Human Capital and Economic Opportunity:
A Global Working Group

Working Paper Series

Working Paper No. 2011-017

Matching with a Handicap: The Case of Smoking in the Marriage Market

Pierre-André Chiappori
Sonia Oreffice
Climent Quintana-Domeque

October, 2011

Human Capital and Economic Opportunity Working Group
Economic Research Center
University of Chicago
1126 E. 59th Street
Chicago IL 60637
humcap@uchicago.edu

Matching with a Handicap:
The Case of Smoking in the Marriage Market*

Pierre-André Chiappori — Columbia University

Sonia Oreffice — Universitat d'Alacant and IZA

Climent Quintana-Domeque — Universitat d'Alacant and IZA

First draft: April 15, 2011.

Abstract

We develop a matching model on the marriage market, where individuals have preferences over the smoking status of potential mates, and over their socioeconomic quality. Spousal smoking is *bad* for non-smokers, but it is *neutral* for smokers, while individuals always prefer high socioeconomic quality. Furthermore, there is a gender difference in smoking prevalence, there being more smoking men than smoking women for all education levels, so that smoking women and non-smoking men are in *short* supply. The model generates clear cut conditions regarding matching patterns. Using CPS data and its Tobacco Use Supplements for the years 1996 to 2007, and proxing socioeconomic status by educational attainment, we find that these conditions are satisfied. There are fewer "mixed" couples where the wife smokes than vice-versa, and matching is assortative on education within smoking types of couples. Among non-smoking wives those with smoking husbands have on average 0.14 fewer years of completed education than those with non-smoking husbands. Finally, and somewhat counterintuitively, we find that, as theory predicts, among smoking husbands those who marry smoking wives have on average 0.16 more years of completed education than those with non-smoking wives.

Keywords: Smoking, education, matching, marriage market.

JEL Codes: D1, J1.

*Paper presented at Columbia University and at the ICREA-MOVE conference on family economics (Barcelona, 2011). We thank Bernard Salanié and Marco Francesconi for helpful comments and suggestions. The usual disclaimers apply. Oreffice and Quintana-Domeque acknowledge financial support from the Spanish Ministry of Science and Innovation (ECO 2008-05721/ECON).

1 Introduction

Cigarette smoking remains the Nation's leading cause of premature, preventable death; during 2000–2004, approximately 443,000 premature deaths in the United States each year were attributed to cigarette smoking (CDC, 2008). Smoking causes deaths from heart disease, stroke, lung and other types of cancer, and chronic lung diseases. In addition, exposure to secondhand smoke is considered an important cause of premature death and disease for individuals who do not smoke themselves, while this secondary effect does not seem to exist for smokers (CDC, 2006; Glymour et al., 2008; Mannino et al., 1997). From a more subjective perspective, a large body of survey evidence around the world shows that the attitudes toward smoking behavior are different between smokers and non-smokers: smokers are less likely than non-smokers to be bothered by secondhand smoke exposure.[1] In a nutshell, there are significant differences between smokers and non-smokers in the health effects of and the attitudes toward being close to a smoker: non-smokers mind being married to a smoker, whereas smokers do not.

In this paper, we investigate the impact of this asymmetry on the *formation* of couples on the marriage market. The issue at stake can be intuitively described as follows. Consider a matching game in which individuals differ both in their 'quality', as summarized by some socioeconomic index (reflecting differences in education, income, social prestige and others), and in their smoking habit. Assume furthermore that the matching tends to be *assortative* in the spouse's quality index (technically, the match of two given spouses generates a surplus that is supermodular in their indices). This tendency is however constrained by the impact of smoking habits. When looking for a mate, non-smoking individuals perceive the smoking habit of a potential partner as a *negative* trait; everything equal, we therefore expect that smokers would be less successful in their quest for a spouse, at least among non-smoking potential

[1] For example, in Ireland, there is evidence based on the Dublin Healthy Cities project (50% vs. 92%). In Australia, data from the 2004 Victorian Population Survey show that smokers are significantly more likely than non-smokers to be 'not at all' concerned about exposure to passive smoke (46% compared with 12%, respectively). See also Pilkington et al. (2006).

partners. However, this potential handicap disappears if their spouse is also a smoker.

In this context, the matching game generates a very simple equilibrium in a particular case - namely if the prevalence of smoking habits is *identical across genders*. Then the stable match is fully symmetric and perfectly assortative along the two dimensions (quality and smoking habits) of heterogeneity; in practice, two marriage markets coexist, and smoking (resp. non-smoking) men marry smoking (resp. non-smoking) women assortatively along the quality dimension. Interestingly, in such a frictionless and perfectly symmetric world, even a small difference in a non smoker's perceived well-being between marrying a smoker or a non-smoker would result in perfectly segregated marriage markets.

Regarding smoking habits, however, symmetry is counterfactual. A striking feature of smoking behavior is precisely the prevailing differences between genders. Male smokers largely outnumber female smokers in the United States, a discrepancy that has remained stable over the last decades. This gender asymmetry has been emphasized by the Surgeon General (e.g., Surgeon General Report, 2001), as well as by several studies in various fields, e.g., Gruber (2001) in economics and Oberg et al. (2010) in medicine. In 2007, in the United States, 26.5% of white men 18–24 years of age and 21.6% of white women 18–24 years of age were current cigarette smokers (NCHS, 2010); the prevalence of smokers among white men 25–34 years of age was 29.0% while it was 21.4% among white women of the same age (NCHS, 2010). The symmetric solution is therefore not attainable, leading to a more complex and more interesting problem, in which some men (the non-smokers) and some women (the smokers) are on the short side of the market.

From a theoretical perspective, our analysis has two specific features. First, it is 'truly' multidimensional, in the sense that agents are characterized by two attributes (quality index and smoking habits) that cannot be summarized into a single indicator; in this respect, it differs from matching models where marital sorting happens along only one characteristic such as education or income (Becker, 1990; Lam, 1988), or where preferences are homogeneous over individual attributes (Chiappori, Oreffice and Quintana-Domeque, 2010, or the search model with two attributes by Coles and Francesconi, 2007). Moreover, one of our characteristics (the quality index) is continuous while the other is discrete; for that reason, the approach developed by Galichon and Salanié (2010) does not apply in our context. Second, and perhaps more interestingly, our framework does not exhibit the kind of single-crossing property that is typically assumed in matching models. Indeed, while the surplus function that determines the outcome of any possible match is assumed supermodular *in qualities*, the interaction with smoking habits seriously complexifies the problem. In particular, one can readily check that our setting does *not*, in general, satisfy the 'twisted buyer' condition that generalizes supermodularity in a multidimensional setting (see Chiappori, McCann and Nesheim, 2010, for a precise statement). As a consequence, the stable match needs not to be *pure*; that is, an open set of agents may at equilibrium be indifferent between several possible matches and randomize between them.[2]

Specifically, we study a model of bidimensional, frictionless matching under transferable utility that captures the main aspects just evoked. In our setting, individual quality indices are normalized to be uniformly distributed over $[0,1]$. A match between a woman with index x and a man with index y generates a surplus $f(x,y)$ which is strictly increasing and supermodular; for practical computations, we follow Chiappori, Iyigun and Weiss (2009) and adopt the quadratic form $f(x,y) = (x+y)^2/2$. If both spouses are non-smokers, they enjoy the full surplus. If, however, one (at least) is a smoker, then the total surplus shrinks by a factor $\lambda < 1$, representing the reduction in life expectancy associated to smoking in marriage (and possibly the non-smokers' distaste for living with a smoker). Although our approach focuses on two specific traits - smoking and 'quality', the framework we develop is suited to investigate matching problems in more general settings, characterized by the presence of several individual attributes, one of which is not necessarily perceived as good or bad by all

[2] Formally, the support of the measure that characterizes the stable match may not be born by the graph of a one-to-one mapping; see Chiappori, McCann and Nesheim (2010) for a detailed discussion of the pureness criterion.

individuals on the marriage market. Moreover, it would be useful in applications where there is a gender asymmetry in the prevalence of a certain attribute in the population. Finally, it is important to note that our approach is purely static. In our context, smoking behavior is viewed as a given characteristic of individuals, and we do *not* model the decision to start or quit smoking. Understanding the dynamics of smoking behavior, especially in its relation to the marriage market, is an important and interesting problem that is left for future work.

Results We fully characterize the stable matching when λ is 'not too small', in a sense we precisely define, and show that it is unique. The corresponding allocation has interesting features. First, it is not pure; an open set of non-smoking women marry either a smoking or a non-smoking partner with positive probabilities. Second, a non-smoking woman with a 'high' index will marry a high index, non-smoking man with probability one; similarly, higher quality, smoking women are deterministically matched with high quality, smoking men, and matching within each category is assortative on quality index. Third, non-smoking women located at the bottom of the quality distribution are randomly matched with either smoking or non-smoking spouses, whereas smoking women in similar situations are all matched with smoking men; again, matching is assortative on quality indices. In particular, among *non-smoking* women, only those with a lower quality index marry a smoking husband with positive probability (and then his index is low), whereas the pattern is inverted for men - only 'lower quality', *smoking* men may marry a (low quality) non-smoker. Finally, there is no 'mixed' marriage between a smoking wife and a non-smoking husband.

In practice, observed matching patterns are largely stochastic. Still, our theoretical analysis suggests that these stochastic patterns should exhibit the corresponding following features. First, mixed couples in which the wife smokes while the husband does not should be less frequent than those in which he smokes and she does not; second, among couples with identical smoking habits matching should be assortative on the quality index; third, non-smoking wives married to a smoking husband should have a lower index than those married to a non-smoking husband; the same should hold for smoking husband married to a non-smoking wife. Fourth, while there is a well-known negative correlation between smoking and education, especially for male, the correlation should be less negative for men married to non-smoking women, when controlling for wives' education.

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

[3] We consider couples whose husbands are between 24 and 34 years old, and wives between 22 and 32 years old.

[4] Section 5 provides a comparison to alternative data sets.

becomes even less negative (and significantly so), in line with the fourth and last prediction of the model.

Related literature Assortative marriage for smoking habits has been previously and extensively documented in the medical and biological literatures (e.g., Sutton, 1980; Venters et al., 1984; Sutton, 1993), although there has been very little economic focus. In the UK, Clark and Etilé (2006) use the British Household Panel Survey from 1991–1999 to document positive sorting by smoking status. Recently, Maralani (2009) shows the existence of assortative mating by smoking across *old* cohorts of Americans born between the 1920s and the 1950s, using data from the Health and Retirement Study. However, there is no matching model investigating the formation of couples by smoking status, no focus on recent years, and even less so any consideration regarding the heterogeneity of preferences between smokers and non-smokers or the gender gap in smoking prevalence.

A key role is also played by educational attainment, an important component of the 'quality' relevant in marital sorting, which is also closely related to cigarette use. This *cigarette connection* is acknowledged by economists at least since the seminal work by Farrell and Fuchs (1982), who document a negative smoking gradient by socioeconomic status (see also Gruber, 2001). Recently, De Walque (2010), using retrospective smoking histories constructed from the smoking supplements of the National Health Interview Surveys conducted between 1978 and 2000, shows that at least among women, college education has a negative effect on smoking prevalence, and that more educated individuals responded faster to the diffusion of information on the dangers of smoking after 1950. Note, however, that the gender gap in smoking prevalence is maintained across all education levels. In 2007, the (age-adjusted) prevalence of smokers by educational level among white men and women 25–years of age and over were as follows: 30.8% vs. 23.9% for those with less than high-school; 29.9% vs. 25.2% for those with high-school; 21.8% vs. 19.6% for those with some college; and 10.5% vs. 8.2% for those with college or above (NCHS, 2010).

The paper is organized as follows. Section 2 presents the model. Section 3 relaxes some assumptions of the model and discusses its empirical implementation. Section 4 describes the data used in the empirical analysis. Section 5 provides the empirical results and sensitivity analysis. Section 6 concludes.

2 The basic framework

2.1 Populations and surplus

We consider two populations (men and women) of equal size, normalized to one. Agents differ in two respects. First, they are characterized by their socioeconomic status; we may think of income, education, prestige, or any combination of those. For simplicity, this status is represented by a continuous index that is uniformly distributed over the interval $[0, 1]$, although more general settings could be considered and will be briefly discussed in the last subsection. Second, some agents smoke while others do not; an agent is thus formally characterized by a pair (x, A) if female and (y, B) if male, where $x \in [0, 1]$ ($y \in [0, 1]$) is the agent's index, and $A, B \in \{N, S\}$ defines the agent's smoking status (non-smoker or smoker). For the time being, we assume that smoking is independent of socioeconomic status. Let s_M and s_W denote the proportion of smokers in the male and female populations, respectively; a crucial feature is that $s_M > s_W$, i.e., the proportion of smokers is larger among men.

We consider a frictionless matching model with transferable utility à la Becker-Shapley-Shubik, in line with recent contributions on similar topics (e.g., Chiappori and Oreffice 2008; Chiappori, Iyigun and Weiss 2009; Galichon and Salanié, 2010). In any married couple, the sum of individual utilities is given by some function of the partner's characteristics; as it is customary in this literature, we define the surplus generated by marriage as the difference between this function and the sum of utility levels the spouses would reach as singles. In our framework, the surplus depends on both the socioeconomic status and the smoking habits of each partner. Following the previous discussion, we assume that a person who smokes does not mind a partner with the same habit; only a non-smoking individual will perceive smoking as a negative attribute of the potential partner, although there are direct health effects of smoking in marriage. Thus, the surplus function Σ has the form:

$$\begin{aligned}\Sigma &= f(x,y) \text{ if both spouses do not smoke}\\ \Sigma &= \lambda f(x,y) \text{ otherwise}\end{aligned}$$

where x and y denote the wife's and husband's respective socioeconomic indices, and where the function f is strictly increasing and supermodular, and satisfies $f(0,0) = 0$. Here, $\lambda < 1$ represents the decrease in surplus generated by the presence of (at least) a smoker in the couple; note that the surplus of a mixed (smoker-non smoker) couple is the same as that of a couple of smokers, but strictly less than for a non-smoking pair.[5] For practical purposes, we will focus on the quadratic case $f(x,y) = (x+y)^2/2$.

2.2 Stable matching

A matching is defined as a measure $d\mu$ on the set $([0,1] \times \{N,S\})^2$ and four functions $u_N(x), u_S(x), v_N(y), v_S(y)$; intuitively, $d\mu[(x,A),(y,B)]$ denotes the probability that agent (x,A) is married with agent (y,B), and $u_A(x)$ (resp. $v_B(y)$) is the utility of a woman (man) with index x (resp. y) and smoking habit A (resp. B). The only constraint on $d\mu$ is that its marginal should equal the initial distributions of individuals (for instance, the marginal on the set $[0,1] \times \{S\}$ of smoking women is uniform with total mass s_W, etc.).

A matching is *stable* if no matched agent would be better off unmatched, and if no two

[5] λ reflects both the distaste d for spousal smoking by a non-smoker individual (indirect effect of smoking on the non-smoker) and the direct health effect of smoking in marriage. To be more specific, we can write $\lambda(d; S_1, S_2)$, where $d = 1$ if one of the spouses is a smoker, and the other is not; $S_1 = 1$ if the wife is a smoker, $S_2 = 1$ if the husband is a smoker. We assume that there are no gender asymmetries in these effects, so $\lambda(1;0,1) = \lambda(1;1,0)$ and that the reduction in surplus due to the distaste effect plus a direct health effect (for one of the spouses) is equal to the reduction in surplus due to two direct health effects when both are smokers, so $\lambda(1;0,1) = \lambda(1;1,0) = \lambda(0;1,1)$.

individuals would prefer being matched together to their current situation. This property can be summarized by the following set of inequalities: for any $(x, A), (y, B)$ we have that

$$\begin{aligned} u_A(x) + v_B(y) &\geq f(x, y) \text{ if } A = B = N \\ &\geq \lambda f(x, y) \text{ otherwise} \end{aligned}$$

where an equality obtains on the support of $d\mu$.

Existence of a stable match stems from general results. Stability, in this transferable utility context, is equivalent to the maximization of aggregate surplus over all possible assignments; therefore the problem boils down to the existence of a solution to a simple maximization problem, for which one can readily check that the standard conditions are satisfied.

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

To prove pureness (or assortativeness), the standard approach, in the one-dimensional case, relies on supermodularity. In a differentiable setting, supermodularity requires that the partial of the surplus function vis a vis one spouse's attribute be strictly increasing (therefore injective) in the other person's attribute. In a multidimensional setting, the natural generalization of supermodularity is the 'twisted buyer' condition[6], which is sufficient to prove pureness of the stable match. The 'twisted buyer' condition states that there exists

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

a set X_L of measure zero such that for each distinct pair (y_1, y_2), any critical points of the function $x \longrightarrow \Sigma(x, y_1) - \Sigma(x, y_2)$ lie in X_L. In our specific context, this would require that for almost all x_0, the partials of the surplus with respect to the index x, computed at two points (x_0, y_1) and (x_0, y_2), cannot be equal unless $y_1 = y_2$. One can easily check that this property does not hold for the model just described. If a woman with index x_0 is a non-smoker, the partial of the surplus with respect to x is $(x_0 + y_1)$ if she marries a non-smoker with index y_1, and $\lambda (x_0 + y_2)$ if she is mated with a smoker with index y_2. For any $y_2 \in \left[\frac{(1-\lambda)x_0}{\lambda}, 1\right]$, if $y_1 = \lambda y_2 - (1-\lambda) x_0$, then the couples (x_0, y_1) and (x_0, y_2) violate the 'twisted buyer' condition, and this construction can be made for an open set of values x_0 - namely $x_0 \in \left[0, \frac{\lambda}{1-\lambda}\right]$. It follows that the stable matching may not be pure in our setting; indeed, we will show below that it is not.

2.3 Characterization

Given the assumptions made, all couples marry, and the resulting stable matching must maximize the total surplus generated over the populations. Regarding smoking habits, four categories of couples (at most) may appear: two non-smokers, two smokers, and a non-smoking wife (husband) matched with a smoking husband (wife). Moreover, within these categories, supermodularity implies that matching will be assortative; i.e., men with a higher socioeconomic index will marry with wives with a higher socioeconomic index.

Our main result is the following:

Proposition 1 *Assume that* $\lambda \geq \frac{s_M}{s_M+1}$. *There exists four values* X, Y, Y' *and* p, *all between* 0 *and* 1, *such that the unique stable match has the following features:*

- *All agents marry*
- *For all* $x \geq X$, *a non-smoking woman with index* x *is matched with probability* 1 *to a*

non-smoking husband with index

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

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

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

In particular, smoking men and non-smoking women marry 'down', whereas non-smoking men and smoking women marry 'up'.

- *For $x < X$, a non-smoking woman with index x is matched:*
  - *with probability p, to a smoking husband with index*

    $$y' = \frac{(1 - s_W)\, p + s_W}{s_M} x < Y'$$

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

    $$y = \frac{(1 - s_W)(1 - p)}{1 - s_M} x < Y$$

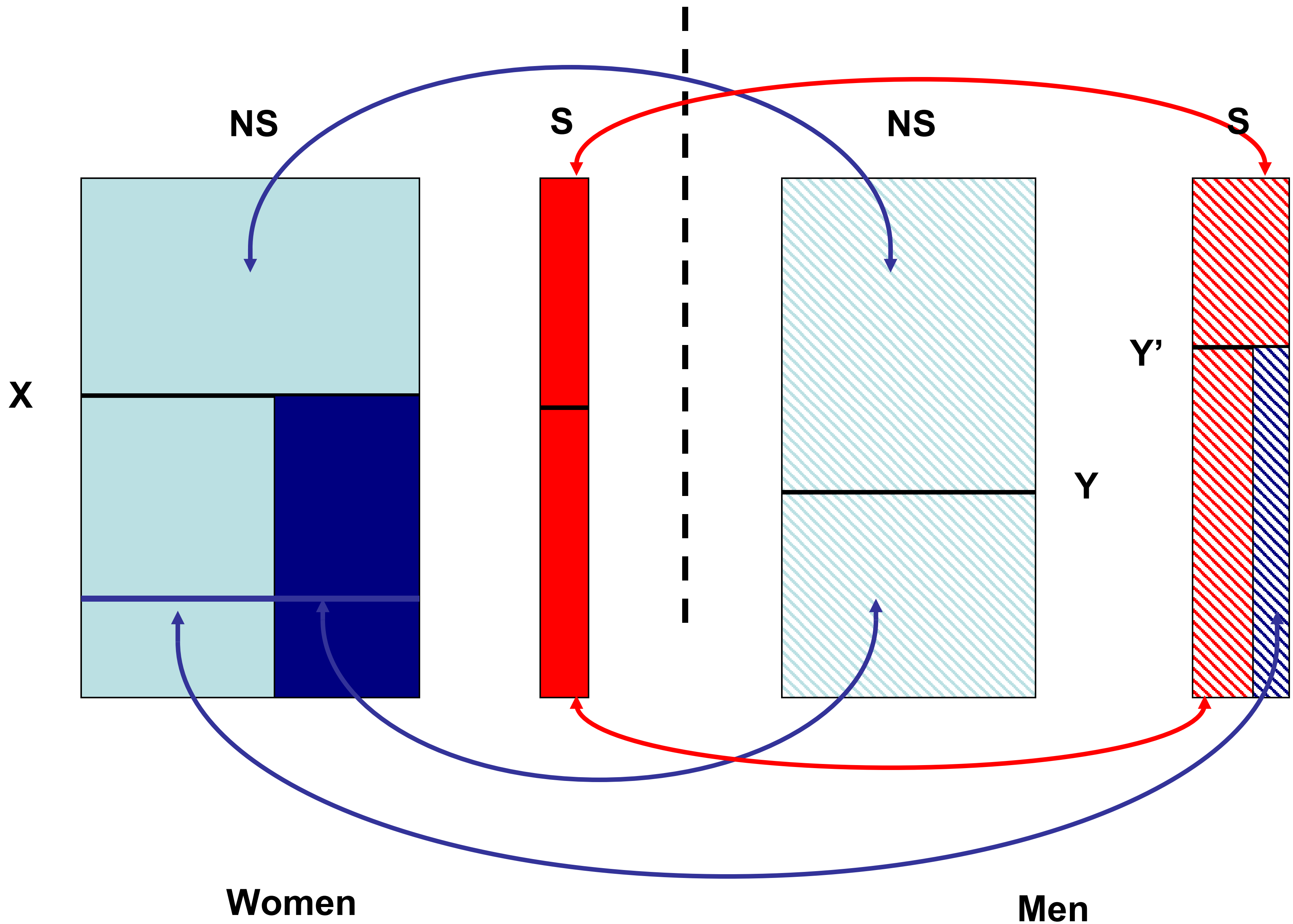
    Moreover, conditional on the index x of the wife, smoking husbands have a higher index than non-smoking ones - i.e., $y' > y$.

- *For $x < X$, a smoking woman with index x is matched with probability 1 to a smoking husband with index:*

  $$y' = \frac{(1 - s_W)\, p + s_W}{s_M} x < Y'$$

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

Proof. *See Appendix* ■


NS
S
NS
S
X
Y'
Y
Women
Men

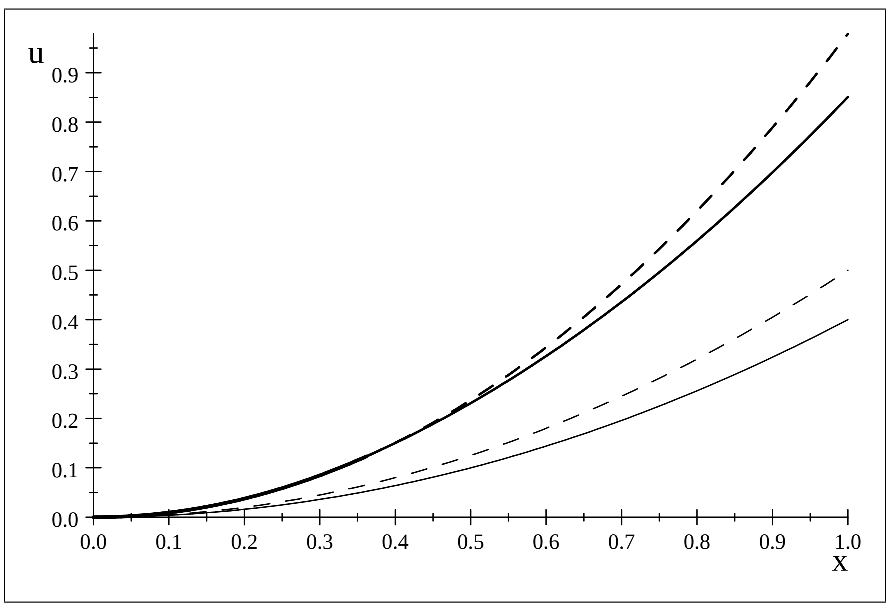
The stable matching is summarized in Figure 1. It can be interpreted as follows. First, high index non-smoking women tend to marry high index non-smoking men, and high index smoking women tend to marry high index smoking men. Such a matching is stable because, for a given index, a non-smoking person views a smoking potential partner as an inferior substitute for a non-smoking one, whereas a smoking person would view them as equivalent. Among these couples, assortative matching requires that, for any couple (x, y), the number of women with an index above x be equal to the number of men with an index larger than y. Since non-smoking women outnumber non-smoking men, non-smoking men and smoking women marry 'up', whereas conversely smoking men and non-smoking women marry 'down'. Below the threshold X, however, the stable match involves randomization: non-smoking women may be married with either a smoker or a non-smoker, while smoking women only marry smokers. Note that non smoking women who randomize at equilibrium must be indifferent between the two potential spouses. Since they dislike smoking habits, this handicap must be compensated by a higher index: of the two potential husbands, the smoker is therefore of higher quality.

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

- Non-smoking husbands always marry a non-smoking wife with probability 1
- Smoking husbands with a higher index ($y' \geq Y'$) marry a high index, smoking wife with probability 1
- Smoking husbands with a lower index ($y' < Y'$) marry either a smoking or a non-smoking wife with positive probability. Here, however, the two potential wives have the same quality index, since a smoking husband is neutral vis a vis his spouse's smoking status.

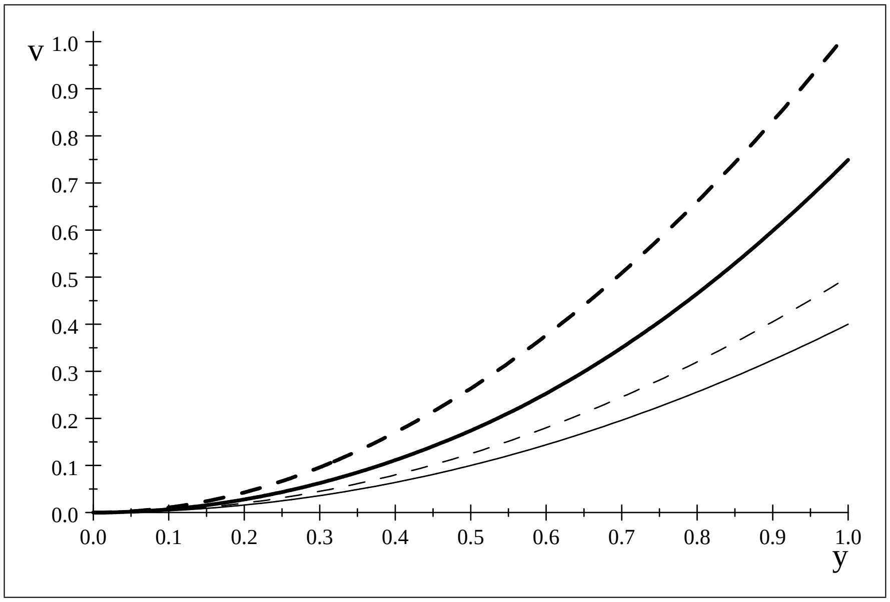
2.4 Utilities and comparative statics

A by-product of the result is a derivation of the intrahousehold allocation of resources implied by equilibrium conditions. In our case, this allocation is exactly pinned down.[7] A precise characterization is given in the Appendix. Figures 2a and 2b represent these utilities as a function of the index for females and males respectively. Both are increasing with the index, and the utility of a married person is always larger than if single. Moreover, the utility of a male smoker is always smaller than for a non smoker. For wives, however, utilities of smokers and non smokers are identical below the threshold X. Non-smoking women with an index smaller than X marry smoking husbands with positive probability; since, from a smoking husband's perspective, a smoking wife with identical index is a perfect substitute, smoking and non-smoking women must have the same utility.



Female utility (dashed = non smokers, thin = singles); $s_M = .25, s_W = .2, \lambda = .8$

[7] See Browning, Chiappori and Weiss (2010, ch. 8) for a detailed presentation of the technique used below.



Male utility (dashed = non smokers, thin = singles); $s_M = .25, s_W = .2, \lambda = .8$

One can also study the main comparative statics properties of the model. They are as follows:

- A larger λ, by reducing the welfare cost of smoking, increases the threshold X, and benefits male and female smokers, but also female non-smokers; however, it hurts male non-smokers by reducing their comparative advantage.
- Increasing the proportion of male smokers s_M also increases the threshold. Its impact on individual welfare is more complex. Among low index individuals, it hurts female and male non-smokers, but benefits male smokers. Regarding high index individuals, a higher s_M favors female smokers and male non-smokers, to the detriment of male smokers and female non-smokers.
- Finally, an increase in s_W has the opposite impact on high-index individuals, but no effect on low-index people.

3 Extensions

3.1 Relaxing some assumptions

The previous model relies on several simplifying assumptions; we briefly discuss the robustness of the main findings when these assumptions are relaxed. Relaxing the assumption of uniform distributions or the functional form of the surplus does not change the qualitative properties of the stable matching; the thresholds X, Y and Y' have to be redefined accordingly, and the probability p of lower index, non-smoking female marrying a smoker, will typically be index-dependent. A second feature is that the education distribution is identical for men and women. If this property does not hold, then it needs not be the case that, among high index individuals, non-smoking men marry 'up'. Indeed, while the basic assortative matching property - for any (x, y) couple, the number of women with an index above x equals to the number of men with an index larger than y - still holds, it no longer implies that $x > y$. However, the other predictions of Proposition 1 remain valid. A third, simplifying assumption used in the previous model is that smoking is independent of socioeconomic status; we discuss its relaxation below.

3.2 Practical implementation

In practice, the frictionless process described in the model is never observed. Marriage markets are not frictionless; moreover, actual matching involves multidimensional characteristics, some of which may actually be unobserved by the econometrician (a direction followed for instance by Chiappori, Oreffice and Quintana-Domeque, 2010, and Galichon and Salanié, 2010), and may furthermore be affected by random shocks à la Shimer and Smith (2000). For all these reasons, observed matching patterns are largely stochastic. Still, the previous analysis suggests that these stochastic patterns should exhibit specific features due to the underlying, competitive structure.[8] Specifically, we expect the following regularities to hold:

1. Mixed couples in which the wife smokes while the husband does not (denoted S-N) should be less frequent than those in which he smokes and she does not (denoted N-S); more precisely, the ratio of S-N to N-S couples should be smaller than implied by the sole difference in relative smoking prevalence, i.e., than the ratio

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

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

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

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

4. In the simplified model, when two non-smoking women with the same (low) index marry respectively a smoker and a non-smoker, the non-smoker should on average be of lower status than the smoker. That is, controlling for the wife's 'quality', the smoking habit of the husband should be *positively* correlated with his status. This prediction is obviously specific to our simplified framework in which smoking prevalence is independent of education. In practice, educated people are less likely to smoke, especially in

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

the male population (Table 4). The prediction should therefore be restated as follows: the *conditional* correlation between male education and smoking habit, *given the education of non-smoking wives,* should be less negative than the unconditional one. Note, furthermore, that this pattern is male-specific; it should *not* hold true for women.

4 Data Description

4.1 The sample

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

The CPS is a series of monthly cross sections, with a short longitudinal component. Individuals in the sample are interviewed eight times—four times, followed by a break of eight months, and then interviewed for the same four months the following year. As such, it is possible to match observations of the same individuals across months, using the household and person identification codes, along with the month-in-sample information. However, several observations are dropped due to the specific design of the rotation samples by 4-month periods. In addition, we also check for age, gender and race, to ascertain that the merged observations consistently belong to the same individual.[9]

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

The TUS-CPS is a National Cancer Institute (NCI)-sponsored survey of tobacco use and policy information that has been administered as part of the Current Population Survey (CPS) since 1992.[10] It is considered a key and reliable source of national, state, and sub-state level data on smoking and other tobacco use in US households, which is widely used in medical research on cancer and other consequences of smoking (e.g., Delnevo and Bauer, 2009; Mills, Messer, Gilpin, Pierce, 2009). It provides data on a nationally representative sample of about 240,000 civilian, non-institutionalized individuals ages 15 years and older.

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

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

Our main sample of husbands and wives consists of white couples, where the wife is between

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

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

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

From the Tobacco Use Supplement, we retrieve information on the smoking status of each individual. Specifically, the respondents are asked whether and how often they smoke, whether they have smoked at least 100 cigarettes in their lifetime, and the actual number of cigarettes they smoke.[11] 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. The TUS vary by year in terms of the battery of questions concerning smoking, at times also covering quitting behavior and smoking-related health problems. However, the type of

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

information needed for our study on smoking status is available in every supplement under analysis.

Finally, each respondent answers these questions, so that each spouse directly *reports* his/her information. Moreover, self-reporting of smoking habits is considered a reliable source of information, as it is found to be validated by measured serum cotinine levels (Caraballo, Giovino, Pechacek, Mowery, 2001).

4.2 A first look at the data

The main characteristics of the data are described in Tables 1-4 and Figures 3 and 4. We present the summary statistics of married and single individuals, and the corresponding ones by smoking type of couple, the observed matching patterns by smoking status, the correlations of smoking status and education by gender and marital status, and the kernel distributions of education by gender and marital status.[12] A preliminary look at the data suggests that the smoking prevalence is higher for men than for women, with 22% of husbands smoking versus 17 % of wives (25 and 21 % for never-married, respectively), consistently with the gender gap reported by the National Center for Health Statistics (NCHS, 2010). Tables 1 and 2 also show that women are more educated than men across all smoking categories. The health status is very similar across spouses, and higher in couples where none is smoking than in those where both spouses are smoking. The average number of children under six years old is about 0.80 for couples and 0.50 for never-married individuals.

Table 3 reports the observed matching by smoking status for husbands and wives. There is strong assortative mating by smoking status: about 72% of couples have non-smoking spouses, and 10% consists of smokers. This is in line with evidence on marital sorting by smoking status in the UK (Clark and Etilé, 2006). However, a new and particularly interesting insight is provided by looking at "mixed" couples where one spouse is a smoker and the other

[12]See also Figures 1A-6A in the Appendix for the distributions of education broken down by smoking status for married individuals.

one is not. Our data reveal that there are fewer "mixed" couples where the wife smokes than vice-versa, 6.50% versus 11.71%, so the ratio is 0.55 (s.e.=0.029), which is statistically significantly lower than the 0.71 (s.e.=0.021) implied by the sole difference in relative smoking prevalence.[13] This finding supports our theoretical framework, in which the matching of a smoking man to a non-smoking woman happens because of the shortage of smoking women relatively to smoking men, given that a smoking man would prefer a smoking spouse. At the same time, the opposite match of a smoking woman to a non-smoking man would be far less frequent, given that all smoking women, who are in short supply, would end up marrying a smoking man.

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

To further explore the relationship of education with smoking status and gender, Figures 3 and 4 and A1-A6, present the kernel distributions of education for men and women, overall and for smokers and non-smokers. They exhibit a tri-modal shape, at high-school degree, 2 years of some college, and college degree for both smokers and non-smokers, even though smokers have lower education. Women appear to be more educated than men, although the two distributions seem quite close to each other, regardless of smoking and marital status.

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

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

Smoking and education are negatively correlated, and the male and female distributions of education are fairly similar to each other, regardless of smoking status. The following Section will use regression analysis to test the main implications of our model.

5 Empirical Analysis

5.1 Main results

Table 5 presents evidence of sorting by education within smoking types of young couples, whose husbands are between 24 and 34 years old, and wives are 22 to 32. In this Table we regress own education on spouse's education controlling for own age, year- and state-fixed effects. For each group of couples, there is assortative mating by education. Although assortative mating by education has been extensively documented in the literature (Lam, 1988; Pencavel, 1998; Qian, 1998; Mare, 2008), here we show that it holds true within each spouses' smoking category, and with magnitudes comparable to the estimated educational sorting in the US (Pencavel, 1998). Perhaps more interesting is the fact that our estimates also suggest that there is a gradient in assortative mating: stronger for couples were none of the spouses smoke (0.65), and weaker for those were both spouses smoke (0.45).

The main empirical results of the paper are presented in Table 6. It contains a series of regressions of young couples in which either wife's or husband's education is the dependent variable and spouse's education and smoking status are the explanatory variables, controlling for own age, year- and state-fixed effects, broken down by own smoking status. A first prediction is that non smoking wives are more likely to have a smoking husband when their 'quality' is low. Indeed, column (1) shows that among non-smoking wives those with smoking husbands have on average 0.14 fewer years of completed education than those with non-smoking husbands. In other words, a smoking husband provides a negative signal about the education of a non-smoking wife. This pattern is not observed for smoking women: column (2) indicates that there is no statistically significant difference in the average years of completed education between those who marry a smoker and those who marry non-smoker ones and the coefficient has a much lower magnitude than in column (1). Note that the estimates in these two columns are consistent with our model's main idea and predictions: own smoking is a negative attribute for someone who marries a non-smoker, but it is neutral when marrying a smoker.

Regarding husbands, however, the prediction was reversed (and somewhat less intuitive): it is now a *non smoking* wife which signals the lower index of a *smoking* husband. This is confirmed by the findings in column (4): among smoking husbands those who marry smoking wives have on average 0.16 *more* years of completed education than those with non-smoking wives. Again, we do not see such a pattern for non-smoking husbands: those with non-smoking wives have on average 0.21 more years of completed education than those with smoking wives, column (3). The magnitudes of the corresponding coefficients all represent sizable correlations, and in particular the sign of the coefficient in column (4) is positive, and opposite to the standard negative gradient between own smoking and own education.

Overall, Table 6 provides support for the two main testable implications of our model, namely: (i) among non-smoking women those who marry smoking men are less educated, and (ii) among smoking men those who marry smoking women are more educated.

Finally, Table 7 focuses on husbands of non-smoking women, and investigates the prediction that, even though smoking habits and education are negatively correlated overall, the magnitude of this effect should be reduced for men once we control for their wives' education (prediction 4). Indeed, such a control significantly decreases the magnitude of the correlation of male education and smoking status. This supports the theoretical prediction that among non-smoking low-quality women, their smoking husbands are "more" educated. Columns (3) and (4) reinforce our evidence, showing that this pattern does not hold for the wives of smoking men; the effect is actually opposite in that case.

5.2 Sensitivity Analysis

We proceed to a few robustness checks. First, we try to take into account unobserved heterogeneity by adding controls for individual and household characteristics, and for the interactions between state and year fixed effects. Our main results (signs, magnitudes, and significance) are robust to the inclusion of health status and number of children. Specifically, we construct a dummy variable for very healthy status (one if the status is excellent or very good, zero if good, fair or poor), and consider the number of own children in the family who are under age 6, as our analysis concerns young couples. The main regression specifications with these additional controls are reported in Table A1: controlling for number of children and health status does not change our estimates. Second, we relax the definition of smoker, by not considering the criterion based on the 100 cigarettes smoked in a lifetime. Although the number of smokers increases, the patterns of assortative mating by smoking status and the relationships between own education and spouse's smoking remain unchanged. In terms of years of schooling, recoding the education variable following Jaeger (1997) confirms our findings, as shown in Table A2.

Third, we slightly modify the age group under analysis, including younger women whose age is between 20 and 30, and younger men whose age is between 22 and 32, to add younger married couples who are likely to be newly-weds. This sample yields the same patterns of results as our main estimates, as shown in Table A3, which reinforces our claim that the young couples in our sample represent recent marriages and the actual matching in the marriage market. The information on duration of marriage or age at marriage is not available in the CPS in any of the years under consideration. However, our choice of very young couples along with the very large sample size of this data set allows us to focus on recently married couples, that is on the matches formed on the marriage market, with which our analysis is concerned.

To further explore the issue of recent marriages, we alternatively examine a very different data source which provides the information on duration of marriage, i.e. the Panel Study of

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

It is important to acknowledge that the PSID is a very different dataset than the CPS. First, the PSID is a panel, not a cross-section. Second, its sample size is much smaller. Moreover, its effective sample size is, given its panel structure, even smaller. Hence, the PSID and its availability of the relevant information on marriage duration comes with the price of a huge reduction in sample size compared to the CPS. This dramatic reduction is exacerbated in our analysis, which is characterized by sub-dividing the sample according to spouse's smoking status.[14] Additionally, in the PSID all the variables are reported by the head of the household, including the information on the wife. The wives' smoking behavior is therefore proxy-reported by their husbands, while in the CPS it is self-reported.

Nevertheless, we replicate our main results on the positive sorting by smoking status, with the asymmetric prevalence of "mixed" couples (Table A4), and on the relationship between husbands' education and spouse's smoking status, and between non-smoking wives' education and their husbands' smoking status (Table A5). The estimates are much noisier and not statistically significant. However, the signs of the coefficients at stake are the same as in our

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

main CPS estimates, and their magnitudes are similar or higher. The observed patterns when using PSID data are consistent with our marriage market predictions and interpretation.[15]

6 Conclusions

We devise a matching model where individuals are characterized by heterogeneous preferences and multidimensional attributes, one of which discrete. The stable match and the testable predictions on who marries whom are derived. In light of the role of smoking in the family, we apply our framework to explore the interaction between smoking status and education at the time of marriage. We study for whom and to what extent smoking is perceived as a personal defect penalized in the formation of couples, on top of the well-known health costs and consequences (CDC, 2008). Considering these additional aspects and consequences of smoking behavior may help health policy makers in the understanding of the actual role of smoking and of smokers. Indeed, the vast socioeconomic literature on the effects of smoking in the family has focused on the intergenerational transmission of smoking habits between parents and children, with distinct gender effects and interactions (e.g., Loureiro, Sanz and Vuri, 2010; Maralani, 2009).

We study the matching between smokers and non-smokers, building a model in which non-smokers prefer to marry non-smokers. In other words, *ceteris paribus*, smoking is perceived as a *bad* characteristic on the marriage market by non-smokers. On the other hand, we assume that smoking is perceived as a *neutral* characteristic on the marriage market by smokers.

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

Individuals sort by education level, but also by smoking status. Given the different preferences for spousal smoking between smokers and non-smokers, and the gender asymmetry by smoking prevalence, with more smoking men than smoking women for all education levels, smoking women and non-smoking men are in *short supply*. We show that at the top of the 'quality' distribution, matching is pure and assortative by index *and* smoking habits; that is, educated non-smoking men marry educated non-smoking women and educated smoking women marry educated smoking men. Below some quality threshold, however, matching patterns become more complex. While non-smoking men still marry a non-smoking spouse, smoking men may be matched with either a smoker or a non-smoker. Equivalently, the husband of a smoking woman is still a smoker spouse; but a non-smoking wife may be married to either a smoker or a non-smoker. In that case, the smoker is typically of better 'quality' than the non-smoker.

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

An obvious limitation of our model is its static nature: we assume that an individual's smoking habit is given and cannot be changed. The next step will be to study the dynamic implications of the model, in a framework where individual may change their smoking habit at some (heterogeneous) cost. It is important to note that the static version provides the basic ingredient that will be needed for that purpose. Indeed, our equilibrium conditions exactly pin down the intrahousehold allocation of welfare as a function of the education and smoking habits of the spouses. We can therefore exactly characterize the *incentives* each agent is faced with, i.e. the expected gain that they would reach on the marriage market by becoming a non smoker (and that will be traded off with the subjective cost of quitting). This is left for future work.

Appendix: Proof of Proposition 1

Since the stable matching maximizes aggregate surplus, its qualitative features, as well as the values of the various thresholds, can be derived from standard, variational calculus arguments. These, however, require long and tedious calculations. Here, we adopt a more direct approach. We first *assume* that the equilibrium is as described in the Proposition, and we provide a complete characterization, including the resulting allocation of surplus between members; we then check that the latter satisfy the stability conditions.

Assuming $\lambda \geq s_M / (1 + s_M)$, define :

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

and

$$p = \frac{s_M(2 - s_M - s_W) - \lambda(1 - s_M)(s_M + s_W)}{(\lambda + s_M - \lambda s_M)(1 - s_W)}$$

One can readily check that all these variables belong to the interval $[0, 1]$.

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

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

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

or equivalently:

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

In particular,

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

- Similarly, a smoking woman with index $x \geq X$ is matched with a smoking man with index y such that

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

In particular,

$$Y' = \frac{s_W}{s_M}X + \frac{s_M - s_W}{s_M}$$

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

$$(1-s_W)(1-p)(X-x) = (1-s_M)(Y-y)$$

or

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

  - the number of smoking men above y equals the number of smoking women above

x plus that of non-smoking women above x who marry a smoker:

$$((1 - s_W)\, p + s_W)\,(X - x) = s_M\,(Y' - y)$$

therefore

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

Women can therefore be classified into four categories, by crossing their index (above versus below X) and their smoking habits; lets us denote these by W_{HN}, W_{HS}, W_{LN} and W_{LS}, where H (resp. L) reads 'above (below) X'. Similarly, non-smoking men are either above or below Y and smoking men are either above or below Y', generating four categories M_{HN}, M_{HS}, M_{LN} and M_{LS}. Note that at the stable matching, couples can only belong to one of the following five pairs of categories: $(W_{HN}, M_{HN}), (W_{HS}, M_{HS}), (W_{LN}, M_{LN}), (W_{LS}, M_{LS})$ and (W_{LN}, M_{LS}).

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

$$u_N(x) + v_N(y) \geq \frac{(x+y)^2}{2}$$

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

$$u_N(x) = \max_y \left(\frac{(x+y)^2}{2} - v_N(y) \right)$$

and from the envelope theorem:

$$u_N'(x) = \frac{\partial}{\partial x} \left(\frac{(x+y)^2}{2} \right)$$

the partial being taken at the point $y = \phi_N(x)$. Therefore for $x < X$:

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

and

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

where K is an integration constant. Then for $y < Y$:

$$\begin{aligned} v_N(y) &= \frac{(x + \phi_N(x))^2}{2} - u_N(x) \\ &= \frac{1}{2} \frac{(2 - p - s_M - s_W + ps_W)}{(1 - s_W)(1 - p)} y^2 - K \end{aligned}$$

Also, agents must prefer marriage to singlehood; this requires:

$$\begin{aligned} \frac{1 - s_M + (1 - s_W)(1 - p)}{1 - s_M} \frac{x^2}{2} + K &\geq \frac{x^2}{2} \\ \frac{(2 - p - s_M - s_W + ps_W)}{(1 - s_W)(1 - p)} \frac{y^2}{2} - K &\geq \frac{y^2}{2} \end{aligned}$$

for all (x, y); this implies

$$\frac{1 - s_M}{(1 - s_W)(1 - p)} \frac{y^2}{2} \geq K \geq -\frac{x^2}{2} \left(\frac{1 - s_W}{1 - s_M} (1 - p) \right)$$

for all (x, y), therefore $K = 0$.

Similar computations give:

- for $x < X$, $y < Y'$:

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

  and

  $$v_S(y) = \frac{\lambda}{2} \frac{p + s_M + s_W - ps_W}{p + s_W - ps_W} y^2$$

and one can readily check that

$$u_S(x) \geq \frac{x^2}{2}, v_S(y) \geq \lambda \frac{y^2}{2}$$

so that all agents prefer marriage than singlehood.

- for $x \geq X$:

$$\begin{aligned} u_N(x) &= x\left(\frac{2-s_M-s_W}{1-s_M}\frac{x}{2}-\frac{s_M-s_W}{1-s_M}\right) \\ &+\frac{1}{2}\frac{(s_M-s_W)^2}{1-s_M}\frac{\lambda+s_M-\lambda s_M}{s_M(2-s_M-s_W)-\lambda(1-s_M)(s_M+s_W)} \\ v_N(y) &= \frac{(1-s_M)(2-s_M-s_W)y^2+2(1-s_M)(s_M-s_W)y}{2(s_M-1)(s_W-1)} \\ &-\frac{1}{2}\frac{(s_M-s_W)^2}{1-s_W}\frac{\lambda-s_M+\lambda s_M}{s_M(2-s_M-s_W)-\lambda(1-s_M)(s_M+s_W)} \end{aligned}$$

and

$$\begin{aligned} u_S(x) &= \lambda x\left(\frac{s_M+s_W}{s_M}\frac{x}{2}+\frac{s_M-s_W}{s_M}\right) \\ &-\frac{1}{2}\frac{\lambda}{s_M}(s_M-s_W)^2\frac{\lambda+s_M-\lambda s_M}{s_M(2-s_M-s_W)-\lambda(1-s_M)(s_M+s_W)} \\ v_S(y) &= y\lambda\left(\frac{s_M+s_W}{s_W}\frac{y}{2}-\frac{s_M-s_W}{s_W}\right) \\ &+\frac{1}{2}\frac{\lambda}{s_W}(s_M-s_W)^2\frac{2-\lambda-s_M+\lambda s_M}{s_M(2-s_M-s_W)-\lambda(1-s_M)(s_M+s_W)} \end{aligned}$$

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

Starting with (W_{HN}, M_{HS}), we must for instance check that for $x \geq X, y' \geq Y'$:

$$P(x,y) = u_N(x) + v_S(y') - \lambda \frac{(x+y')^2}{2} \geq 0$$

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

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

which gives

$$\begin{aligned} x &= (s_M - s_W) \frac{\lambda + s_M - \lambda s_M}{s_M(2 - s_M - s_W) - \lambda(1 - s_M)(s_M + s_W)} = X \\ y &= (s_M - s_W) \frac{2 - \lambda - s_M + \lambda s_M}{s_M(2 - s_M - s_W) - \lambda(1 - s_M)(s_M + s_W)} = Y' \end{aligned}$$

Since $P(X, Y') = 0$ by definition, the condition is satisfied.

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

References

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

[2] Browning, M., P-A. Chiappori and Y. Weiss (2011), *Family Economics*, Cambridge University Press.

[3] 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, 807–814.

[4] CDC (2008) Smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 2000–2004. MMWR 2008;57(45):1226–8. Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5745a3.htm.

[5] 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.

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

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

[8] Chiappori, P-A., S. Oreffice and C. Quintana-Domeque (2010) Fatter attraction: anthropometric and socioeconomic matching on the marriage market. IVIE WP-AD 2010-23.

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

[10] Coles, M., M. Francesconi (2007) On the emergency of toyboys: equilibrium matching with ageing and uncertain careers, IZA discussion paper 2612, Germany: Bonn.

[11] Delnevo, C.D., Bauer, U.E. (2009) Monitoring the tobacco use epidemic III: The host: data sources and methodological challenges, *Preventive Medicine*, 48, 16-23.

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

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

[14] Fox, J.T. (2010), Identification in matching games, *Quantitative Economics*, 1, 203-254.

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

[16] Glymour, M. Maria, Triveni B. DeFries, Ichiro Kawachi, and Mauricio Avendano (2008) Spousal Smoking and Incidence of First Stroke The Health and Retirement Study, *American Journal of Preventive Medicine*; September, 35(3): 245–248.

[17] 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.

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

[19] Lam, D. (1988) Marriage Markets and Assortative Mating with Household Public Goods, *Journal of Human Resources*, 23,462–487.

[20] Loureiro, M.L:, Sanz-de-Galdeano, A:, Vuri, D. (2010) Smoking Habits: Like Father, Like Son, Like Mother, Like Daughter?, *Oxford Bulletin of Economics and Statistics*, in press.

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

[22] Mannino DM, Siegel M, Rose D, et al. (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*; 6:296–305.

[23] Maralani, V. (2009) An Unequal Start: The Alignment of Education and Smoking in Families of Origin, unpublished manuscript.

[24] Mare, Robert, D. (2008) Educational Assortative Mating in Two Generations, University of California at Los Angeles, unpublished manuscript.

[25] Mills, A.L., Messer, K., Gilpin, L.A., Pierce, GP. (2009) The effect of smoke-free homes on adult smoking behavior: a review, *Nicotine Tobacco Research*, 1131–41.

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

[27] Oberg, M., Jaakkola, M., Woodward, A., Peruga, A., Pruss-Ustun, A. (2010), Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries. *Lancet*; 1-8

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

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

[30] Qian, Z. (1998) Changes in Assortative Mating: The Impact of Age and Education, *Demography*, 279–292.

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

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

[33] Sutton, G. C. (1980), Assortative marriage for smoking habits, *Annals of Human Biology*, 7, 449.

[34] Sutton, G.C. (1993), Do men grow to resemble theri wives or vice versa?, *Journal of Biological Science,* 25, 25–29.

[35] Venters Maurine H., David R. Jacobs Jr., Russell V. Luepker, Lois A. Maimaw and Richard F. Gillum (1984) Spouse concordance of smoking patterns: the Minnesota heart survey. *American Journal of Epidemiology*, 120, 4: 608–616.

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

	Married	
	Men	Women
Age	29.44	27.80
	(2.80)	(2.77)
Education	13.66	13.82
	(2.38)	(2.30)
Smoke	0.22	0.17
	(0.41)	(0.37)
# Children under age 6	0.82	
	(0.85)	
N	12,035	
	Singles (never married)	
	Men	Women
Age	29.22	27.17
	(3.12)	(3.14)
Education	13.63	13.78
	(2.41)	(2.33)
Smoke	0.25	0.21
	(0.43)	(0.41)
# Children under age 6	0.50	0.55
	(0.76)	(0.78)
N	28,086	29,102

Note: Sampling weights are used.

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

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

Note: Sampling weights are used.

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

Weighed % and (unweighed number of observations)

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

Note: Sampling weights are used.

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

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

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

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

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

Figure 3:
Kernel densities of education for singles by gender

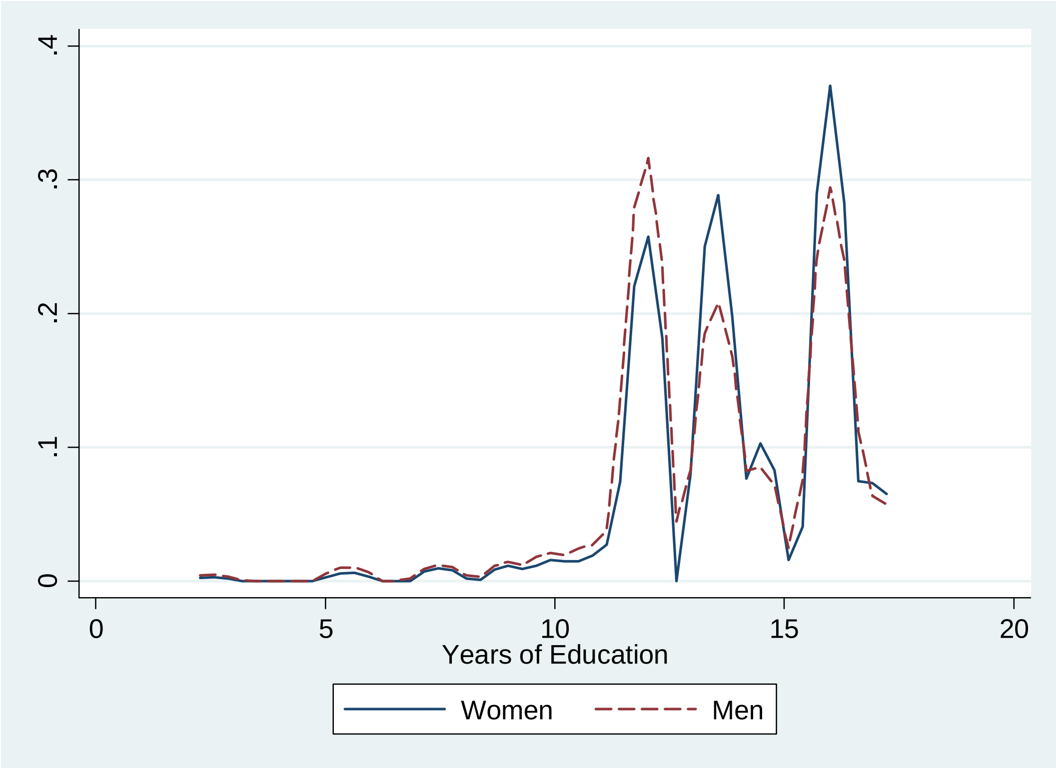


Figure 4:
Kernel densities of education for married couples by gender

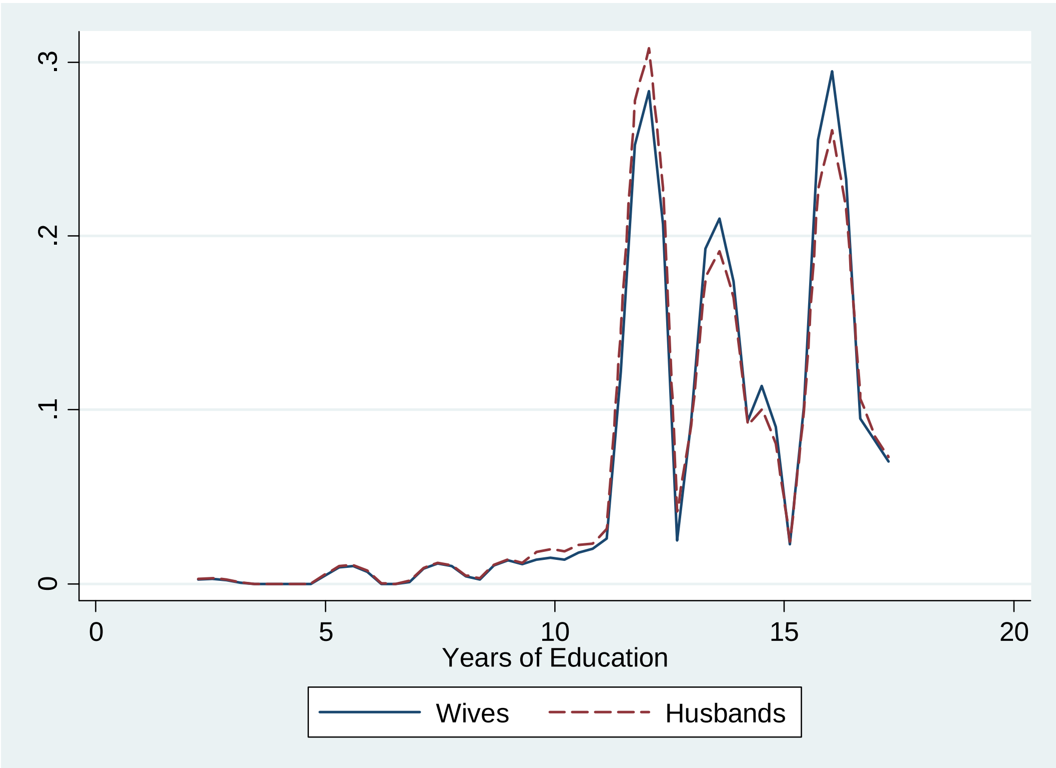


Table 5:
Sorting by education
Husband's age 24-34, Wife's age 22-32.
CPS 1996–2007.

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

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

Note: All regressions include: own age, year and state fixed effects. Reference categories: 2007 and District of Columbia. Sampling weights are used. Robust standard errors.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

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

	Wife's Education		Husband's Education	
	Non-Smoker	Smoker	Non-Smoker	Smoker
Spouse's Education	0.630***	0.473***	0.684***	0.556***
	(0.012)	(0.027)	(0.013)	(0.026)
Spouse Smokes	−0.141**	−0.025	−0.209***	0.160**
	(0.060)	(0.086)	(0.074)	(0.076)
N	10118	1917	9477	2558
R^2	0.47	0.29	0.46	0.36

Note: All regressions include: own age, year and state fixed effects. Reference categories: 2007 and District of Columbia. Sampling weights are used. Robust standard errors.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

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

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

Note: All regressions include: own age, year and state fixed effects. Reference categories: 2007 and District of Columbia. Sampling weights are used. Robust standard errors.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

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

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

Note: All regressions include: own age, year and state fixed effects. Spouse's health status is controlled for by a dummy variable: 1 if excellent or very good health, 0 if good, fair or poor. Reference categories: 2007 and District of Columbia. Sampling weights are used. Robust standard errors.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

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

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

Note: All regressions include: own age, year and state fixed effects. Spouse's health status is controlled for by a dummy variable: 1 if excellent or very good health, 0 if good, fair or poor. Reference categories: 2007 and District of Columbia. Sampling weights are used. Robust standard errors.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

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

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

Note: All regressions include: own age, year and state fixed effects. Reference categories: 2007 and District of Columbia. Sampling weights are used. Robust standard errors.
*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1

Table A4:
Observed Matching

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

Weighed %

I. Full sample

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

II. Recently married: marital duration ≤ 4 years

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

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

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

I. Full sample

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

II. Recently married: marital duration ≤ 4 years

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

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

Figure 1A

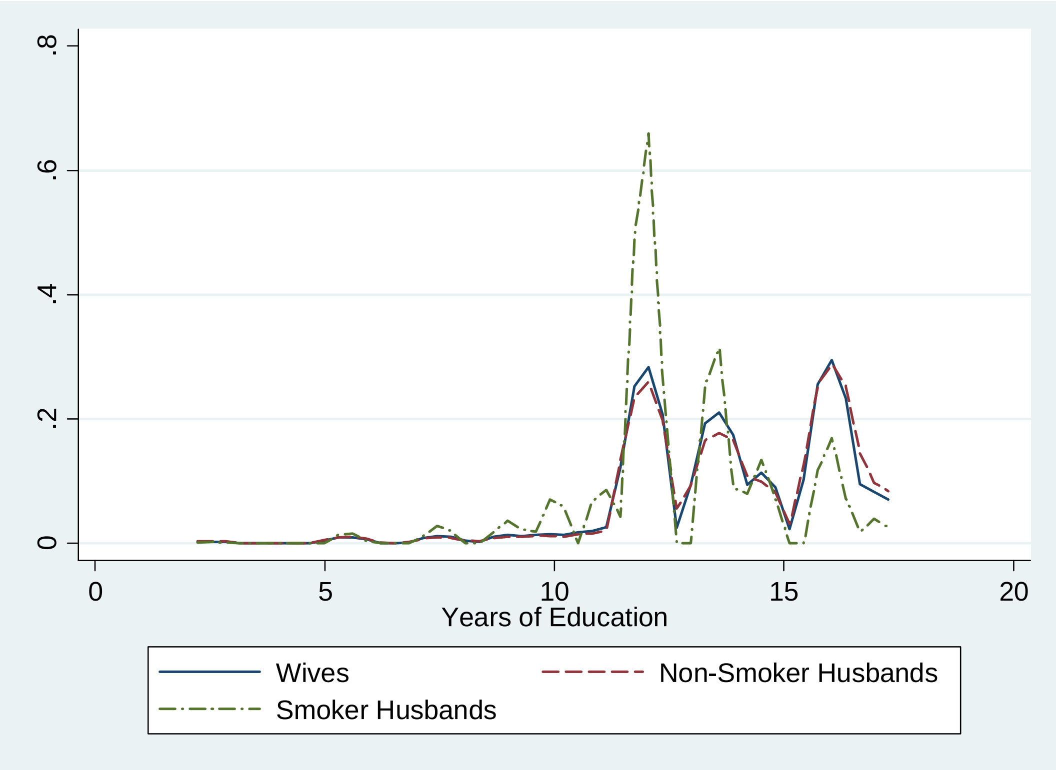


Figure 2A

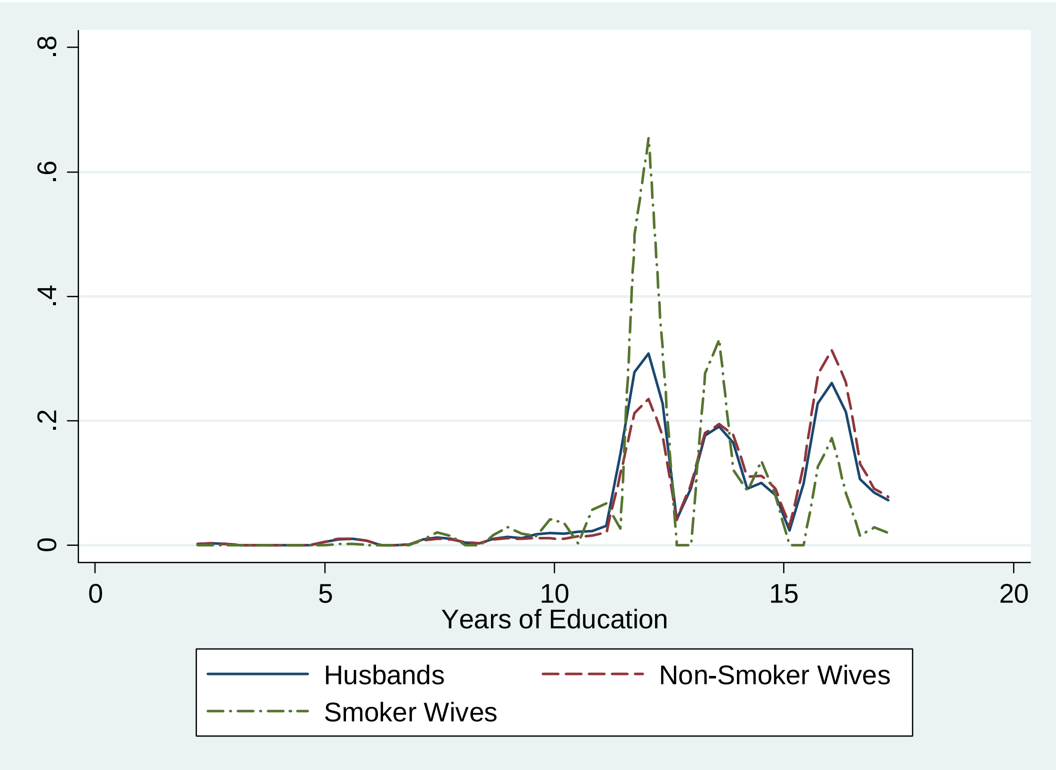


Figure 3A

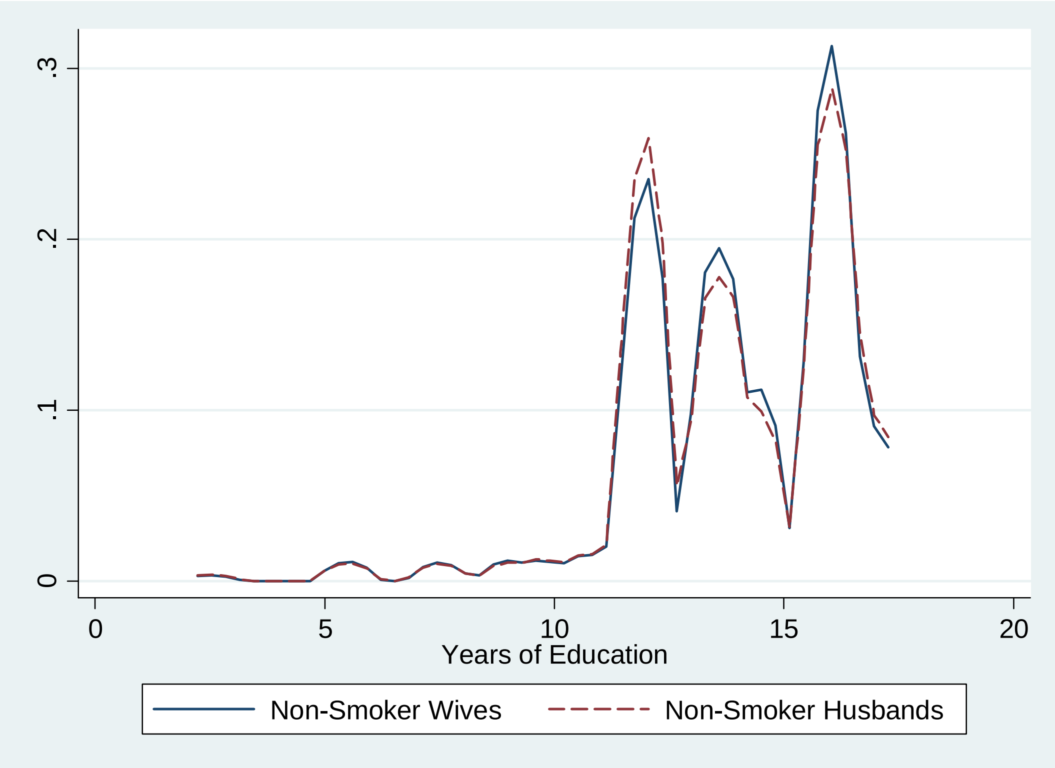


Figure 4A

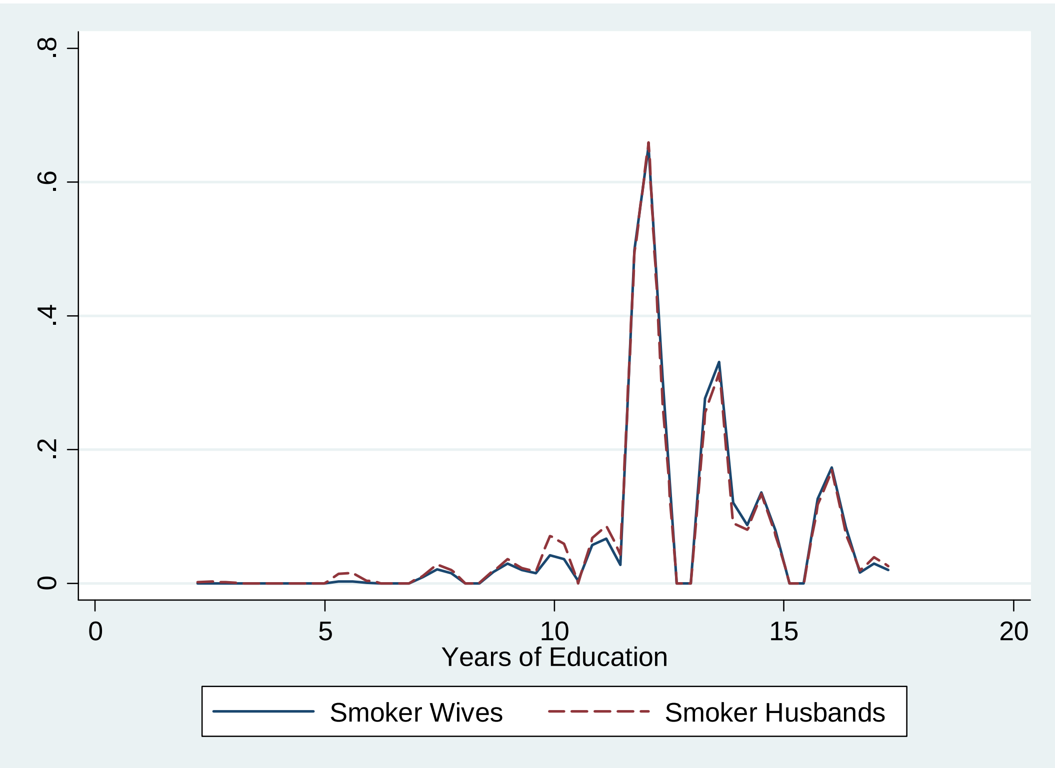


Figure 5A

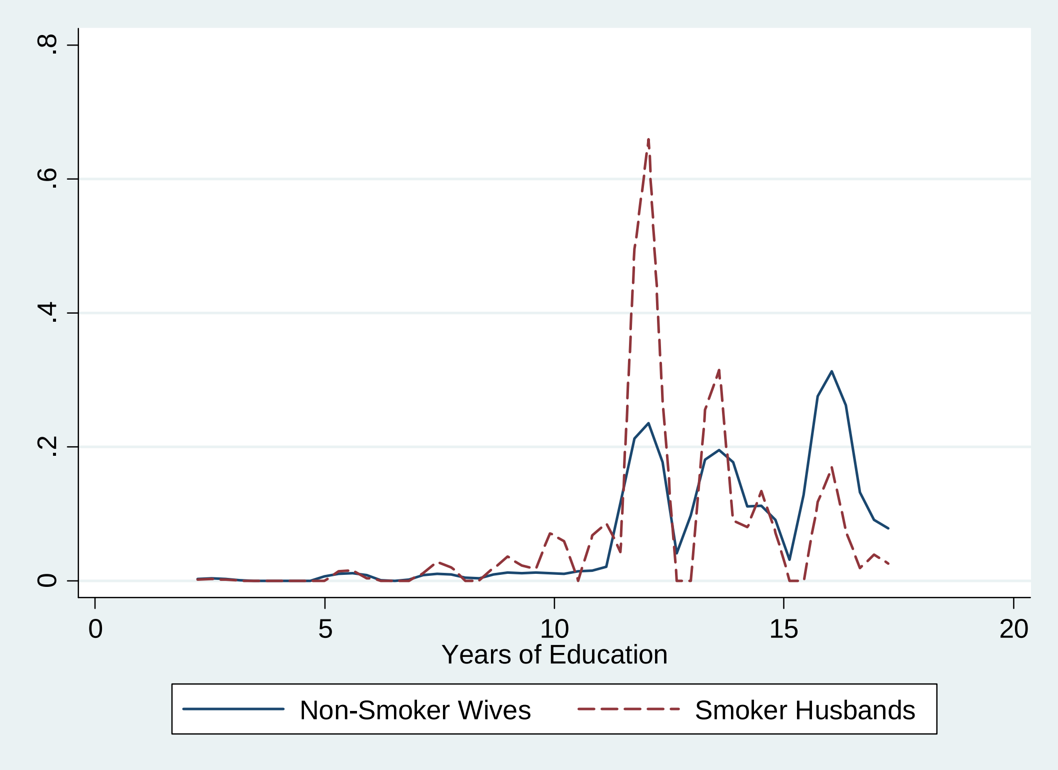


Figure 6A