Bidimensional Matching with Heterogeneous Preferences: Education and Smoking in the Marriage Market[*]

Pierre-André Chiappori   Sonia Oreffice   Climent Quintana-Domeque

Columbia University   University of Surrey and IZA   University of Oxford and IZA

July 2016

Abstract

We develop a frictionless matching model under transferable utility where individuals are characterized by a continuous trait and a binary attribute. The model incorporates attributes for which there are heterogeneous preferences in the population regarding their desirability, that is, the impact of the traits cannot be summarized by a one-dimensional attractiveness index. We present a general resolution strategy based on optimal control theory, and characterize the stable matching. We then consider education and smoking status, further specify the model by observing that there are more male than female smokers above each education level, and derive additional predictions about equilibrium matching patterns and how individuals with different smoking habits "marry down" or "marry up" by education. Using the CPS March and Tobacco Use Supplements for the period 1996 to 2003, we find that the hypotheses based on our model predictions are borne out in the data.

Keywords: multidimensional matching, marital surplus, smokers, socioeconomic status, transferable utility.
JEL Codes: D1, J1.

1 Introduction

Empirical evidence strongly suggests that matching processes in the marriage market are multidimensional. Spouses tend to be similar in a variety of characteristics, including age, body mass index, education, race, religion, and smoking status (Becker, 1991; Oreffice and Quintana-Domeque, 2010; Qian, 1998; Silventoinen et al., 2003, Sutton, 1980; Weiss and Willis, 1997). New developments in the matching literature have started to address the multidimensional nature of attractiveness by considering settings with two or more attributes. In particular, several recent studies[1] consider frameworks where multiple characteristics can be summarized by a single, one-dimensional attractiveness index, so that, technically, the matching process is de facto one-dimensional. Such a single-index approach is a powerful tool, and it allows to apply the standard one-dimensional matching techniques to settings with multiple attributes. However, its validity relies on a strong homogeneity assumption: it must be the case that the trade-offs between individual female traits are identically perceived by all men (and similarly for male characteristics). This approach, appealing as it may seem, is not always appropriate.

When preferences for any of the relevant matching characteristics are *heterogeneous* within either the male or the female populations (or both), matching becomes multidimensional, as long as the traits at stake are not perfectly correlated. That is, the corresponding traits cannot be collapsed into a single index, and one-dimensional matching techniques are not well-suited to characterize the corresponding stable matches. In the marriage market, several characteristics can be heterogeneously assessed by potential mates, including race, ethnicity, age, health attributes, or smoking status. For instance, we know that smokers are more likely to marry smokers (Clark and Etilé, 2006; Maralani, 2009; Sutton, 1980; Venters et al., 1984), and that a large medical and public health literature essentially shows that non-smokers mind their partner's smoking status, while smokers do not.[2]

The goal of this paper is to extend the standard matching model under Transferable Utility to a particular case of multidimensional matching, where one of the characteristics is assessed *heterogeneously* by potential mates. In particular, we present a general resolution strategy

---

[1] See for instance Coles and Francesconi (2013) for a search Non-Transferable Utility (NTU) approach, Chiappori, Oreffice and Quintana-Domeque (2012) for a general investigation.

[2] There is ample medical evidence that second-hand smoking has detrimental health effects on non-smokers but not on smokers (ASH, 2011; CDC, 2006; Glymour et al., 2008; Mannino et al., 1997). In addition, the attitude of smokers towards smokers is much more permissive than that of non-smokers (ASH, 2011; Lader, 2009; Pilkington et al., 2006).

and characterize the stable matching when individuals differ in two characteristics.[3] We also suggest a new way of testing matching models. In most one-dimensional matching models, theoretical implications regarding who marries whom are straightforward; essentially, they boil down to supermodularity implying assortative matching. In multidimensional contexts, however, things are much more complex, since theory may imply specific properties on matching patterns along the various dimensions under consideration. We argue that such predictions may be taken to the data and tested in reduced form. Specifically, in this paper we study a parsimonious model and derive a series of predictions. We then test hypotheses based on these predictions, and find they are indeed supported by the data. While our hypotheses tests can be formally derived from several specific stochastic structures found in the literature, we argue that they do not require explicit assumptions on the stochastic structure of unobservables; therefore, and in the same spirit as previous tests of index-based models (e.g., Chiappori, Oreffice and Quintana-Domeque, 2012), such tests can be viewed as a useful complement to more structural approaches.

In our model, individuals are characterized by two dimensions. 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. We suggest to interpret the second characteristic as the individual's smoking status. The marital surplus function is assumed to be differentiable and supermodular in the continuous indices, as is common in the literature, and to be multiplicatively impacted by the discrete characteristic: it may be diminished by the presence of a smoker in the couple. As long as smoking and socioeconomic status are not perfectly correlated (and they are not), this heterogeneity in preferences is a key feature of our setting: it rules out a single-index representation, since the trade-off between the two characteristics is perceived differently among potential spouses, which is precisely what index models forbid.

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 dis-

---

[3]Among multidimensional empirical works, Lindenlaub (2014) considers a specific model in which the distribution of types is normal and the payoff is quadratic to study multidimensional sorting between workers and jobs in the labor market, while Galichon and Salanié (2010) and Dupuy and Galichon (2014) use a Choo and Siow (2006) framework. For the analysis of dating patterns using a Gale-Shapley NTU approach, see Banerjee et al. (2013) and Hitsch, Hortaçsu and Ariely (2010).

tributed. We show that the resulting stable matching exhibits full segregation, in the sense that smokers exclusively marry smokers. This result is interesting per se, since it does not depend on the magnitude of the negative impact of smoking. In our "pure" frictionless framework, therefore, even minor differences in preferences (or surplus) may result in large-scale segregation; moreover, the latter is efficient, in the sense that it does maximize total welfare.

We then consider the general asymmetric case. We prove existence and generic uniqueness of the stable matching and present a general resolution strategy, which can be used in any discrete or continuous multidimensional framework. Even in our simple framework, a closed-form characterization of the stable matching cannot be obtained in the general case; however, we derive general predictions on the form that the stable matching may take. The analysis of a specific, quadratic case in which a closed form solution exists is provided in our Supplementary Material.

Next, we further specify the model by assuming that, for any SES level, there are more male smokers than female smokers among individuals above that SES level, a pattern that we actually observe in the data. This generates a set of additional predictions regarding the nature of the stable matching. First, there are *no* 'mixed' couples in which she smokes and he does not, whereas the opposite pattern –he smokes and she does not– happens with positive probability. Second, smoking husbands married to smokers are of *higher* "quality" (i.e., higher SES) than those married to non-smokers: smoking "premium" for smoking wives. Third, and conversely, non-smoking wives married to smoking husbands have a *lower* SES than those married to non-smokers: smoking "penalty" for smoking husbands. Fourth, there is positive assortative matching on SES among couples with identical smoking habits. Fifth, positive assortative matching on SES by smoking status is stronger at the top of the SES distribution: smoking husbands (wives) at the top of the SES distribution are more likely to marry smoking wives (husbands) than those with lower SES. Similarly, non-smoking husbands (wives) at the top of the SES distribution are more likely to marry non-smoking wives (husbands) than those with lower SES. Finally, some female non-smokers may marry either a smoker or a non-smoker with positive probability; then it must be the case that the smoking husband has a higher SES. Similarly, some male smokers may marry either a smoker or a non-smoker; but then both women have the *same* SES.

We use the Current Population Survey March Supplements data combined with the To-

3

bacco Use Supplements (TUS) for the period 1996 to 2003. These TUS supplements, which are widely used in medical research on smoking, provide the largest representative sample of the US population, and, crucially, allow us to retrieve information on *both* spouses.

We follow a reduced-form approach to test the hypotheses based on the theoretical predictions on the stable matching derived from our deterministic model, acknowledging that our general model could not be solved under a general stochastic structure. Our approach is compatible with a Choo-Siow specification, that is, a framework where unobserved heterogeneity can be fully captured by an additive random term, which is moreover the sum of two random variables with type 1 extreme value distributions; but it could also be used in alternative settings.[4] In our deterministic model, stability conditions imply that there should be no "mixed" couples in which she smokes and he does not, whereas the opposite pattern –he smokes and she does not– should happen with positive probability (the asymmetry between genders being due to the larger prevalence of male smoking). In a more complex setting, the presence of either search frictions (à la Shimer and Smith, 2000) or unobserved matching characteristics (or both) typically results in positive probabilities for all possible couples. Still, a natural hypothesis is that mixed couples where the wife smokes should be much less frequent than vice-versa; i.e., the ratio of the two subpopulation sizes should be significantly lower than what would be implied by the sole difference in relative smoking prevalence.[5] To test this hypothesis, we compare the actual subpopulations ratio to what it should be under independence for each of the 9 US Geographical Census divisions; interestingly, we find that the hypothesis is satisfied in *each* of them, despite significantly different smoking prevalences across regions. This is consistent with our first prediction.

Similarly, our regression analysis shows that among smoking husbands, those who marry smoking wives exhibit (on average) 0.15 *more* years of completed education (or about a 1.2 percent higher annual earnings) than those with non-smoking wives, consistent with our second prediction. Conversely, our evidence reveals that among non-smoking wives, those with smoking husbands exhibit (on average) 0.11 to 0.13 *fewer* years of completed education than those with non-smoking husbands, supporting our third prediction. Consistent with our fourth

---

[4]Taking the model to the data in a structural way is far beyond the scope of this paper. Extending the Choo and Siow (2006) methodology to a multidimensional setting with discrete and continuous characteristics is still an open question, despite recent and promising advances (Chiappori, Salanié and Weiss, 2014; Dupuy and Galichon, 2014; Galichon and Salanié, 2010).

[5]In a pure Choo-Siow setting, this hypothesis would follow from a result by Graham (2011).

prediction, we also find positive assortative matching on education for each type of couple, and in particular for couples with identical smoking habits. In addition, at the top of the SES distribution, positive assortative matching on education by smoking status is stronger, supporting our fifth prediction. Finally, the well-known negative correlation between education and smoking is confirmed in our data, and we estimate that for men this correlation becomes less negative if one further controls for the wife's education, while this pattern does not appear for women; we argue that this fact is in line with our last prediction.

Perhaps the most informative prediction of our bidimensional matching model is the one stating that smoking husbands married to smokers have higher SES than those married to non-smokers. Such a prediction is typical of a matching logic, in which female smokers eventually benefit from being on the short side of the market. A potential concern here is that such empirical patterns (even when controlling for additional covariates), instead of being a consequence of matching mechanisms on the marriage market, may reflect the fact that smoking behavior is *endogenous* to marital status. However, we doubt this can explain our findings. Using information on engaged couples, newlyweds, and couples married for over 5 years, Sutton (1993) finds that similarities in smoking status were already present about the time of marriage. More recently, Banks, Kelly and Smith (2013), who use retrospective evidence from the Health and Retirement Study, show that most smoking behavior is initiated before marriage, so that smoking is *predetermined* with respect to marital status. In addition, we use a simple and popular estimator in program evaluation (see Wooldridge, 2002) obtained from an OLS regression of education on spousal smoking status controlling for the predicted probability that the spouse is a smoker (and its square). The key variable(s) used in predicting such a probability is (are) the age at which the spouse started smoking regularly (and its square). Reassuringly, the estimates adjusted by the spousal propensity to smoke give a similar picture to our main estimates.

Finally, we present two refutability tests in our Supplementary Material: one based on *randomly* sorting married men and women to couples, which shows that the predictions of our model are rejected; another based on testing a *single-index* model in education and smoking, which is rejected by the data.

The next Section contains the model. Section 3 describes the data. Section 4 tests the hypotheses based on the theoretical predictions of our model. Section 5 presents a brief

discussion, while Section 6 concludes.

2 The Model

We present a parsimonious preference model given the two types of characteristics at stake, defining the surplus structure and the corresponding stable matches under Transferable Utility (from now on TU).[6] After characterizing the equilibrium under a benchmark "symmetric" case (where male and female characteristics play the same role in the surplus function and are identically distributed), we analyze the general case deriving predictions on the form that the stable matching may take, which can be used in any discrete or continuous multidimensional framework, although a closed-form characterization of the stable matching cannot be obtained. Finally, we introduce the male prevalence condition into the model (for any SES level, there are more male smokers than female smokers among individuals above that SES level), and derive several predictions that will be the basis of the hypotheses in our empirical analysis.

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 possible. An agent is thus formally characterized by a pair (x, X) if female and (y, Y) if male, where

---

[6]The theoretical analysis of matching under TU is typical in marriage market analysis, whereas models analyzing dating typically consider NTU. 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.

x or $y \in [0, 1]$ is the agent's continuous index (i.e., SES), and $X, Y \in \{N, S\}$ defines the agent's discrete characteristic (i.e., smoking status). 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 habits X and Y.

2.1.2 Surplus

In any married couple, the sum of individual utilities is given by some function of the partner's characteristics; as it is customary, 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. We assume that the surplus Σ generated by a match between (x, X) and (y, Y) has the form:

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

The function S is strictly increasing, continuously differentiable and supermodular; moreover, $\lambda < 1$. This means 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. Hence, in our framework, the surplus of a mixed (smoker, non-smoker) couple is the same as that of a couple of smokers, but strictly less than of a non-smoking pair.

Restrictive as it may seem, we believe that the common λ assumption is a reasonable approximation. The multiplicative nature of the impact of λ reflects the fact that smoking, by reducing life expectancy, decreases the present discounted value of future welfare proportionally. We believe that this is more realistic than an additive parameter, as the latter would imply that those couples with lower socioeconomic resources (and surplus) would be disproportionately penalized by having (at least) a smoker in the couple. In addition, a large medical

and public health literature[7] shows that a smoking partner decreases life expectancy of a non-smoker but not of a smoker, which is compatible with our assumption.[8] Last but not least, while the single parameter λ assumption cannot be directly tested (since the marital surplus is not observed), one can evaluate its plausibility using auxiliary conditions of the Choo-Siow type, such as unobserved heterogeneity of type 1 extreme value distribution form. We perform this exercise and find empirical support for it (see Table S1 in the Supplementary Material).

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

---

[7]See for instance ASH (2011), CDC (2006), Glymour et al. (2008), Mannino et al. (1997), amongst others. Moreover, the attitude of smokers towards smokers is much more permissive than that of non-smokers (ASH, 2011; Lader, 2009; Pilkington et al., 2006), suggesting that beyond the impact on life expectancy, the psychological costs of a smoking partner are large for non-smokers but negligible for smokers.

[8]A natural extension would allow the surplus to be decreased by a smaller amount when both spouses smoke than when only one is a non-smoker; i.e., the surplus would be discounted by $\lambda < 1$ if both are smokers, and by $\lambda' < \lambda$ if only one is a smoker. We briefly discuss the implications in the concluding section (see footnote 15).

$(x, X), (y, Y)$ we have that

$$u_X(x) \geq 0, v_Y(y) \geq 0 \text{ and} \tag{1}$$

$$u_X(x) + v_Y(y) \geq \Sigma((x, X), (y, Y))$$

therefore

$$u_X(x) + v_Y(y) \geq S(x, y) \text{ if } X = Y = N \tag{2}$$

$$\geq \lambda S(x, y) \text{ otherwise}$$

where an equality obtains on the support of μ. The first constraints in (1) reflect the require-
ment 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 state 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.[9] 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:

$$\min_{u_N, u_S, v_N, v_S} \int_{[0,1] \times \{N,S\}} (\mathbf{1}[X = S] u_S(x) + \mathbf{1}[X = N] u_N(x)) \, dF(x, X)$$

$$+ \int_{[0,1] \times \{N,S\}} (\mathbf{1}[Y = S] v_S(y) + \mathbf{1}[Y = N] v_N(y)) \, dG(y, Y)$$

---

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

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.

Let us note that there may exist matches involving "mixed strategies", whereby an open set of agents are each matched to several agents with positive probabilities. In our setting, an individual may be indifferent between two types of mates. For instance, 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 $\lambda < 1$.

2.3   The symmetric case

As a benchmark, we consider 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. It is well known that a matching is stable if and only if the corresponding measure maximizes total surplus. In this context, 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 marry smokers. Assume it is; then an equal measure open set of non-smokers of the opposite gender also marry smokers. But since $\lambda < 1$, the total surplus is then less than in the completely assortative matching, a contradiction. ∎

While the symmetric case is obviously very specific, it constitutes an interesting benchmark. A first lesson that can be drawn from it is that bidimensional 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 (see Appendix).

The preceding conclusions, however, heavily rely on the very specific features of the symmetric framework. In particular, 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

We now consider the general case. 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. These probabilities are endogenous, and are determined by the equilibrium (stability) conditions.

2.4.1 Qualitative results

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. 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*

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

11

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

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

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.* ∎

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

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

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

Proof. *See 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.

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.

Note that, in particular, male smokers who match with female non-smokers must have a higher socioeconomic status than their wife; indeed, the only reason why a female non-smoker

would agree to marry a smoker (thus accepting a shrinkage of total surplus) is the shortage of male non-smokers. This particular pattern is *not* directly linked to the multiplicative nature of the impact of smoking; as soon as marrying a smoker involves a loss in surplus, we expect similar patterns. What is *specific* of the multiplicative setting, however, is that these phenomena do not take place at the top of the distribution –where the loss is larger.

This same logic also 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 and is 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 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;[10] ultimately, the matching patterns depend on the distributions F and G.

2.4.2 Surplus maximization

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

---

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

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.

- 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 $(\phi_N, \phi_S, \psi_N, \psi_S)$.

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:

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

$$+ \lambda \int_0^1 \left[(1 - p_S(t)) S(t, \phi_S(t)) + p_S(t) S(t, \psi_S(t)) \right] dF_S(t)$$

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. 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 (see Appendix).

2.5 The constrained general case

We now introduce the additional assumption 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:

15

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 non-smoker 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 non-smoker and smoker mate of (y', S). Then*

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

Proof. *See Appendix* ■

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, the smoker has a higher SES than the non-smoker.

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, 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. In addition, the intracouple allocation of the surplus can be recovered (see Appendix).

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

2.5.1 Theoretical Predictions

In summary, the theoretical predictions of the constrained general case are:

1. There are no "mixed" couples in which she smokes and he does not, whereas the opposite pattern –he smokes and she does not– happens with positive probability.

2. Smoking husbands married to smokers are of higher "quality" (i.e., higher SES) than those married to non-smokers: smoking "premium" for smoking wives.

3. Non-smoking wives married to smoking husbands have a lower SES than those married to non-smokers: smoking "penalty" for smoking husbands.

4. There is positive assortative matching on SES among couples with identical smoking habits.

5. Positive assortative matching on SES by smoking status is stronger at the top of the SES distribution: smoking husbands (wives) at the top of the SES distribution are more

likely to marry smoking wives (husbands) than those with lower SES. Similarly, non-smoking husbands (wives) at the top of the SES distribution are more likely to marry non-smoking wives (husbands) than those with lower SES.

6. When a female non-smoker marries both a smoker and a non-smoker with positive probability, the smoking husband has a higher SES than the non-smoking one. When a male smoker marries both a smoker and a non-smoker with positive probability, both women have the same SES.

3   Data Description

3.1   CPS March and TUS data

Our empirical application uses years of education as the measure of SES and smoking status for the binary attribute. We use data from the US Current Population Survey (CPS), specifically its annual March supplements and the Tobacco Use Supplements (TUS), for the years 1996 to 2003, 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; and February and June 2003.

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.[11]

The TUS-CPS is a National Cancer Institute sponsored survey of tobacco use and policy

---

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

information that has been administered as part of the CPS since 1992. 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 are 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 the 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, and 2003. 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), and November (2001 and 2003) 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 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. In the US the median age at first marriage was around 27 for men and 25 for women for the period 1996-2007. On the other hand, a lower bound of 22 and 24 years old also allows us to include college graduates after they have completed their schooling. The additional two years in the husbands' bounds are based on the standard median / mean age difference of two years between male and female spouse (Chiappori, Iyigun and Weiss, 2009). Note that the March CPS does not record the duration of marriage; in particular, the June Fertility Supplements that used to provide the age at (first) marriage, do not contain it any longer in the most recent years that our study is concerned about.

In addition to individual age, we use the state of residence, year of interview, sample household weight and education of the individual. From 1992, the CPS records education as degrees attained rather than years of schooling completed, we thus assign the number of years of schooling to the corresponding degrees. We will also consider information on self-reported health status, number of children under age six and male-head indicator. 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, we focus on the respondents' answers to whether and how often they smoke, whether they have smoked at least 100 cigarettes in their lifetime, and at which age they started smoking regularly. 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). Age at which they started smoking regularly, as a *predetermined* explanatory variable of current smoking status with respect to marital status, will prove useful in dealing with the potential *endogeneity* of smoking with respect to marital status. 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 et al., 2001).

3.2 Smoking Prevalence by Gender

An interesting aspect of the smoking attribute 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's Report, 2001), as well as by several studies in various fields (e.g., Gruber (2001) in economics and Öberg et al. (2011) in medicine). While a negative smoking gradient is observed by education (e.g., Gruber, 2001), the gender gap in smoking prevalence is maintained across all education levels. For instance, in 2007, the 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).

We further investigate this gender smoking asymmetry by computing the number of smokers by gender and educational attainment (years of education) in our data. Figure 1 shows that the number of male smokers is *always* higher than the number of female smokers, for each educational attainment (years of education). Hence, the **MP assumption** is satisfied in our data.

[**Figure 1 about here**]

3.3   Descriptive Statistics

Table 1 displays the summary statistics of the main variables of interest in our sample. Panel A reveals that women are slightly more educated and slightly less likely to be in excellent or very good health than their husbands, while their average number of children under six years old is about 0.8. Panel B reports the same statistics for never married individuals, showing again that men are healthier and less educated than women, and with a negligible number of young children. Finally, note the asymmetry in the smoking prevalence across genders: the smoking prevalence is 22% for husbands versus 17% for wives (29% and 26% for never-married men and women).

[**Table 1 about here**]

4   Empirical Analysis

4.1   From Theory to Data

It is well known by now that matching models are difficult to test from the sole observation of matching patterns. For instance, the Choo-Siow formulation, where the matching problem is analyzed as a series of discrete choice models, is exactly identified and therefore cannot be rejected by the data, even with its strong parametric assumption, i.e., that unobserved heterogeneity follows a type 1 extreme value distribution. Here, we suggest that an alternative approach may be feasible. In more complex settings, including multidimensional contexts, matching theory can generate strong *qualitative* predictions. In some cases –including the

21

present one– hypotheses based on these predictions can be tested in reduced form. Specifically, the theoretical analysis suggests that matching patterns should exhibit specific features due to the underlying competitive structure. These features could in principle be formally derived from an explicit, *specific* stochastic structure, which can be based on a Choo-Siow setting (following a line of research initiated by Graham, 2011), but *not* in general; an alternative justification would involve a "rank order" property à la Fox (2010), or possibly search models. In any case, we argue that if the MP assumption is satisfied, then we expect the following regularities to hold:

Hypothesis 1: Relative Prevalence of Mixed Couples. Mixed couples in which the wife smokes (denoted (0,1)) should be less frequent than those in which the husband smokes (denoted (1,0)). Actually, the MP assumption by itself implies an asymmetry between the proportions of mixed couples: if there are more male than female smokers, then one expects more (1,0) 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) for all SES levels; then if matching were random with respect to smoking habits, the ratio of (0,1) to (1,0) couples should be

$$r_{\text{observed}} < r_{\text{random}} = \frac{s_W\,(1 - s_M)}{s_M\,(1 - s_W)}$$

In our total 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 .72; we expect the observed ratio to be significantly smaller than this threshold. Moreover, the same exercise can be performed for each "marriage market" (or region).

Hypothesis 2: Smoking "Premium" for Smoking Wives. Among smoking husbands, those who marry smoking wives should have (on average) a *higher* SES than those who marry non-smokers.

Hypothesis 3: Smoking "Penalty" for Smoking Husbands. Among non-smoking wives, those who marry smoking husbands should have (on average) a *lower* SES than those who marry non-smokers.

Hypothesis 4: Assortativeness by SES among couples with identical smoking habits.

22

Among couples with identical smoking habits (i.e., both smokers or both non-smokers), matching should be assortative on SES.

Hypothesis 5: Stronger Assortativeness by Smoking Status at the Top of the SES Distribution. Smoking husbands (wives) at the top of the SES distribution are more likely to marry smoking wives (husbands) than those with lower SES. Similarly, non-smoking husbands (wives) at the top of the SES distribution are more likely to marry non-smoking wives (husbands) than those with lower SES.

Hypothesis 6: Conditional versus Unconditional Correlation between Smoking and SES. When two non-smoking women with the same 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 (non-smoking) 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 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 hypotheses.

4.2 Testing the Hypotheses

4.2.1 Relative Prevalence of Mixed Couples

Table 2 reports both the observed (Panel A) and random (Panel B) matching patterns by smoking status for husbands and wives. Panel A reveals that there is strong assortative

mating by smoking status: about 71% of couples have non-smoking spouses, and roughly 10% consists of smokers. This is in line with previous evidence on marital sorting by smoking status (e.g., Clark and Etilé, 2006). In addition, there are fewer mixed couples where the wife smokes than vice-versa, 6.7% versus 11.93%. These figures are very different from the percentages arising from random matching that are reported in Panel B, where sorting is weaker (the percentage of couples in the main diagonal is now less than 70%) and the two types of mixed couples are much more prevalent and similarly represented. Indeed, Panel C shows that the ratio of observed mixed couples where the wife smokes than vice-versa is 0.56 (s.e.=0.03), which is statistically significantly lower than the 0.72 (s.e.=0.02) implied by the sole difference in relative smoking prevalence, consistently with our hypothesis 1.[12] As a last test, in Panel D, we perform the same exercise at the US Geographical Census division level (which consists of 9 regions). The results are quite explicit: we find that the prediction is satisfied in *each* of these regions. Under the null of random matching, the probability of such an outcome would be extremely small, i.e., $2^{-9} \cong 2 \cdot 10^{-3}$.

[**Table 2 about here**]

4.2.2 Smoking Premium and Smoking Penalty

We investigate hypotheses 2 and 3 in Tables 3 and 4, respectively. They contain a series of regressions in which either the husband's or wife's education is the dependent variable while spouse's education and smoking status are the main explanatory variables, for the samples of smoking husbands or smoking wives in Table 3, and of non-smoking husbands or non-smoking wives in Table 4. Two specifications are presented: a standard one, with controls for own age, year and state fixed effects; and another one with additional controls, where we also include a spousal very healthy indicator, number of children under age six and a male-head indicator.

[**Table 3 about here**]

Starting from the left-block of regressions, columns (1) and (2) show that among smoking wives there is *no* statistically significant difference in the average years of completed education between those who marry smoking men and those who marry non-smoking ones, and that

---

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

the point estimates are virtually zero. The action is concentrated in columns (3) and (4): among smoking husbands those who marry smoking women have on average 0.15 *more* years of completed education (or about a 1.2 percent higher annual earnings[13]) than those with non-smoking ones. This supports our hypothesis 2.

Our 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, which is what we observe in the "placebo" columns (1) and (2). Smoking men have to compete for smoking partners with their own education. In other words, smoking women "marry up" with respect to her own education; 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 wife's smoker coefficient in columns (3) and (4) is opposite to the well-known negative gradient between own smoking and own education.

The evidence presented in Table 3 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 columns (3) and (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 columns (3) and (4) either.

We now turn to investigate hypothesis 3. Table 4 inquires about the average education among non-smoking wives and non-smoking husbands depending on the smoking status of their spouses. Consistent with hypothesis 3, columns (1) and (2) show that among non-smoking wives those with smoking husbands have on average 0.13 to 0.11 *fewer* years of completed education (or about a 0.9 percent lower annual earnings) 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

---

[13]Assuming that each extra year of education an individual gets is worth about an eight percent increment to their annual earnings.

it, in terms of lower socioeconomic standards. Finally, columns (3) and (4) show that among non-smoking husbands those with non-smoking wives have on average 0.24 to 0.21 *more* years of completed education than those with smoking wives. Thus, a non-smoking wife marries, on average, a better non-smoker spouse in terms of education.

[**Table 4 about here**]

4.2.3 Smoking Premium and Smoking Penalty controlling for the Propensity Score

Overall, the estimates reported in Tables 3 and 4 are consistent with our model, and support hypotheses 2 and 3. A potential concern is that smoking status captures some characteristics that we do *not* observe in our data (e.g., personality traits) that would explain the above empirical findings. Although smoking is correlated with self-control and impatience, recent empirical evidence shows that some non-cognitive skills and personality traits exhibit *negative* sorting in the marriage market, opposite to the strong positive sorting observed for smoking status. Dupuy and Galichon (2014) and Lundberg (2012) find that there is *negative* sorting in autonomy, that male extraversion is *negatively* correlated with female conscientiousness and risk attitudes, and male agreeableness is *negatively* correlated with female extraversion.[14]

Another concern is whether such empirical patterns (even when controlling for additional covariates) are the mere reflection of smoking behavior being *endogenous* to marital status, rather than a confirmation of the postulated marriage market mechanism. We explore such a possibility by using a simple and popular estimator in program evaluation (see Wooldridge, 2002) obtained from an OLS regression of education on spousal smoking status controlling for the predicted probability that the spouse is a smoker (the propensity score). The spousal propensity score is the probability that the spouse is a smoker given a set of observable characteristics. To make this approach as appealing as possible (at least one of) the predictors must be *predetermined* with respect to marital status, and the *age at which the spouse started smoking regularly* seems a natural candidate. Hence, we predict the probability that the spouse j is a smoker ($Smoking_j = 1$) given a set of spousal characteristics ($\mathbf{X}_j$) using a probability

---

[14]Lundberg (2012) also finds that extraversion and neuroticism only matter for women, while conscientiousness only for men, in terms of marriage probabilities.

model (Probit):

$$Smoking_j = \Phi(\mathbf{X}_j\beta) + \varepsilon_j \tag{9}$$

where Φ is the standard normal *cdf* and $\mathbf{X}_j$ contains the following spousal characteristics: age (in categories), years of education, year fixed effects, state fixed effects, and age at which the spouse started smoking regularly. We code this last variable as 0 for never smokers or never regular smokers. For this reason, and the fact that people who start regularly using tobacco when they are younger are more likely to have trouble quitting than people who start later in life (Surgeon General's Report, 2012), we expect a non-linear relationship between smoking status today and age at which the individual started smoking regularly, so that we also consider the *square* of this variable as an additional predictor.

Once we have estimated the propensity scores $\widehat{Smoking}_j$ for men and women (see Table A1), we then estimate the following OLS regressions

$$Education_i = \alpha + \delta Smoking_j + \gamma \widehat{Smoking}_j + u_i \tag{10}$$

for four groups of individuals, smoking wives, smoking husbands, non-smoking wives, and non-smoking husbands, but also

$$Education_i = \alpha + \delta Smoking_j + \gamma_1 \widehat{Smoking}_j + \gamma_2 \widehat{Smoking}_j^2 + u_i \tag{11}$$

to account for potential *nonlinearities* of the conditional expectation function of education in the propensity score. The estimated propensity score plays the role of a *control function.*

[Table 5 about here]

[Table 6 about here]

Reassuringly, the estimates displayed in Tables 5 and 6 give a similar picture to our main estimates in Tables 3 and 4, and are indeed consistent with existing research. Using information on engaged couples, newlyweds, and couples married for over 5 years, Sutton (1993) finds that similarities in smoking status were *already* present about the time of marriage. More recently, Banks, Kelly and Smith (2013), who use retrospective evidence from the Health and Retirement Study, show that most smoking behavior is initiated *before* marriage, so that

smoking is *not* endogenous to marital status. Controlling for own age in equations (10) and (11) is immaterial for our findings (results available upon request).

4.2.4 Assortativeness by SES among couples with identical smoking habits

In Table 7 we investigate the degree of assortativeness by education among different types of couples depending on spouses' smoking status. For each type of couple, we regress own education on spouse's education controlling for own age, year and state fixed effects. Our estimates reveal positive assortative matching by education for each type of couple. Although assortative mating by education has been extensively documented in the literature (Lam, 1988; Pencavel, 1998; Qian, 1998; Mare, 2008), here we show that it holds true within each spouses' smoking category. This is consistent with hypothesis 4.

[**Table 7 about here**]

4.2.5 Stronger Assortativeness by Smoking Status at the Top of the SES distribution

In Table 8 we investigate how the degree of assortativeness by smoking status varies across the SES distribution. We run regressions of a spousal smoking status indicator on individual smoking status, individual SES and the interaction of these two variables. We measure SES in two different ways: years of schooling and high-school degree or above (12 years of schooling or more). The results of these regressions suggest that, if anything, "mixed" couples (i.e., where one spouse is a smoker and the other is not) are less prevalent at high SES, supporting hypothesis 5.

[**Table 8 about here**]

4.2.6 Conditional versus Unconditional Correlations between Smoking and SES

Finally, in Table 9 we run regressions of education on smoking status for two samples, non-smoking women and smoking men, controlling or not for spousal SES. Our findings support hypothesis 6. Among men married to non-smoking women, smoking men tend to have (on average) a lower SES (1.3 fewer years of education) but this difference decreases (to 0.7 years less of education) when controlling for their wives' SES. Among women married to smoking

men, smoking women tend to have (on average) a lower SES (around 0.3 fewer years of education) *regardless* of their husbands' SES.

[**Table 9 about here**]

4.3   Refutability Tests

We implement two types of refutability tests (see Angrist and Krueger, 2001). We first provide placebo tests based on the idea that the predictions of our model should not be borne out in the data when married men and women are randomly assigned to couples. In our second approach, we test the proportionality constraints implied by an alternative single-index model.

4.3.1   Placebo Tests

We reshuffle our observed married individuals into new randomly married couples. The way we do it is in two stages. We first keep married men in our sample and generate a new id drawn from a uniform distribution for each of these men, and then rank them according to that pseudo-random id. We do the same for married women: we generate a new id drawn from a uniform distribution for each of these women, and then rank them according to that pseudo-random id. We then merge men and women according to that pseudo-random id. This constitutes our random matching sample. After that, we investigate whether the hypotheses based on the predictions of our model are observed in these randomly generated data. The answer is clearly negative:

- Table S2 in the Supplementary Material shows that with randomly generated couples, it is *not* true that among smoking husbands those who marry smoking wives have (on average) a higher SES.

- Table S3 in the Supplementary Material shows that with randomly generated couples, it is *not* true that among non-smoking wives, those who marry smoking husbands have (on average) a lower SES.

- Table S4 in the Supplementary Material shows that with randomly generated couples, it is *not* true that "mixed" couples are less prevalent at high levels of SES.

29

- Finally, Table S5 in the Supplementary Material reveals that with randomly generated couples, hypothesis 6 is *not* borne out in the data.

All in all, it seems that our model cannot explain the marriage patterns of randomly generated couples.

4.3.2 Alternative Single-Index Model

Our bidimensional model of smoking and SES rules out a single index representation: the trade-off between these two characteristics is perceived *differently* among potential spouses, which is precisely what index models forbid. Chiappori et al. (2012) have shown that index models are testable; they must satisfy a set of proportionality restrictions. In Table S6 in the Supplementary Material, we show that these predictions are rejected by the data when the two dimensions are smoking status and SES, providing additional support for the contention that, albeit very appealing, the single-index model representation is not always appropriate, and that, in particular, it does not apply to our case.

5   Discussion

Our model parsimoniously assumes that the presence of a smoking spouse decreases surplus by the same factor irrespective of the partner's smoking status. This assumption could be relaxed; one could assume, for instance, that the surplus discount is larger for mixed couples. The corresponding, more general model would still be solvable using the same technique. Most general conclusions would remain valid. In particular, there would be assortative matching at the top of the distribution; if male smokers outnumbered female ones, then mixed couples would consist of a male smoker and a non-smoking wife; and in the latter case the wife would "marry up". However, other predictions would be lost. For instance, the last statement of Proposition 5 would no longer hold; when a smoking man randomizes between two potential spouses, they will not have the same socioeconomic index (although, interestingly, they do in the data, see Table 6). In addition, the closed-form solution of the uniform-quadratic example provided in Supplementary Material would no longer be valid; its resolution, although still feasible, would become more complex (and tedious). For instance, it would require considering

30

more cases. All in all, the new, and less tractable, model, would lose in empirical content (although the predictions that would be discarded are supported by the data), with no significant gain in terms of its basic insights. We therefore believe that parsimony considerations suggest concentrating on the single λ case.

6   Conclusions

We develop a bidimensional frictionless matching model on the marriage market under transferable utility, 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. As long as traits are not perfectly correlated, this heterogeneity is key: it rules out a single-index representation. The trade-off between the two characteristics is perceived differently among potential spouses, which is precisely what index models forbid. That is, the corresponding one-dimensional matching techniques are not well-suited to characterize the stable matches in this setting, and this paper precisely characterizes a specific extension to allow for heterogeneous preferences.

The main message of our paper is twofold. First, specific multidimensional models of matching, although intrinsically more complex than one-dimensional ones, are by no means intractable. We actually describe a general strategy for tackling problems of this type, and in particular how the characterization of the equilibrium can be formulated as an optimal control problem; we also show how such a theoretical approach can generate strong predictions on matching patterns. While we concentrate on smoking in our empirical application, other aspects could readily be considered. Race or ethnicity are important examples (e.g., Chiappori, Oreffice and Quintana-Domeque, 2016), or characteristics such as age or health attributes.

Secondly, while one-dimensional models are difficult to test, at least in the absence of additional information on either match-specific surpluses or transfers between agents, in more complex settings, and particularly in multidimensional contexts, the strong *qualitative* predictions that matching theory generates can be taken to the data. Hypotheses based on these predictions can be tested in reduced form. In our case, the model predicts a series of specific patterns (from the relative scarcity of couples where the wife smokes and the husband does not, to the opposite association of a smoking spouse on one's education for wives and hus-

bands) that are difficult to justify otherwise, and can be directly tested in a simple and robust way. We do not view such a strategy as a substitute for more explicitly structural approaches, but we believe it provides an informative complement.

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

$$\int_{1-\varepsilon}^{1} dF_N(s) = \int_{1-\eta(\varepsilon)}^{1} dG_N(s). \tag{12}$$

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 (12) 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 (12), 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

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

while matching x and y' would generate a surplus

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

By definition of ε, $\Sigma_1 > \Sigma$, 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

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

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

$$\Sigma_1 - \Sigma = S\left(x, y'\right) - S\left(x', y'\right) - \lambda\left(S\left(x, y_S\right) - S\left(x', y_S\right)\right)$$
$$> S\left(x, y'\right) - S\left(x', y'\right) - \left(S\left(x, y_S\right) - S\left(x', y_S\right)\right) > 0$$

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

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

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

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

therefore

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

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*

$$u(x) = \max_s \left(S(x, s) - v_N(s) \right)$$
$$= \max_{s'} \left(\lambda S(x, s') - v_S(s') \right)$$

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:

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

Since $\partial S / \partial x$ is strictly increasing in y, $\lambda < 1$ 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

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

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

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

and $\Sigma_1 > \Sigma$ 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

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

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

$$\Sigma_1 = S(x, \bar{y}) + \lambda S(x', y)$$

Since

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

we have that $\Sigma_1 > \Sigma$; again, an open set of marriages satisfying this pattern would violate surplus maximization. The proof of the last statement is identical.

C   Proof of Proposition 5

The first statement is a direct consequence of Proposition 3. Similarly, the last two statements directly follow from Lemma 1 above. To establish the second, we need to show that any mixed marriage can only involve a non-smoking woman and a male smoker. Under the Male Prevalence assumption, this directly follows from the proof of Proposition 4.

D   Recovering individual utilities in the stable match

D.1   The symmetric case

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:

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

D.2 The general case

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.

D.3 The constrained general case

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,*

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

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).

Proof. *Assume Ms. x marries either Mr. y (a non-smoker) or Mr. y' The first inequality comes from the fact that, for any given SES, a non-smoker can only be either an equivalent or a better partner than a smoker. Next, if $p_N(x) > 0$, both a smoking and a non-smoking wives marry the same smoking husband with positive probability. Since the total surplus is the same in both cases, their utility must be the same. Lastly, first order conditions give that*

$$u'_N(x) = \frac{\partial}{\partial x} S(x, \phi_N(x)) > 0$$
$$u'_S(x) = \lambda \frac{\partial}{\partial x} S(x, \psi_S(x)) > 0$$

so that

$$(u_N(x) - u_S(x))' = \frac{\partial}{\partial x} S(x, \phi_N(x)) - \lambda \frac{\partial}{\partial x} S(x, \psi_S(x))$$

For x large enough, both $\phi_N(x)$ and $\psi_S(x)$ are close to 1, and that difference is positive. ■

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 the uniform-quadratic example we present in the Supplementary Material, 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.

References

[1] Angrist, J., Krueger, A. (1999) "Empirical Strategies in Labor Economics," in *Handbook of Labor Economics*, O. Ashenfelter and D. Card (Eds.), Elsevier, Volume 3A.

[2] ASH (2011) "Secondhand Smoke," Research Report.

[3] 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.

[4] 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.

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

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

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

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

[11] 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.

[12] 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.

[13] Chiappori, P.-A., Oreffice, S., Quintana-Domeque, C. (2016) "Black-White Marital Matching: Race, Anthropometrics, and Socioeconomics," *Journal of Demographic Economics*, forthcoming.

[14] Chiappori, P.-A., Salanié, B. (2015) "The Econometrics of Matching Models", *Journal of Economic Literature*, forthcoming.

[15] Chiappori, P.-A., Salanié, B., Weiss, Y. (2014) "Partner Choice and the Marital College Premium," mimeo, Columbia University.

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

[17] 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.

[18] Coles, M., Francesconi, M. (2013) "Equilibrium Search and the Impact of Equal Opportunities for Women," Economics Discussion Papers 742, University of Essex, Department of Economics.

[19] 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.

[20] Dupuy, A., Galichon, A. (2014) "Personality traits and the marriage market," *Journal of Political Economy*, 122(6):1271–1319.

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

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

[23] Galichon, A., Salanié, B. (2010) "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models", Columbia University, Discussion Paper: 0910-14.

[24] 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.

[25] Graham, B. (2011) "Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers," in J. Benhabib, M. Jackson and A. Bisin eds: *Handbook of Social Economics*, North-Holland, Vol. 1B, 965-1052.

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

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

[28] 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.

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

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

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

[32] Lader, D. (2009) "Smoking-related behaviour and attitudes, 2008/09." Opinions survey report 40.

[33] 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.

[34] Lundberg, S. (2012) "Personality and marital surplus," *IZA Journal of Labor Economics*, 1:3.

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

[36] 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.

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

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

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

[40] 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.

[41] 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.

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

[43] 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.

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

[45] 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.

[46] 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.

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

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

[49] Shimer, R., Smith, L. (2000) "Assortative matching and search," *Econometrica*, 68(2):343–369.

[50] 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.

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

[52] Surgeon General's Report (2012) "Preventing Tobacco Use Among Youth and Young Adults," US Department of Health and Human Services.

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

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

[55] 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.

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

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

[58] Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

Table 1: Summary Statistics

Means (Standard Deviations)

A. Married	Husbands	Wives
Age (years)	29.48	27.81
	(2.80)	(2.78)
Education (years)	13.65	13.79
	(2.37)	(2.30)
Smoke (= 1 if every day / some day smoker and 100 cigarettes in their lifetime)	0.22	0.17
	(0.42)	(0.38)
Very Healthy (= 1 if excellent or very good self-reported health)	0.83	0.80
	(0.38)	(0.40)
Number of children under age 6	0.83	
	(0.85)	
Number of couples	10,305	
B. Never Married	Men	Women
Age (years)	28.24	25.91
	(3.12)	(3.06)
Education (years)	13.77	13.99
	(2.28)	(2.22)
Smoke (= 1 if every day / some day smoker and 100 cigarettes in their lifetime)	0.29	0.26
	(0.45)	(0.44)
Very Healthy (= 1 if excellent or very good self-reported health)	0.79	0.77
	(0.41)	(0.42)
Number of children under age 6	0.04	0.15
	(0.24)	(0.44)
Number of individuals	8,990	9,361

Note. CPS 1996-2003, Men aged 24-34, Women aged 22-32. Sampling weights are used.

Table 2: Matching Patterns by Smoking Status

A. Observed	Non-Smoking Wife	Smoking Wife
Non-Smoking Husband	70.93% (7,390)	6.70% (683)
Smoking Husband	11.93% (1,211)	10.44% (1,021)

B. Random	Non-Smoking Wife	Smoking Wife
Non-Smoking Husband	64.32%	13.30%
Smoking Husband	18.53%	3.83%

C. Prevalence Ratios of Mixed Couples	Observed	Random
Ratios	$\frac{6.70}{11.93} = 0.56$ [0.03]	$\frac{13.30}{18.53} = 0.72$ [0.02]

D. Prevalence Ratios of Mixed Couples by Census Division	Observed	Random
Ratios in Census Division 1 (New England)	0.57 [0.12]	0.72 [0.08]
Ratios in Census Division 2 (Middle Atlantic)	0.63 [0.09]	0.73 [0.07]
Ratios in Census Division 3 (East North Central)	0.67 [0.09]	0.80 [0.06]
Ratios in Census Division 4 (West North Central)	0.62 [0.11]	0.75 [0.08]
Ratios in Census Division 5 (South Atlantic)	0.43 [0.07]	0.63 [0.05]
Ratios in Census Division 6 (East South Central)	0.58 [0.12]	0.75 [0.08]
Ratios in Census Division 7 (West South Central)	0.48 [0.08]	0.67 [0.06]
Ratios in Census Division 8 (Mountain)	0.55 [0.09]	0.70 [0.07]
Ratios in Census Division 9 (Pacific)	0.56 [0.10]	0.69 [0.08]

Note. Sampling weights are used. Weighted % and (non-weighted number of observations). "Delta method" standard errors in brackets.

Table 3:

Regressions of Education on Spousal Smoking Status for Smokers

Dependent variable:	Smoking Wives		Smoking Husbands	
	Wife's Education		**Husband's Education**	
	(1)	(2)	(3)	(4)
Spouse is a Smoker	−0.002	−0.006	0.139*	0.151*
	(0.092)	(0.092)	(0.080)	(0.080)
Spouse's Education	0.471***	0.461***	0.555***	0.542***
	(0.029)	(0.030)	(0.028)	(0.028)
Standard controls	Yes	Yes	Yes	Yes
Additional controls	No	Yes	No	Yes
N	1,704	1,704	2,232	2,232
R^2	0.28	0.29	0.37	0.37

Sample spans the columns "Smoking Wives" and "Smoking Husbands".

Note. Standard controls: own age, year and state fixed effects.
Additional controls: spousal very healthy indicator, number of children under age 6 and a male-head indicator.
Sampling weights are used.
Robust standard errors in parentheses.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table 4:

Regressions of Education on Spousal Smoking Status for Non-Smokers

Dependent variable:	*Sample*			
	Non-Smoking Wives		**Non-Smoking Husbands**	
	Wife's Education		**Husband's Education**	
	(1)	(2)	(3)	(4)
Spouse is a Smoker	−0.133**	−0.108*	−0.235***	−0.208***
	(0.063)	(0.063)	(0.080)	(0.080)
Spouse's Education	0.636***	0.608***	0.684***	0.669***
	(0.013)	(0.013)	(0.014)	(0.015)
Standard controls	Yes	Yes	Yes	Yes
Additional controls	No	Yes	No	Yes
N	8,601	8,601	8,073	8,073
R^2	0.48	0.49	0.46	0.47

Note. Standard controls: own age, year and state fixed effects.
Additional controls: spousal very healthy indicator, number of children
under age 6 and a male-head indicator.
Sampling weights are used.
Robust standard errors in parentheses.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table 5:
Regressions of Education on Spousal Smoking Status for Smokers
Controlling for Spousal Propensity to Smoke $\widehat{P}(Spouse's\ characteristics)$

	Sample			
	Smoking Wives		**Smoking Husbands**	
Dependent variable:	**Wife's Education**		**Husband's Education**	
	(1)	(2)	(3)	(4)
Spouse is a Smoker	−0.009	−0.004	0.250*	0.238*
	(0.151)	(0.165)	(0.130)	(0.138)
$\widehat{P}(Spouse's\ characteristics)$	−0.568**	2.93***	−0.743***	3.12***
	(0.245)	(0.816)	(0.207)	(0.704)
$\widehat{P}(Spouse's\ characteristics)^2$	—	−5.10***	—	−5.77***
		(0.999)		(0.895)
N	1,656	1,656	2,204	2,204
R^2	0.01	0.03	0.01	0.03

Note. $\widehat{P}(Spouse's\ characteristics)$ is the predicted probability of a spouse smoking status using a Probit.
Spouse's characteristics: age (in categories), years of education,
age when he/she started smoking, and its square (in columns (2) and (4)), 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:

Regressions of Education on Spousal Smoking Status for Non-Smokers

Controlling for Spousal Propensity to Smoke $\widehat{P}(Spouse's\ characteristics)$

	Sample			
	Non-Smoking Wives		**Non-Smoking Husbands**	
Dependent variable:	**Wife's Education**		**Husband's Education**	
	(1)	(2)	(3)	(4)
Spouse is a Smoker	−0.230**	−0.236**	−0.450***	−0.435***
	(0.101)	(0.096)	(0.112)	(0.111)
$\widehat{P}(Spouse's\ characteristics)$	−1.42***	2.57***	−1.02***	0.654
	(0.132)	(0.588)	(0.145)	(0.540)
$\widehat{P}(Spouse's\ characteristics)^2$	—	−5.87***	—	−2.63***
		(0.893)		(0.834)
N	8,529	8,529	8,038	8,038
R^2	0.04	0.04	0.02	0.02

Note. $\widehat{P}(Spouse's\ characteristics)$ is the predicted probability of a spouse smoking status using a Probit.

Spouse's characteristics: age (in categories), years of education,

age when he/she started smoking, and its square (in columns (2) and (4)), 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: Assortative Matching by Education

Regressions of Education on Spousal Education by Type of Couple

	Both Non-Smokers		Both Smokers	
	Wife's Education	Husband's Education	Wife's Education	Husband's Education
Spouse's Education	0.641***	0.695***	0.464***	0.440***
	(0.014)	(0.015)	(0.037)	(0.035)
N	7,390	7,390	1,021	1,021
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	0.604***	0.619***	0.482***	0.488***
	(0.034)	(0.040)	(0.047)	(0.044)
N	1,211	1,211	683	683
R^2	0.44	0.44	0.33	0.31

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 8: Assortativeness in Smoking Status by SES

Regressions of Spousal Smoking Status

	Wife is a Smoker		Husband is a Smoker	
	(1)	(2)	(3)	(4)
Education	−0.014***	—	−0.022***	—
	(0.001)		(0.002)	
Smoker	0.239***	0.311***	0.389***	0.374***
	(0.074)	(0.032)	(0.103)	(0.045)
Education × Smoker	0.009	—	0.004	—
	(0.006)		(0.008)	
High School (or more)	—	−0.034**	—	−0.105***
		(0.016)		(0.020)
High School (or more) × Smoker	—	0.073**	—	0.092*
		(0.035)		(0.047)
N	10,305	10,305	10,305	10,305
R^2	0.20	0.19	0.20	0.19

Note. All regressions include spouse's 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 9: Conditional and Unconditional Correlations

Regressions of Education on Smoking Status

Dependent variable:	Sample			
	Non-Smoking Women		**Smoking Men**	
	Husband's Education		**Wife's Education**	
	Unconditional	Conditional	Unconditional	Conditional
Smoker	−1.34***	−0.657***	−0.339***	−0.314***
	(0.072)	(0.055)	(0.087)	(0.073)
Spouse's Education	—	0.683***	—	0.559***
		(0.008)		(0.018)
Adjusted Wald Test				
Test of Equality (p-value)	0.0000		0.6291	
N	8,601	8,601	2,232	2,232
Adjusted R^2	0.08	0.48	0.08	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 A1: Propensity to Smoke

Probit Models of Smoking Status on Explanatory Variables

	Wives		Husbands	
	(1)	(2)	(3)	(4)
Age individual started smoking regularly	0.145***	0.272***	0.157***	0.290***
	(0.003)	(0.010)	(0.003)	(0.013)
Age individual started smoking regularly^2	—	−0.006***	—	−0.007***
		(0.0004)		(0.0006)
Education	−0.124***	−0.106***	−0.141***	−0.131***
	(0.010)	(0.010)	(0.010)	(0.011)
Age fixed effects	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES
State fixed effects	YES	YES	YES	YES
N	10,835	10,835	10,790	10,790
Pseudo-R^2	0.50	0.52	0.54	0.55

Note. Sampling weights are used. Robust standard errors in parentheses.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Figure 1: Number of Smokers by Educational Level and Gender. Married.