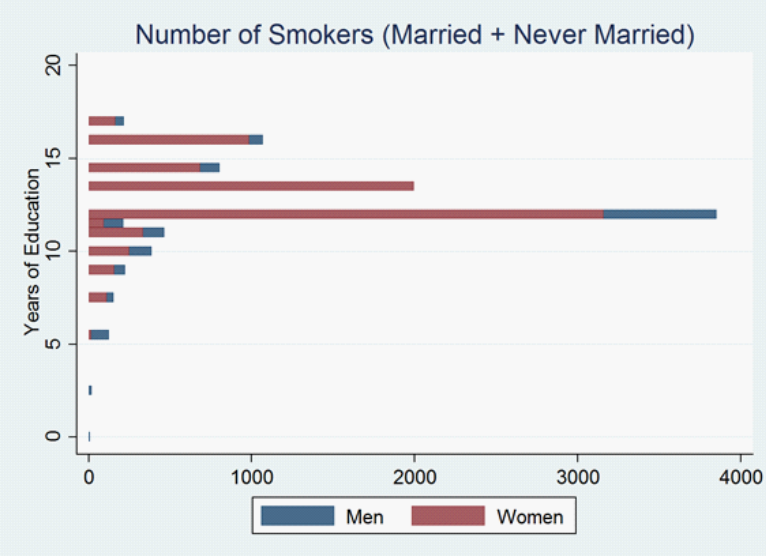


Figure 3: Number of smokers by years of education. Married and never-married men and women.

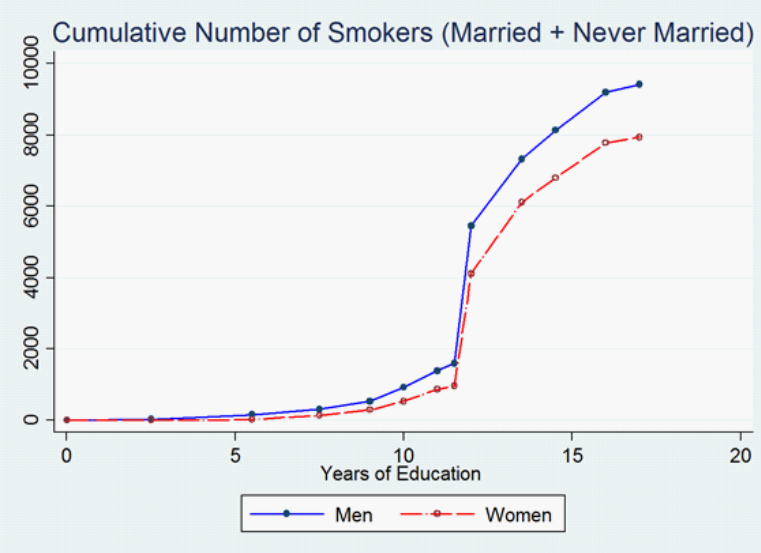


Figure 4: Cumulative number of smokers by years of education. Married and never-married men and women.

Matching patterns and correlation between smoking and education 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 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). In addition, *prediction 1* is empirically supported: 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.[11] Indeed, these figures are very different from the prevalences arising from "random matching" in Panel B. 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.

[Table 3 about here]

In Table 4 we investigate the correlation between smoking and education. For both men and women smoking status and education are negatively correlated, although the correlations appear to be different by gender, with the male gradient being significantly larger (in absolute value) than the female one. These findings are in line with the literature on smoking and education (see De Walque, 2010). In addition, these patterns are present for both married and never-married individuals.

[Table 4 about here]

Positive assortative matching on education Table 5 presents evidence of positive assortative matching by education among couples with identical smoking habits. 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 matching by education. Although assortative mating by education has been extensively documented in the literature (Lam, 1988; Pencavel,

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

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*.

[Table 5 about here]

"Marrying up" or "marrying down" *Prediction 3* is investigated in Table 6. It contains a series of regressions 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: smoking husbands, who are on the long side of the market, 'marry down', that is, they 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).

[Table 6 about here]

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 partners with their own education. In other words, smoking women 'marry up', they 'benefit' from the fact that they are on the short side of the market by marrying a higher 'quality' spouse. Note that the sign of the coefficient in column (4) is opposite to the well-known negative gradient between own smoking and own education.

The evidence presented in Table 6 is also supportive of our assumptions on the surplus reduction due to smoking. If, for instance, smokers also preferred 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.

Unconditional versus conditional correlations between smoking and education Finally, we investigate our last prediction, *prediction 4*, 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: among non-smoking low-quality women, their smoking husbands are "more" educated. In addition, columns (3) and (4) show that this pattern does not hold for the wives of smoking men.

[Table 7 about here]

4.5 Sensitivity Analysis

4.5.1 Additional controls, alternative definitions, and different age groups

We proceed to a few robustness checks. First, we assess the role of omitted "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 (results available on request). 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.[12]

4.5.2 Alternative data sets

To further explore the issue of recent marriages, we alternatively examine a very different data source which provides 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 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.

[12] 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.

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.[13] 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.[14]

[13] 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.

[14] 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 develop a bidimensional matching model under transferable utility on the marriage market, where individuals are characterized by a continuous trait (e.g., socioeconomic status) and a binary attribute (e.g., smoking status) which is heterogeneously (dis)liked in the population: individuals with the binary attribute (equal to one) do not care about the spousal binary attribute, while those without it (equal to zero) are negatively affected. In the benchmark, symmetric case, when male and female characteristics play the same role in the surplus function and the distributions are identical, the stable matching is pure: matching is perfectly assortative on the continuous trait, and perfect segregation obtains regarding the binary attribute. In the general case, the stable matching needs not be pure. We present a general resolution strategy based on optimal control theory. While a closed-form characterization is typically not available, we can derive general results: assortative matching on the continuous trait among couples with identical binary attributes, purity at the top of the distributions of the continuous traits, and possible randomization at the bottom of the distribution of the continuous traits. Under an additional (prevalence) assumption, two additional predictions arise: one regarding the randomization at the bottom of the distributions of the continuous trait; the other regarding the relationship between the continuous trait and the binary attribute.

This last case is not a theoretical curiosity. Actually, in our empirical application with US data from the CPS and its Tobacco Use Supplements for the years 1996 to 2007, male smokers outnumber female smokers across the distribution of education. We then empirically analyze the predictions of our model regarding matching patterns by smoking status, assortative matching on education, how individuals 'marry down' or 'marry up' on education, and the correlation between education and smoking. Remarkably enough, our predictions are borne out in the data. We find, for instance, that among smoking husbands those who marry smoking wives have on average 0.16 more years of education than those with non-smoking wives, so that smoking women, who are in the short side of the market, "marry up". Instead, among non-smoking wives those with smoking husbands have on average 0.14 fewer years of completed education than those with non-smoking husbands, so that smoking men, being in the long side of the market, "marry down".

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 homogeneously assessed in the population. 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.

APPENDIX

A Proof of Proposition 3

Take some small $\varepsilon > 0$ such that

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

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

$$S = \lambda s(x,y) + \lambda s\left(x', y'\right)$$

while matching x and y' would generate a surplus

$$S_1 = s\left(x, y'\right) + \lambda s\left(x', y\right)$$

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

$$\Sigma = s\left(x', y'\right) + \lambda s\left(x, y_S\right)$$

whereas mixing matches would generate

$$\Sigma_1 = s\left(x, y'\right) + \lambda s\left(x', y_S\right)$$

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

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

by supermodularity, which contradicts surplus maximization. Define

$$\phi\left(s\right) = s\left(x, s\right) - s\left(x', s\right)$$

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

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

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

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

therefore

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

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*

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

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:

$$u'(x) = \frac{\partial}{\partial x} s(x,y) = \lambda \frac{\partial}{\partial x} s\left(x,y'\right)$$

Since $\partial s/\partial x$ 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

$$S = \lambda s\left(x'', y\right) + \lambda s\left(x', y'\right)$$

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

$$S_1 = \lambda s\left(x', y\right) + \lambda s\left(x'', y'\right)$$

and $S_1 > S$ by supermodularity of s; 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

$$S = s(x, y) + \lambda s\left(x', \bar{y}\right)$$

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

$$S_1 = s(x, \bar{y}) + \lambda s\left(x', y\right)$$

Since

$$s(x, \bar{y}) - s(x, y) = s\left(x', \bar{y}\right) - s\left(x', y\right) > \lambda\left(s\left(x', \bar{y}\right) - s\left(x', y\right)\right)$$

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.

References

[1] Banerjee, A., Duflo, E., Ghatak, M., Lafortune, J. (2013) "Marry for What? Caste and Mate Selection in Modern India," *American Economic Journal: Microeconomics*, 5(2): 33–72.

[2] Banks, J., Kelly, E., Smith J.P. (2013) "Spousal health effects: the role of selection," in *Discoveries in the Economics of Aging*, D.A. Wise (ed)., NBER Book Series, University of Chicago Press.

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

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

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

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

[7] 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.

[8] Carlier, G. (2001). "A general existence result for the principal-agent problem with adverse selection", *Journal of Mathematical Economics* 35, 129-50.

[9] 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`

[10] 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

[11] 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

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

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

[14] 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.

[15] Chiappori, P.-A., and McCann, R. (2013a) "Investing in matching", mimeo, Columbia University

[16] Chiappori, P.-A., and McCann, R. (2013b) "Multidimensional matching: the continuous case", mimeo, Columbia University

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

[18] Chiappori, P.-A., Oreffice, S., Quintana-Domeque, C. (2012) "Fatter attraction: anthropometric and socioeconomic matching on the marriage market," *Journal of Political Economy*, 120(4):659–695.

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

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

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

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] Dupuy, A., Galichon, A. (2012) "Personality traits and the marriage market," IZA Discussion Paper No. 6943.

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

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

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

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

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

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

[33] 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.

[34] Gretsky,N., J. Ostroy, and W. Zame (1992). "The nonatomic assignment model", *Econonic Theory* 2, 103-27

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

[36] 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.

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

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

[39] 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.

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

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

[42] 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.

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

[44] V. Levin (1999), "Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich Problem", *Set-Valued Analysis* 7, 7-32.

[45] Low, C. (2013),"Pricing the Biological Clock: Reproductive Capital on the US Marriage Market", Mimeo, Columbia University

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

[47] 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.

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

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

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

[51] 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.

[52] 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.

[53] 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.

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

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

[56] 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.

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

[58] 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.

[59] 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.

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

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

[62] 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.

[63] 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.

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

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

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

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

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

[69] 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.

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

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

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

[73] Wong, L. (2003b) "Why Do Only 5.5% of Black Men Marry White Women? ," *International Economics Review* 21(3):699–728.

Table 1: SUMMARY STATISTICS. MARRIED VS. 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 FOR 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: MATCHING PATTERS BY SMOKING STATUS

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%**
	(8,710)	(767)
Smoking Husband	**11.71%**	**10.01%**
	(1,408)	(1,150)

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: CORRELATION BETWEEN SMOKING AND EDUCATION

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***	−0.026***
	(0.001)	(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***	−0.030***
	(0.001)	(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: ASSORTATIVE MATCHING BY EDUCATION

Regression of Education on Spouse's 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	8,710	8,710	1,150	1,150
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	1,408	1,408	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: "MARRYING UP" AND "MARRYING DOWN"

Regression of Education by Smoking Status on Spouse's Education and Smoking Status

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	10,118	1,917	9,477	2,558
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: UNCONDITIONAL AND CONDITIONAL CORRELATIONS BETWEEN SMOKING AND EDUCATION

Regression of Smoking Status on Education (and Spouse's Education) for subsamples

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	10,118	10,118	2,558	2,558
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: ADDITIONAL CONTROLS

Regression of Education by Smoking Status on Spouse's Education and Smoking Status 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: ALTERNATIVE DEFINITION OF YEARS OF EDUCATION

Regression of Education by Smoking Status on Spouse's Education and Smoking Status 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: ALTERNATIVE AGE GROUPS

Regression of Education by Smoking Status on Spouse's Education and Smoking Status

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	7,861	1,506	7,252	2,115
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: MATCHING PATTERNS BY SMOKING STATUS

PSID DATA: 1999-2007

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

Weighed observations

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: "MARRYING UP" AND "MARRYING DOWN"

PSID DATA: 1999-2007

Regression of Education by Smoking Status on Spouse's Education and Smoking Status

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

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	2,035	350	1,873	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

Bidimensional Matching with Heterogeneous Preferences: Smoking in the Marriage Market.

SUPPLEMENTARY MATERIAL

1 A uniform-symmetric quadratic example

1.1 Specification

As stated in the paper, 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 in subsection 2.3. of the paper. 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:

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

While this form can be given micro foundations[1], we use it mostly for

[1] 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.

practical purposes: it allows the derivation of a closed form solution.

1.2 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, supermodularity implies that matching will be assortative; i.e., men with higher SES will marry wives with higher SES.

1.2.1 Marital patterns

Our main result is the following:

Proposition 1 *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.*

After the interpretation of this result, two proofs will be provided, one using the surplus optimization program and the other relying on a direct approach.

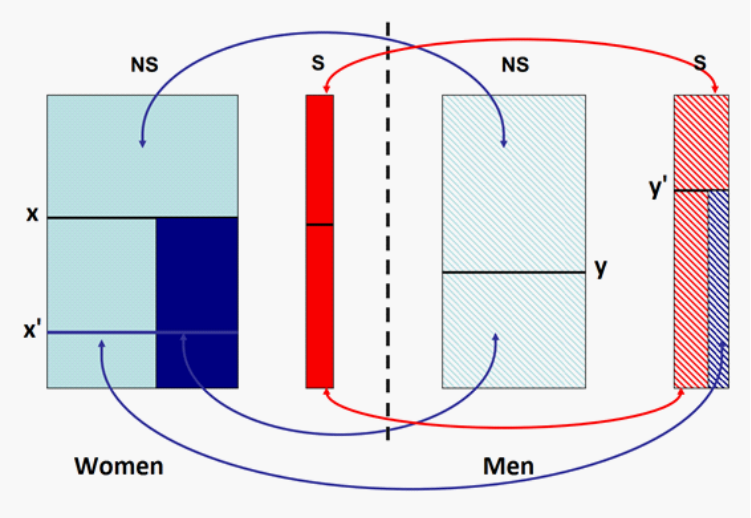


Figure 1A: Qualitative properties of the stable matching.

1.2.2 Interpretation

The stable matching is summarized in Figure 1A. It can be interpreted as follows. First, in the line of Proposition 1, 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. Lastly, 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.

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.

1.2.3 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.[2] Figures 2A and 3A 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 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

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

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.

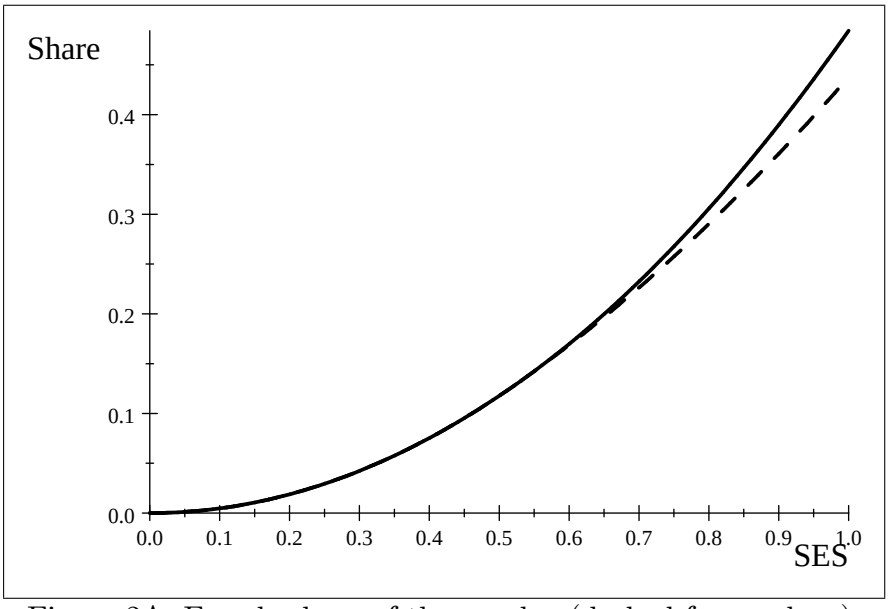


Figure 2A: Female share of the surplus (dashed for smokers).

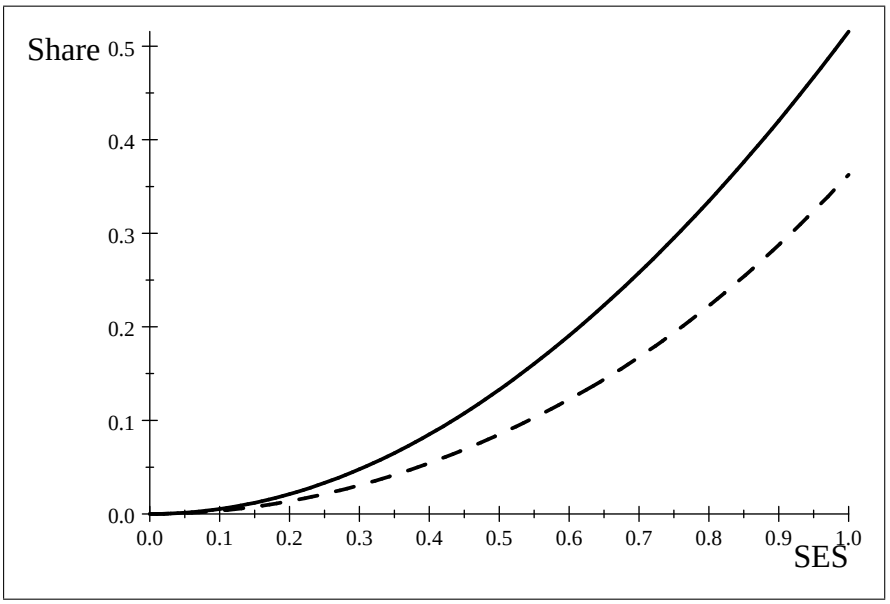


Figure 3A: Male share of the surplus (dashed for smokers).

1.3 Proof of Proposition 1: surplus maximization

We now provide two proofs of Proposition 1, aimed at illustrating the various ways the problem can be approached. We start with surplus

maximization. From the general characterization provided in the main paper, we know that $q_N(y) = 0$ and $p_S(x) = 1$ for almost all x and y, and that moreover $\psi_N(x) = \psi_S(x)$. Measure constraints become:

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

for non-smoking couples, and

$$s_W(1 - x) = s_M \int_{\psi_S(x)}^1 q_S(s)\, ds = s_M \int_{\psi_N(x)}^1 q_S(s)\, ds \tag{2}$$

for smokers. Finally, regarding mixed marriages:

$$(1 - s_W) \int_x^1 p_N(t)\, dt = s_M \int_{\psi_N(x)}^1 (1 - q_S(s))\, ds \tag{3}$$

which, using (2), becomes

$$(1 - s_W) \int_x^1 p_N(t)\, dt = s_M(1 - \psi_N(x)) - s_W(1 - x) \tag{4}$$

and in particular, adding up (1) and (4):

$$x = (1 - s_M)\phi_N(x) + s_M \psi_N(x) \tag{5}$$

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

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

and (4) 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:

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

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

$$S = (1-s_W)\int_0^1 (1-p_N(t))\,t\left(1-\frac{1-s_W}{1-s_M}(1-t-P(t))\right)dt$$
$$+(1-s_W)\,\lambda\int_0^1 p_N(t)\,t\left(1-\frac{1-s_W}{s_M}P(t)-\frac{s_W}{s_M}(1-t)\right)dt \qquad (6)$$
$$+\lambda s_W\int_0^1 t\left(1-\frac{1-s_W}{s_M}P(t)-\frac{s_W}{s_M}(1-t)\right)dt$$

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:

$$H(t,p_N,P) = (1-s_W)(1-p_N)\,t\left(1-\frac{1-s_W}{1-s_M}(1-t-P)\right)$$
$$+(1-s_W)\,\lambda p_N t\left(1-\frac{1-s_W}{s_M}P-\frac{s_W}{s_M}(1-t)\right) \qquad (7)$$
$$+\lambda s_W t\left(1-\frac{1-s_W}{s_M}P-\frac{s_W}{s_M}(1-t)\right)+\zeta p_N$$

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.

1.4 Proof of Proposition 1: 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.

1.5 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:

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

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

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

- 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

    $$p\widetilde{x} = \frac{s_M - s_W}{1 - s_W} \tag{8}$$

  - 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

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

    which, as above, implies that

    $$\widetilde{x} = \frac{q(s_M - s_W)}{p(1-s_W) - q s_W} \tag{9}$$

- 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

$$\widetilde{x} = \left(\frac{s_M}{s_W} - 1 \right) \frac{(1-q)}{q} \tag{10}$$

One can readily check that (8), (9) and (10) are redundant: if some numbers $(p, q, \widetilde{x})$ satisfy any two of them, they also satisfy the third.

1.6 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:

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

1.7 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 (8), (9) and (10), these conditions imply that:

$$\begin{aligned}\widetilde{x} &= \frac{(s_M + \lambda(1-s_M))(s_M - s_W)}{s_M(1-s_W) - \lambda s_W(1-s_M)} \leq 1\\ p &= \frac{s_M(1-s_W) - \lambda s_W(1-s_M)}{(s_M + \lambda(1-s_M))(1-s_W)} \leq 1\\ q &= \frac{s_M(1-s_W) - \lambda s_W(1-s_M)}{s_M} \leq 1\end{aligned} \tag{11}$$

The indifference condition also applies to male smokers; again, it leads to conditions (11).

1.8 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:

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

1.9 Comparative statics

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

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

1.10 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 $x \geq \widetilde{x}, y' \geq \widetilde{y}'$:

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

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

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

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.

1.11 Numerical simulations

In the numerical simulations, we take:

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

Then the functions take the following form:

$$u_N\,(x) = \begin{cases} 0.470\,59x^2 & \text{if } x < 0.531\,25 \\ 0.533\,33x^2 - 6.\,666\,7 \times 10^{-2}x + 0.017\,71 & \text{if } 0.531\,25 \leq x \end{cases}$$

$$u_S\,(x) = \begin{cases} 0.470\,59x^2 & \text{if } x < 0.531\,25 \\ 0.32x^2 + 0.16x - 4.\,250\,0 \times 10^{-2} & \text{if } 0.531\,25 \leq x \end{cases}$$

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

$$v_S\,(y) = \begin{cases} 0.34y^2 & \text{if } y < .625 \\ 0.5y^2 - 0.2y + 0.062\,5 & \text{if } 0.625 \leq y \end{cases}$$

References

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

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