Don't give up on me baby: spousal correlation in smoking behaviour

Andrew E. Clark, Fabrice Etilé

DON'T GIVE UP ON ME BABY:

SPOUSAL CORRELATION IN SMOKING BEHAVIOUR*

Andrew Clark
CNRS and DELTA (Joint Research Unit CNRS, EHESS, ENS), Paris, France

Fabrice Etilé**
INRA – CORELA, Paris, France

November 2003

Abstract

We use nine waves of BHPS data to examine interactions between spouses in terms of a behaviour with important health repercussions: cigarette smoking. Partners' behaviours may be correlated due to matching in the marriage market, bargaining within marriage, or information revealed by others' behaviour. Simple probit and bivariate probits reveal a positive correlation between partners' smoking participation, which is consistent with both matching and bargaining. Controlling for fixed effects allows us to distinguish between opposing interpretations. In our preferred specification, a bivariate probit with random effects, partners' behaviours are statistically independent: all of the correlation in smoking status works through the correlation in individual fixed effects. As such, we believe that the correlation in the raw smoking data reflects matching on the marriage market, rather than bargaining within the couple.

JEL Codes: C33, D83, I12, I18.

Keywords : Smoking, Matching, Bargaining, Learning, Health.

***The BHPS data were made available through the ESRC Data Archive. The data were originally collected by the ESRC Research Centre on Micro-social Change at the University of Essex. Neither the original collectors of the data nor the Archive bear any responsibility for the analyses or interpretations presented here. We thank Andrew Jones, François Legendre, Maarten Lindeboom, Nigel Rice, Michael Visser, Frank Windmeijer, and seminar participants at INRA-CORELA, the 20th Journées de Micro-économie Appliquée (Montpellier), the York Seminar in Health Econometrics, and the 12th European Workshop on Econometrics and Health Economics (Menorca). We are particularly grateful to David Madden for detailed comments.

****Corresponding author** : Fabrice Etilé, INRA-CORELA, 65 Boulevard de Brandebourg, 94205 Ivry-sur-Seine cedex, France. Tel: (33)-1-49-59-69-86. Fax: (33-1-49-59-69-90. E-mail: etile@ivry.inra.fr.

Don’t Give Up on Me Baby: Spousal Correlation in Smoking Behaviour

Andrew Clark and Fabrice Etilé

1 Introduction

Over the past fifteen years, economists have started to relax the assumption that the most salient level of analysis is that of the individual. This evolution has been felt particularly strongly in the areas of labour supply, where the collective model has made great inroads into the established literature (see Chiappori, 1992), and in political welfare economics, where the issue of fungibility of welfare receipts has come to the forefront (see Blank, 2002).

In this light, we consider interactions between spouses in terms of an (*a priori*) observable behaviour with important repercussions on health: cigarette smoking. There are a number of reasons why partners’ behaviours may be correlated: these are broadly divided into matching in the marriage market[1], strategy within marriage, and the information revealed by others’ behaviour when there is uncertainty.

There are at least two important consequences of interactions between household members. The first is purely descriptive or positive: if household members interact then we build better, in the sense of more realistic and accurate, economic models if we take this fact into account. The second is normative: absent special cases, optimal policy is bound to depend on the nature of these interactions. In terms of the current paper’s subject matter, we can ask whether it is sufficient, or more efficient, to target one person per household in terms of health education (or some other intervention), as opposed to all household members.

This paper uses British household panel data to look at the correlation in smoking behaviour between partners. We consider both participation in smoking and, using the panel aspect of the data, the quit decision. We uses nine waves of British Household Panel Survey (BHPS) data to measure the extent of this correlation, both in cross-section and, accounting for individual heterogeneity, panel regressions.

If partners act strategically within marriage, then we expect their behaviours to be correlated, even when individual fixed effects are controlled for (these latter will only pick up the time-invariant correlation between spouses’ preferences). We will estimate a number of different specifications to test for this type of influence. The first, simple, specifications

[1] We use the term “marriage” loosely here, as we consider both legally married couples and those who live together.

(simple and bivariate probits) will not control for individual fixed effects. We then control for unobserved individual heterogeneity by considering quit equations, random effect probits, and (our preferred specification) a random effect bivariate probit.

The paper is organised as follows. We first run briefly through some ways in which to think of correlation between partners' behaviour: matching, household bargaining and social learning. Section 3 then presents the data, and Section 4 the econometric approach. Our main findings are reported in Section 5. We show that matching explains the spousal correlation in smoking behaviour. Section 6 proposes further empirical results regarding learning and the presence of children. Section 7 concludes.

2 Theory

We briefly present here a number of arguments which may account for correlation in partners' smoking behaviour. These are not specific to cigarette consumption, and may apply more generally to other risky behaviours. They fall into three different groups: individual fixed effects; interactions due to bargaining within marriage; and social learning about smoking's health risks from the observation of one's partner.

2.1 Matching in the Marriage Market

The first theoretical consideration refers to the process of matching on the marriage market. It seems uncontroversial to say that people tend to marry those who think and behave like them. In this optic, smoking is an easily observed signal of (less-easily observed) general preferences with respect to other activities and goods, such as parties, concerts, healthy foods and sport[2]. This correlation can also be interpreted with respect to preferences for the present (*i.e.* discount rates). Contoyannis and Jones (2001) show that a number of such lifestyle variables are correlated with each other. This type of matching with respect to (unobserved) individual fixed effects (or lifestyle preferences) likely applies especially to teenagers and young adults.

[2] Recently, Farrell and Shields (2002) and Leonard and Mudar (2003) have uncovered empirical evidence consistent with spousal matching with respect to sporting activity and drinking respectively. Latkin *et al*. (1995) find evidence of spousal correlation in injecting habits amongst drug users.

Equally, matching may concern observable variables like income or social class. Here, a number of studies have found a wage penalty for smokers (see Levine *et al.*, 1997, and Van Ours, 2002). This will make non-smokers more attractive on the marriage market, due to their higher expected life-cycle income. The resulting matches will then exhibit a certain degree of homogamy, whereby smokers are matched with smokers, and non-smokers with non-smokers; there will still be smoker: non-smoker matches, of course, as this is not the only relevant criterion (see also Wilson, 2001). One reason for mixed matches is that the marriage market is not gender-neutral. In general, the impact of health on labour market outcomes is more pronounced for men than for women (see Currie and Madrian, 1999). As such, male smokers will be more heavily penalised than female smokers when there is matching by income.

A last related point concerns preferences over time spent together with one's partner. Aversion to spending time alone in widowhood, will also lead to preferences for partners whose life expectancy coincides with ones own. Here again, preferences for the present come into play.

The implication of matching in the marriage market is that, first, both male and female smoking equations must include individual fixed effects. Second, we expect these fixed effects to be correlated.

2.2 Collective Decision-Making

The second interpretation of the correlation between partners' behaviour relies on some ongoing collective decision-making process within the household. The key point here, and that which allows this process to be distinguished from matching, is that the outcome of this bargaining, measured by cigarette consumption, is susceptible to change over time.

The set-up is a battle of the sexes, where partners do not have the same preferences, but value choosing the same outcome. The payoffs will then depend on both the private pleasure from smoking and the value of homogamy. The latter can be understood as the value of time spent together in good health.

The Nash equilibrium can be anywhere in the Smoking-No Smoking space, depending on the relative size of the private and public payoffs. There are two implications. First, the value of making the same choice will yield some correlation in partners' smoking status. Second, life-cycle events can alter the payoffs. In particular, the arrival of children may raise the value of time spent together. Hence, we expect the presence of children within the household to increase adults' incentives to remain in good health. This shift in the relative size of the public and private payoffs makes it more likely that the Non-Smoker/Non-Smoker solution Pareto-dominates[3]. However, adjustment costs will dampen these incentives. In addition, gender differences may push towards lower correlation in spousal smoking: the mother's smoking may have a greater impact on children's health (especially during pregnancy) than the father's smoking.

Another application of the battle of the sexes is in terms of endogenous life expectancy. Aversion to spending time on one's own in widowhood (which should increase with the duration of married life, as the hypothetical event approaches) may lead spouses to under-invest in health. If partner's health investments are not perfectly observable, smoking may be interpreted as a signal of the partner's commitment to healthy behaviours. In this signalling game, spouses will again tend to co-ordinate on equilibria where both smoke or no-one smokes[4].

The empirical implication of household bargaining is that partners' smoking statuses may be correlated, even after controlling for individual fixed effects.

2.3 Social Learning

The last explanation applies in a world where there is both uncertainty about smoking's dangers and social learning. Health changes for one partner who smokes may change the perceived riskiness of smoking for both. The correlation in one of the key parameters determining smoking leads naturally to a correlation in observed behaviour (see Clark and Etilé, 2002).

2.4 Empirical predictions

These different arguments yield three main empirical predictions:

Prediction 1 (Matching and Bargaining): *Smoking status will be positively correlated between partners*. However, detailed information on fixed individual characteristics are required to distinguish matching from bargaining. As this kind of information is generally only partially present in surveys, we appeal to panel regression techniques to control for fixed unobservables.

---

[3] In non-cooperative couples there is of course no guarantee that the Pareto-dominant point is chosen. Also note that the basic battle of the sexes set-up only predicts that choices will be correlated, rather than saying which of the homogamous solutions will result.

[4] Formal models of strategic health investments within the household are developed in Bolin *et al*, (2001, 2002). However, issues such as interactions in the choice of the length of life are not investigated.

Prediction 2 (Bargaining and Learning): *If both A and B smoke, a negative health shock for A will have an ambiguous effect on B's consumption*: the expected future value of time spent with one's partner falls, which increases B's consumption, but B's subjective evaluation of the dangers of smoking (weakly) rises. An unanticipated negative health shock for a non-smoking A will increase B's consumption.

Prediction 3 (Bargaining): *The presence of children may change the degree of correlation between partners' smoking behaviour.*

We will attempt to test these predictions on long-run British panel data.

3 Data

3.1 The British Household Panel Survey

The British Household Panel Survey is an annual panel of roughly 10 000 individuals in around 5 000 different households in Great Britain. We use the first nine waves (1991-1999). All adults in the household are interviewed separately with respect to their socio-demographic characteristics, income, employment, and health. Further details of this survey are available at the following address: http://www.iser.essex.ac.uk/bhps.

3.2 Smoking in the BHPS

We consider all individuals who are observed over the nine waves, to estimate the probability of being a smoker as a function of past smoking status. This initial unbalanced subsample (Sample 1) includes 63530 observations (12467 individuals) of which:

- 21172 are not in a couple at both t-1 and t.
- 2814 change from not being in a couple to being in a couple (married or living together), or vice versa, between t-1 and t[5].
- 39544 remain with their partner between t-1 and t[6].

Smoking participation by marital status and sex is summarised in Table 1 (where "single" refers also to those who did not remain with the same partner between t-1 and t).

[5] There will likely be a number of couples who do not live together at t-1, and who subsequently live in the same house (either cohabiting or married) at t. As we cannot identify the first status, these individuals will be counted in the second category.

[6] A couple who live together at t-1, and are married at t, will appear in the third category.

Table 1: Smoking participation by marital status and sex.

	Smokers (%)	All
Men in couples	5143 (25.8%)	19890 (67.0%)
Single Men	3248 (33.1%)	9802 (33.0%)
All Men	**8391 (28.3%)**	**29692 (100%)**
Women in couples	4771 (24.3%)	19654 (58.1%)
Single Women	4261 (30.0%)	14184 (41.9%)
All Women	**9032 (26.7%)**	**33838 (100%)**
All	***17423 (27.4%)***	***63530***

Note that whilst 67% of men interviewed in this sample are in couples, this is only true for 58% of women. There is an 7.3% difference in the smoking participation of singles versus those in couples for men; the analogous figure for women is 5.7% per cent. It is possible that this reflects an indirect widowhood effect. Men in couples are about 9.5 years older than single men, while single women are 3 years older than women in couples. Widowhood explains these differences (32% of single women are widowed, against only 11% of single men).

Consider now couples who stay together for two consecutive years and for whom information on both partners (of different gender) is available (N = 19307 x 2). We use this subsample (Sample 2) to run bivariate probit estimations of current smoking behaviour as a function of spouses' past smoking decisions. We use lagged spouse smoking to avoid problems of simultaneity and identification.

Smoking participation for men and women in this sample is 25.9% and 24.2% respectively. Other descriptive statistics are presented in Appendix A. The crosstabulation of couples' smoking statuses in Sample 2 is shown in Table 2 below.

Table 2: Couples' smoking statuses.

		Male	
		Smoker	*Non-Smoker*
Female	*Smoker*	13.9%	10.3%
	Non-Smoker	12.0%	64.0%

The table reveals that there are more mixed couples (one smokes, the other does not) than there are matched smoking couples, suggesting that many factors, in addition to smoking, are important in couple formation. We can calculate conditional probabilities to illustrate the correlation in spouses' behaviours. Table 3 shows these for men, conditional on their partner's smoking status. The table should be read as follows: given that the woman smokes, the probability that the man smokes is 57.4%; given that the female partner does not smoke, the probability that the male partner smokes is only 15.8%.

Table 3: Conditional Probabilities.

		Female	
		Smoker	*Non-Smoker*
Male	*Smoker*	57.4%	15.8%
		Male	
		Smoker	*Non-Smoker*
Female	*Smoker*	53.9%	13.7%

The conditional probabilities for women tell almost the same story. Nevertheless, the descriptive odds ratio for women is slightly higher than that for men (3.93 and 3.63 respectively)[7]: male smoking is somewhat more of a risk factor for women than is female smoking for men. There is obviously a positive correlation between partners' smoking statuses, with something of a gender asymmetry[8]. In the remainder of this paper, we try to investigate this correlation in the light of section 2's theoretical considerations.

3.3 Accounting for selection bias

We note that there is likely some selection bias involved in moving from sample 1 to sample 2, as matching and smoking are correlated. We try to control for this by estimating a marital status selection equation on sample 1 as a function of education (3 dummies), labour force status (10 dummies), region and year, and regional unemployment rates by sex and year. These last variables are used to satisfy the exclusion restrictions. Unemployment rates are expected to be correlated with marital status as (i) being in couple is a form of insurance against well-being losses caused by future unemployment (positive correlation); and (ii) the average quality of offers on the marriage market falls as unemployment rises – therefore, the implicit cost of divorce increases. However, the rate of new marriages may fall as unemployment rises, and the net effect depends on the demographic structure.

The dependent variable equals 1 if the individual remains with his/her partner between *t-1* and *t*, and 0 otherwise. From this selection equation (estimated by gender) we compute a

[7] The descriptive odds ratio is the ratio of the conditional probability of positive outcome when the conditioning variable is active to the conditional probability of positive outcome when the conditioning variable is inactive. Here: $\frac{\Pr(\text{male smokes} \mid \text{female smokes})}{\Pr(\text{male smokes} \mid \text{female does not smoke})} = \frac{0.574}{0.158} = 3.63$.

[8] This (small) difference in the descriptive odds ratio is consistent with the lower wage penalty for women smokers described above.

Mills ratio, which is introduced in the standard way into the regressions carried out using sample 2. In most of our regressions we use the balanced sub-sample (Sample 3) extracted from sample 2. As we need to observe one-period lagged smoking, this leaves us with roughly 10500 couples observed over eight periods. These couples are stable for all nine periods. Therefore, their match is likely to be better than that of couples who are excluded from sample 3. To correct for this selection bias, we compute a Mills ratio using a selection variable that equals 1 if the individual remains with the same partner over all of the 9 periods (this reflects selection from sample 1 into sample 3).

4 Econometric Modelling: A Dynamic Bivariate Probit Model

Let $Y_{i,t}$ be a binary indicator for smoking by individual *i* during period *t*, $X_{i,t}$ a vector of exogeneous individual and household covariates. The agent decides to smoke ($Y_{i,t}$=1) if the latent variable $Y^*_{i,t}$ is positive:

$$Y_{i,t} = \begin{cases} 1 & \text{if} \quad Y^*_{i,t} = \alpha Y_{i,t-1} + \beta Y_{-i,t-1} + \gamma X_{it} + c_i + \tilde{\varepsilon}_{it} > 0 \\ = 0 & \text{otherwise} \end{cases} \qquad \textbf{(1)}$$

where $Y_{-i,t-1}$ is the lagged smoking status of her/his partner (if any). If she/he does not live in a couple, $Y_{-i,t-1}$ is set to 0. The residual has two components: an individual specific effect c_i and a time-varying random shock $\tilde{e}_{i,t}$. A number of notes are in order. First, we would, in principle, prefer to estimate $Y_{i,t}$ as a function of $Y_{-i,t}$. However, such a model suffers from simultaneity bias. The variable $Y_{-i,t-1}$ in equation (1) above therefore acts as an instrument for $Y_{-i,t}$. Second, we do not control for cost-of-adjustment effects and past level of own consumption in this specification (Suranovic *et al.*, 1999). Third, in terms of the three arguments presented in section 2, the fixed effect c_i will capture matching, whereas β will measure the bargaining or information effects.

Now, consider the following bivariate probit specification (which simply develops **(1)**):

$$Y_{i,t} = \begin{cases} 1 & \text{if} \quad Y^*_{i,t} = a_1 Y_{i,t-1} + ß_1 Y_{-i,t-1} + ?_1 X_{i,t} + c_i + \tilde{e}_{i,t} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$Y_{-i,t} = \begin{cases} 1 & \text{if} \quad Y^*_{-i,t} = a_2 Y_{i,t-1} + ß_2 Y_{-i,t-1} + ?_2 X_{-i,t} + c_{-i} + \tilde{e}_{-i,t} > 0 \\ 0 & \text{otherwise} \end{cases} \qquad \textbf{(2)}$$

where ($Y_{i,t}$,$Y_{-i,t}$) describes the joint current smoking status of the couple, c_i and c_{-i} are individual specific fixed effects, and $(\tilde{\varepsilon}_{i,t}, \tilde{\varepsilon}_{-i,t})$ are bivariate normally distributed (with variances normalised to 1).

Index by *h,t* all row vectors whose two components are the variables describing the behaviours or characteristics of each spouse at time *t*, for instance $Y^*_{h,t} = (Y^*_{i,t}, Y^*_{-i,t})$. We are interested in the identification of the α's and γ's in the following model:

$$\begin{aligned}\Pr(Y_{h,t} \mid Y_{h,t-1}, X_{h,t}, c_h) = F_2[&(2Y_{i,t}-1)(\lambda_1 Y_{h,t-1} + \gamma_1 X_{i,t} + c_i, \\ &(2Y_{-i,t}-1)(\lambda_2 Y_{h,t-1} + \gamma_2 X_{-i,t} + c_{-i}), (2Y_{i,t}-1)(2Y_{-i,t}-1)\rho] \end{aligned} \quad \textbf{(3)}$$

where $\lambda_1=(\alpha_1, \beta_1)$ and $\lambda_2=(\alpha_2, \beta_2)$.

Assume that the vector of fixed effects has a conditional density $f(c_h|Y_{h,0},X_h)$, where X_h is the vector $(X_{h,1},X_{h,2},\ldots,X_{h,T})$. We can write the log-likelihood of observations $Y_{h,t}$, *t=1* to *T* as follows (Wooldridge, 2002*a*, 2002*b*[9]):

$$\Pr(Y_{h,1}, Y_{h,2}, \ldots, Y_{h,T} \mid Y_{h,0}, X_h) = \int_{\mathrm{supp}(c_h)} \left[\prod_{t=1}^{T} \Pr(Y_{h,t} \mid Y_{h,t-1}, X_{h,t-1}, c_h) \right] f(c_h \mid Y_{h,0}, X_h)\, dc_h \quad \textbf{(4)}$$

To estimate this model, we will assume that *f* is a bivariate discrete distribution. Hence c_i takes a finite number S_i of values c_{ij} on the real line. Accordingly, S_{-i} is the number of support points c_{-ik} for the marginal distribution of c_{-i}.

The conditional probability P_{kj} that $c_h=(c_i,c_{-i})$ is equal to $c_{jk}=(c_{ij}, c_{-ik})$ is modelled as a multinomial logit, with $Z_h=(Y_{h,0}, X_h)$ as regressors (Greene, 2002). Hence, there are vectors of coefficients $\delta_{11}, \delta_{12},\ldots, \delta_{S_i S_{-i}}$ such that:

$$\pi_{jk} = \Pr(c_h = c_{jk}) = \frac{\exp(\delta_{jk} Z)}{\sum_{j=1}^{S_i} \sum_{k=1}^{S_{-i}} \exp(\delta_{jk} Z)} \quad \textbf{(5)}$$

with δ_{11} normalised to 0. The final household likelihood is:

$$\Pr(Y_{h,1}, Y_{h,2}, \ldots, Y_{h,T} \mid Y_{h,0}, X) = \sum_{s=1}^{S} \pi_{jk} \left[\prod_{t=1}^{T} \Pr(Y_{h,t} \mid Y_{h,t-1}, X_{h,t-1}, c_{jk}) \right] \quad \textbf{(6)}$$

The parameters c_{jk}, α, β, δ and γ can then be identified from the data, by maximisation of the log-likelihood. The model is estimated by the Simulated Annealing EM algorithm (Celeux *et*

[9] This conditioning technique is proposed by Wooldridge to deal with the "initial conditions" problem that arises as a consequence of the correlation between $Y_{h,0}$ and the fixed effect in $\Pr(Y_{h,1}| Y_{h,0},X_{h,1},c_h)$.

al., 1995), which is a stochastic version of the EM algorithm of Dempster *et al.* (1977). The EM algorithm is increasing in lnL but is known to lead often to poor local maxima or saddle points. The SAEM algorithm was designed to overcome this limitation. Following other authors, we compare information criteria such as the BIC, and the AIC[10] for 2, 4 or more points c_{jk} in order to find out the optimal number S* of points of support (see for instance Deb and Trivedi, 1997 or Wedel *et al.*, 1993). An entropy-based measure is also used as an indicator for the satisfactory distribution of the individual heterogeneity over the support points[11]. The closer this is to 0, the more inaccurate is the classification of observations into distinct homogeneous groups (Jedidi *et al.*, 1997). Our model may be misspecified if two-period lagged decisions affect current smoking statuses. As we use a finite discrete distribution to model household fixed effects, there is no test of omitted variables: standard tests do not account for uncertainty on the optimal number of support points S, which may vary with the set of regressors. However, we will compute LM statistics for omission of $Y_{h,t-2}$, but they have to be considered cautiously. Last, in bivariate probits the use of distinct control variables in each equation (labour force status, age etc.) allows the robust identification of the correlation coefficient (Keane, 1992).

For the sake of comparison, we will also estimate a simple random-effect probit model of equation **(1)**, where the individual specific effects follow a Normal distribution. Further, the conditioning technique used to produce equation **(4)** is applied to avoid the initial conditions problem. We will also estimate bivariate probit models without individual or household fixed effects (c_h=0: cross-section regressions). These may be argued to be more descriptive than conclusive, as lagged consumption is likely to be endogenous in these regressions.

---

[10] AIC=LnL-p where LnL is the log-likelihood and p the number of parameters in the model. BIC=LnL-LnN*p/2 where N is the number of observations. There is no test of the optimal number of classes as no point of support can have no mass i.e. the conditions of compactness of the parameter space are not met.

[11] Entropy= $1-\frac{\sum_h \sum_j \sum_k -\varpi_{hjk}\ln(\varpi_{hjk})}{N*\ln(S^*)}$ where ϖ_{hjk} is the probability that c_h=c_{jk} given the information available in the dataset.

5 Matching or Bargaining?

5.1 Descriptive analysis

Table 4 presents the results of probit estimation of smoking status using Samples 1 and 2. It has three columns. The first refers to the simplest probit model estimated on sample 1. Current smoking is presumed to depend only on lagged smoking participation. Columns 2 and 3 show the results from a bivariate probit model without fixed effects, where partner's past smoking behaviour is added as an explanatory variable.

Table 4: Simple probit models.

Specification	**1: Probit**	**2: Bivariate Probit**	
Sample	Sample 1	Sample 2	
Equation	*Pooled*	*Male*	*Female*
Past participation: $Y_{i,t-1}$	3.181*** (0.019)	3.081*** (0.035)	3.331***(0.039)
Partner's past participation $Y_{-i,t-1}$	No	0.328*** (0.037)	0.271*** (0.041)
Age/10	-0.008 (0.037)	0.109 (0.096)	0.069 (0.086)
$Age^2/100$	-0.007* (0.004)	-0.018* (0.010)	-0.009 (0.009)
Sex	0.069*** (0.020)	No	No
Log (real income)	-0.054*** (0.010)	-0.077*** (0.025)	-0.083*** (0.020)
Has at least one child at home	0.037 (0.029)	0.002 (0.057)	0.084 (0.059)
Newborn child between *t-1* and *t*	-0.043 (0.063)	-0.026 (0.085)	0.047 (0.111)
Education ≥A-level	-0.151*** (0.021)	-0.148*** (0.038)	-0.122*** (0.044)
Married	-0.182*** (0.032)	Reference	Reference
Living together	-0.025 (0.042)	0.062 (0.060)	0.245*** (0.065)
Separated	0.188*** (0.070)	No	No
Never married	Reference	No	No
Other controls: Labour Force Status, Household size, Other marital statuses, Year, Region.			
Mills Ratio (selection into stable couples)	No	1.333*** (0.480)	2.518** (1.151)
Partner's Mills ratio	No	0.102 (0.094)	0.132 (0.130)
Constant	-1.067*** (0.121)	-2.324*** (0.493)	-3.070*** (0.772)
Rho	No	0.480*** (0.033)	
N	63530	19307	

Note: Standard errors in parentheses, adjusted for clustering at the household level. *=significant at the 10% level, **=significant at the 5% level, ***=significant at the 1% level. LnL_0 = log-likelihood for the constant-only model.

The main conclusion from this table, in terms of our paper's subject, is that in multivariate regressions individual smoking participation depends on partner's behaviour. We also uncover significant evidence that smoking and couple duration are correlated. The positive significant Mills ratio shows that those whose couple duration is largely determined by unobservables (for instance lifestyle choices) are more likely to smoke[12].

Other results in Table 4 (see also Table B1 in Appendix B) show no correlation between smoking and the age cohort. Smoking is more prevalent amongst the poorer, the less-educated, and lower amongst housewives. The income effects may be counteracted by “stress effects” for those with managerial responsibilities. Although household size has some effect on smoking, the presence of children as such does not affect behaviour.

However, these probit models suffer from severe misspecification bias, due to the omission of the fixed effect. To this extent, the results that we have presented up to this point may be argued to be more descriptive than conclusive.

5.2 Controlling for Fixed Effects

Table 6 below reports results from different specifications controlling for unobserved individual heterogeneity. From now on, we drop a number of variables that were insignificant in the previous estimates or do not vary greatly over time: this concerns most of the regional, occupation and household size dummies. The table has 8 columns. Columns 1 and 2 report benchmark estimates from a bivariate probit specification without fixed effects. Columns 3 and 4 estimate quit equations by gender. Using transitions from smoking to not smoking is a way to control for individual effects. Own past smoking status is replaced by own past consumption level in this specification, in order to control for adjustment costs. Columns 5 and 6 show estimated results from <u>individual</u> smoking equations (see equation **(1)**, where there is no spousal correlation in the residuals). We used for these estimates a random-effect probits (controlling for the initial condition problem as in equation **(4)**)[13].

The dynamic bivariate probit in columns 7 and 8 adds an individual fixed effect for both partners to the standard bivariate probit model. This is our preferred specification. The individual random effects here are distributed as a finite discrete distribution (see Section 4.1).

---

[12] Our instruments are insignificant in the marriage selection equation, although the power of the full set of variables is correct (see Table B2 in Appendix B). This problem of weak instrumentation disappears in the next section when we introduce individual fixed effects, dropping a number of Table B1’s explanatory variables.

[13] The usual information criteria (AIC or BIC) favour the random effect probit over the simple bivariate probit in columns 1 and 2.

Table 5 below illustrates our search for the optimal number of support points in the discrete distribution of individual heterogeneity. Information criterion with sample penalties (BIC) suggest four (2 x 2) as the “best” number of points, whereas AIC favours nine (3 x 3) support points. However, the entropy is higher with four support points, and with nine support points, two classes turn out to be only sparsely populated (with a mass of about 2%) and several class memberships are not well identified (the variance of the β_{jk} in equation **(5)** is very high). Hence, we retain four support points.

Table 5: Random-effect Bivariate Probit Model – selection of the number of mass points.

S_i	1		2			3	
S_{-i}	1	2	1	2	3	2	3
Log-likelihood	-2774	-2586	-2573	-2378	-2334	-2334	**-2271**
BIC	-2988	-2868	-2856	**-2823**	-2941	-2941	-3118
AIC	-2817	-2647	-2635	-2474	-2465	-2465	**-2454**
Entropy	100%	93.0%	92.1%	**94.9%**	90.2%	90.8%	80.7%

In Table 6’s specification 6, the variances of the coefficients are computed under the assumption that the optimal number of support points is not a random variable. A rigorous computation of the variance-covariance matrix should consider this issue. To our knowledge, this has not been treated in the existing literature, and following the tradition we use the standard information matrix.

Table 6: Dynamic equations with individual effects.

Specification	**3: Bivariate Probit**		**4: Quit Probits**		**5: Gaussian Random-Effect Probits**		**6: Discrete Random Effect Bivariate Probit**	
Sample	Sample 3							
Equation	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
Past participation: $Y_{i,t-1}$ (Past consumption in specification 4)	3.110*** (0.050)	3.526*** (0.058)	-0.034*** (0.006)	-0.045*** (0.008)	1.543*** (0.098)	1.438*** (0.132)	1.926*** (0.125)	2.070*** (0.181)
Partner's past participation: $Y_{-i,t-1}$ (Partner's past quit in specification 4)	0.307*** (0.053)	0.300*** (0.063)	-0.383 (0.316)	-0.010 (0.239)	-0.011 (0.149)	0.284* (0.154)	0.097 (0.422)	0.125 (0.536)
Mills Ratio	0.283** (0.135)	0.024 (0.150)	-0.546** (0.247)	0.013 (0.226)	-0.090 (0.273)	0.019 (0.310)	0.174 (0.220)	-0.209 (0.386)
Partner's Mills ratio	0.248* (0.132)	0.361** (0.180)	-0.118 (0.251)	-0.255 (0.290)	-0.157 (0.264)	0.488 (0.324)	0.100 (0.245)	0.280 (0.323)
Age/10	0.304** (0.143)	0.095 (0.147)	-0.114 (0.232)	-0.236 (0.216)	1.356 (1.561)	1.609 (2.111)	0.694*** (0.170)	0.159 (0.275)
$Age^2/100$	-0.038*** (0.014)	-0.014 (0.014)	0.007 (0.024)	0.007 (0.023)	-0.095* (0.054)	-0.150** (0.068)	-0.068*** (0.019)	-0.017 (0.032)
Log (real income)	-0.133*** (0.036)	-0.096*** (0.029)	0.013 (0.060)	0.075 (0.054)	-0.137** (0.069)	-0.074 (0.068)	-0.065 (0.044)	-0.058 (0.036)
Education ≥ A-level	-0.112** (0.051)	-0.105* (0.057)	0.208** (0.084)	0.020 (0.099)	-0.426 (0.294)	0.051 (0.336)	-0.140*** (0.052)	0.003 (0.072)
Has at least one child at home	0.009 (0.065)	0.143* (0.075)	-0.023 (0.103)	-0.320*** (0.123)	-0.005 (0.143)	0.004 (0.180)	-0.007 (0.075)	0.132 (0.116)
Newborn child between *t-1* and *t*	-0.026 (0.153)	-0.343 (0.216)	0.201 (0.193)	0.653*** (0.203)	-0.017 (0.176)	-0.567** (0.219)	-0.012 (0.372)	-0.538* (0.307)
Living together (ref: married)	0.067 (0.128)	0.174 (0.142)	-0.067 (0.175)	-0.492*** (0.173)	0.475 (0.306)	0.761* (0.411)	0.447** (0.183)	0.232 (0.207)
Mother/father at home	-0.679** (0.267)	-0.313*** (0.080)	0.507 (0.365)	0.212* (0.124)	-0.803 (0.543)	-0.597*** (0.151)	-0.649 (0.491)	-0.373*** (0.127)
Household size=5	0.090 (0.084)	0.089 (0.083)	-0.001 (0.132)	-0.358** (0.164)	-0.066 (0.149)	0.086 (0.203)	-0.056 (0.101)	0.147 (0.137)
Year = 1993 (ref: 1992)	0.090 (0.097)	-0.019 (0.118)	-0.170 (0.124)	0.126 (0.145)	-0.132 (0.191)	-0.248 (0.249)	-0.057 (0.141)	-0.200 (0.279)
Year = 1994 (ref: 1992)	0.183* (0.099)	-0.032 (0.115)	-0.232* (0.132)	0.075 (0.152)	-0.183 (0.323)	-0.503 (0.439)	-0.055 (0.209)	-0.392 (0.262)
Year = 1995 (ref: 1992)	0.254*** (0.097)	0.281*** (0.106)	-0.443*** (0.144)	-0.458** (0.183)	-0.171 (0.465)	-0.195 (0.634)	-0.029 (0.128)	-0.085 (0.253)

Year = 1996 (ref: 1992)	0.281*** (0.097)	0.132 (0.109)	-0.460*** (0.139)	-0.269* (0.157)	-0.174 (0.603)	-0.392 (0.826)	-0.001 (0.136)	-0.228 (0.244)
Year = 1997 (ref: 1992)	0.117 (0.097)	0.127 (0.108)	-0.264* (0.139)	-0.165 (0.160)	-0.447 (0.749)	-0.488 (1.026)	-0.198 (0.159)	-0.253 (0.260)
Year = 1998 (ref: 1992)	0.116 (0.099)	0.182* (0.105)	-0.222 (0.136)	-0.358** (0.173)	-0.460 (0.896)	-0.417 (1.228)	-0.208 (0.160)	-0.145 (0.331)
Year = 1999 (ref: 1992)	0.154 (0.098)	0.090 (0.112)	-0.310** (0.138)	-0.008 (0.156)	-0.490 (1.038)	-0.583 (1.423)	-0.178 (0.114)	-0.257 (0.197)
Constant	-1.918*** (0.525)	-1.962*** (0.496)	0.289 (0.805)	-0.118 (0.687)	-4.317*** (1.307)	-3.318** (1.406)	-1.701** (0.714)	-0.205 (1.064)
Constant ($\bar{?}_{1.2}$ =13.3%)	No	No	No	No	No	No	-1.701** (0.714)	-2.413** (0.986)
Constant ($\bar{?}_{2.1}$ =11.9%)	No	No	No	No	No	No	-3.665*** (0.709)	-0.205 (1.064)
Constant ($\bar{?}_{2.2}$ =65.8%)	No	No	No	No	No	No	-3.665*** (0.709)	-2.413** (0.986)
Individual effect variance	No	No	No	No	1.153***	1.511***	No	No
Rho / Gender correlation in time-varying random errors	0.521*** (0.046)		Independent regressions by sex		Independent regressions by sex		0.591*** (0.129)	
Gaussian random-effect: correlation with $\tilde{\varepsilon}_{i,t}$	No		No		0.571***	0.695***	No	
Controls for initial conditions: $Y_{h,1}$, X_{ht} (t=2,…,9)	No	No	No	No	Yes	Yes	Yes	
N	1321 households observed on 8 periods.		2290	2182	1321 households observed on 8 periods.			
Wald Chi-2 (#df)	8252 (40)		76 (20)	115 (20)	1304 (31)	802 (31)	_	
LM statistics for omission of $Y_{h,t-2}$	No		No	No	No	No	0.303 (5.99)	
Log Likelihood	-2749		-756	-531	-2434		-2378	

Note: standard errors in parentheses, adjusted for clustering on households. *=significant at the 10% level, **=significant at the 5% level, ***=significant at the 1% level,. In the quit regressions, observations corresponding to individuals who don't smoke at *t-1* have been dropped. The Z vector of conditioning variables was reduced to ($Y_{h,1}$,$X_{h,2}$) due to weak within-group variance of the X variables. The controls that were insignificant in the previous regressions are dropped (region, household size, other labour force statuses). Wave dummies control for price variations.

There are two striking results. First, as expected, in specifications **5** and **6** the coefficient on own lagged participation drops sharply, although it remains significant at all normal levels. The omitted individual fixed effect in specification **3** biases upwards the coefficient on lagged participation in the usual way.

Second, individual smoking participation is statistically independent of partner's smoking participation in a number of specifications. The strong effects of partner's lagged smoking in specification **3** (a bivariate probit <u>without</u> individual fixed effects) entirely disappear in specification **6** where individual effects are modelled in a flexible manner.

This result also pertains in simple probit equations (where there is no correlation between the error terms). Specification **4**'s quit estimates (which can be thought of as first-difference regressions) show no effect of partner's past quitting decision[14]. Last, specification 5 shows the results of simple probits with Gaussian random effects. Here, there is some correlation between women's smoking status and partner's lagged smoking, but only at the ten per cent level. Again, this specification is dominated by specification **6**'s bivariate random effects probits.

The absence of correlation between partners' behaviours reveals some deeper structure in the descriptive results presented above (see Table **4**). Partners' behaviours are indeed correlated in the raw data, but only because their associated fixed effects are not independent. In fact, the estimated individual fixed effects in columns 7 and 8 are positively correlated, with a correlation coefficient of 0.523. This is consistent with a matching model of marriage, but does not support a bargaining model of health behaviour determination within couples.

Three additional points can be highlighted in Table 6. The Mills ratio is insignificant in the fixed effect specifications, casting some doubt on Table 4's simpler results. Contrary to Prediction 3, the presence of children has no significant effect on smoking participation. There is however a "pregnancy" effect on women's smoking. The fact that this only appears for women is consistent with an asymmetric gender effect of parental smoking on child health. Last, we note that specification **6** is not rejected by an LM test for the omission of two-period lagged participation.

---

[14] It is actually difficult to identify the effect of partners' quits yesterday on quits today since, even in a large sample like the BHPS, we have very few observations on this phenomenon. Use of sample 2 does not yield different results.

6 Further results

The results in section 5 have allowed us to make progress in distinguishing between the matching and bargaining explanations of spousal correlation in smoking. This section proposes a test of the second and third predictions. We would like to know how agents react to past health changes, and whether being a parent reinforces the correlation in smoking statuses.

6.1 Health

Our second prediction is that smoking status may be positively or negatively correlated with partner's past health changes. This can be explained by both social learning about smoking's dangers (or more generally managing health capital) and, in a bargaining framework, partner's health developments affecting the anticipated value of time spent together.

We report results from probit quit equations and random effects probits on individual smoking status. These regressions include health development variables for both the individual and his or her partner between *t-2* and *t-1*, as measured by changes in subjective health status[15]. We denote all of *i*'s past health developments while smoking (both own and partner's) within a time period *j* (i.e. from *j-1* to *j*) by ΔH_{ij}. We explain quit decisions as a function of past health changes ΔH_{ij}, and current smoking status as a function of the sum of all past health changes: $\sum_{j=1}^{t-1} \Delta H_{ij}$. A positive correlation between these variables and smoking may be interpreted as learning about smoking's dangers.

Previous work using the same methods and dataset (Clark and Etilé, 2002) found some evidence that individual cigarette consumption reacts to own past health changes, but is largely independent of partner's health changes. Here we will reproduce this exercise, but with a far cruder binary measure of smoking. Table 7 shows the results.

[15] The subjective health status variable has five categories in the BHPS (excellent, good, fair, poor, very poor). We recoded excellent and good to "good", and fair, poor and very poor to "poor". At wave nine, the categories were somewhat different (excellent, very good, good, fair, poor), but the distribution of replies led us to keep the same grouping. We use this binary health variable (whereas Clark and Etilé, 2002, kept four categories) as our qualitative regressions do not allow efficient identification of a large number of health change dummies.

Table 7: Smoking Equations with Partner's Health Changes.

Specification	**7: Quit probit**		**8: Quit probit**		**9: Gaussian Random-Effect Probits**		**10: Discrete Random-Effect Bivariate Probit**	
Sample	Sample 2		Sample 3					
Equation	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
Past participation: $Y_{i,t-1}$ (Past consumption level in specification 7 and 8)	-0.028*** (0.004)	-0.049*** (0.006)	-0.032*** (0.007)	-0.045*** (0.009)	1.488*** (0.114)	1.288*** (0.153)	1.982*** (0.184)	2.031*** (0.323)
Partner's past participation: $Y_{-i,t-1}$ (Partner's past quit in specification 7 and 8)	0.065 (0.162)	-0.073 (0.184)	-0.471 (0.322)	-0.058 (0.248)	0.230 (0.152)	0.539*** (0.181)	-0.020 (0.914)	0.184 (0.706)
Health status stayed good between t-2 and t-1	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Health status changed from good to poor between t-2 and t-1	0.051 (0.095)	0.050 (0.102)	-0.103 (0.143)	0.032 (0.145)	-0.037 (0.095)	-0.070 (0.120)	-0.036 (0.084)	-0.009 (0.168)
Health status changed from poor to good between t-2 and t-1	0.021 (0.084)	-0.081 (0.096)	0.025 (0.121)	-0.090 (0.138)	-0.028 (0.069)	-0.005 (0.087)	0.004 (0.049)	0.039 (0.110)
Health status stayed poor between t-2 and t-1	0.136 (0.090)	0.108 (0.098)	0.122 (0.122)	0.116 (0.135)	-0.074 (0.062)	-0.084 (0.072)	-0.041 (0.043)	-0.066 (0.083)
Partner smoked at t-1, partner's health status stayed good between t-2 and t-1	-0.290*** (0.077)	-0.173** (0.083)	-0.262** (0.112)	-0.088 (0.120)	0.032 (0.046)	-0.039 (0.051)	0.074 (0.098)	-0.018 (0.095)
Partner smoked at t-1, partner's health status changed from good to poor between t-2 and t-1	-0.049 (0.176)	0.069 (0.198)	-0.035 (0.262)	-0.350 (0.301)	-0.157 (0.161)	0.362 (0.226)	-0.242 (0.307)	0.276 (0.331)
Partner smoked at t-1, partner's health status changed from poor to good between t-2 and t-1	0.016 (0.144)	0.265 (0.196)	0.349 (0.232)	0.325 (0.357)	-0.145 (0.111)	-0.025 (0.163)	-0.063 (0.223)	-0.027 (0.443)
Partner smoked at t-1, partner's health status stayed poor between t-2 and t-1	-0.143 (0.160)	0.180 (0.194)	0.085 (0.229)	0.163 (0.239)	0.074 (0.112)	-0.064 (0.118)	0.053 (0.222)	-0.135 (0.318)
Partner's health status stayed good between t-2 and t-1	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Partner's health status changed from good to poor between t-2 and t-1	-0.222 (0.134)	-0.187 (0.158)	-0.288 (0.207)	0.018 (0.201)	0.019 (0.114)	-0.283 (0.175)	0.050 (0.327)	-0.072 (0.320)
Partner's health status changed from poor to good between t-2 and t-1	-0.225* (0.117)	-0.452*** (0.169)	-0.516*** (0.190)	-0.519 (0.322)	0.100 (0.082)	0.117 (0.139)	0.081 (0.251)	0.073 (0.426)
Partner's health status stayed poor between t-2 and t-1	-0.069 (0.128)	-0.169 (0.166)	-0.115 (0.198)	0.0001 (0.209)	0.039 (0.084)	0.006 (0.090)	-0.037 (0.221)	0.026 (0.289)
Other controls: same as in Table 6's regressions								
Rho / Gender correlation in time-varying random errors	Independent regressions by sex							
Controls for initial conditions	Irrelevant				Yes			
N	4045	3782	2290	2182	1321 households observed on 7 periods			

Note: Standard errors in parentheses, adjusted for clustering on households. *=significant at the 10% level, **=significant at the 5% level, ***=significant at the 1% level.

We find no correlation between own past health developments and current smoking status (controlling for own and partner's past participation). In the quit equations, continued good health for a partner who smokes is associated with a lower probability of quitting. However, better health for a non-smoking partner is also negatively correlated with quitting. It is difficult to tell a consistent story here, and most of the estimated coefficients are insignificant. These regressions reveal little evidence of either social learning or of private learning.

This latter contrasts with the findings in Clark and Etilé (2002), where one's own health changes did matter. There are at least two potential reasons for this: Clark and Etilé consider four categories of health developments, interacted with age group and sex; and, perhaps most importantly, they model the level of daily cigarette consumption, not the participation decision in its own right. With respect to the last point, convex adjustment costs will render the level of consumption more malleable than the decision to smoke itself.

6.2 Children

Our prediction 3 was that the presence of children would increase the correlation between smoking behaviours, by emphasising the public good value of time spent with one's partner in good health. The (current or past) presence of children may also act as an indicator of the quality of the match, which again implies a correlation of preferences[16].

Table 8 below interacts lagged smoking statuses with dummies for the presence of children, in specifications **4**, **5** and **6** of table 6. The results do not support our third prediction. The estimated correlations with partner's lagged smoking status are lower for couples with children in specification **12**, but for women only, and our interaction variables remain insignificant in specification **13**'s estimates.

[16] Using the marital and fertility history information in Wave 2 (including the number of adopted, step- and natural children), and information about the number of children at home, we know whether 98% of the individuals in sample 2 have already had a child at home (natural or "adopted"/step children are treated equally).

Table 8: Spousal correlation & children

Specification	**11: Quit probits**		**12: Gaussian Random-Effect Probits**		**13: Discrete Random Effect Bivariate Probit**	
Sample	Sample 3					
Equation	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
Past participation & a child	-0.033*** (0.006)	-0.045*** (0.008)	1.557*** (0.103)	1.521*** (0.145)	1.959*** (0.146)	2.142*** (0.195)
Past participation & no child	-0.028*** (0.009)	-0.058*** (0.017)	1.298*** (0.234)	1.030*** (0.257)	1.575*** (0.321)	1.853*** (0.257)
Partner's past participation & partner has a child	-0.104 (0.089)	0.015 (0.096)	-0.006 (0.161)	0.244 (0.180)	0.041 (0.463)	0.051 (0.623)
Partner's past participation & partner doesn't have child	-0.254 (0.206)	0.220 (0.242)	0.091 (0.342)	0.660* (0.355)	0.287 (0.479)	0.205 (0.556)

Note: Standard errors in parentheses, adjusted for clustering on households. *=significant at the 10% level, **=significant at the 5% level, ***=significant at the 1% level.

6.3 The matching process

One may argue that within-household interactions concerning lifestyle choices take place mainly during the matching process or during the first years of the couple. Hence, the lifestyle choices of stable couples should already be well matched when we observe them in Table 6's regressions. To test this argument, we estimate specifications **4** and **5** on sample 2 with partner's past participation interacted with an indicator for the duration of the cohabitation spell (at least three years vs. less than three years)[17].

Table 9: Old Couples & New Couples

Specification	**14: Quit probits**		**15: Gaussian Random-Effect Probits**	
Sample	*Sample 2*			
Equation	*Male*	*Female*	*Male*	*Female*
Past level of consumption	-0.031*** (0.004)	-0.044*** (0.005)	No	No
Past participation	No	No	1.483*** (0.087)	1.427*** (0.103)
Partner's past participation & cohabitation less than three years.	-0.183** (0.081)	-0.039 (0.087)	0.208 (0.171)	0.301 (0.187)
Partner's past participation & cohabitation more than three years.	-0.211*** (0.060)	-0.070 (0.064)	0.087 (0.135)	0.299** (0.138)
N	5115	4760	13908	13908

There is no difference in the estimated interaction effect between old and new couples in the quit equation. The estimates in specifications **14** and **15** do not provide any significant evidence of stronger partner influence in new couples, although the estimated coefficients on

[17] We cannot estimate discrete bivariate fixed effect probits for the short duration couples as identification of this model requires transitions between smoking statuses.

partner's past participation are higher for new couples. Although we find that male quitting is lower when the female smokes, comparing the results of Table 6 and Table 9 is hazardous, since the estimation samples are different. This emphasises the need for a structural model of the marriage market accounting for the various selection biases, wherein smoking may be considered by spouses and potential partners as a signal for lifestyle preferences. This appealing, although complicated, subject is left for future research.

7 Conclusion

This paper has used nine waves of BHPS data to examine intra-spousal correlations in smoking behaviour, which can result from matching, bargaining, public goods, or social learning. We first note that there is indeed a correlation in smoking status in the raw data, although there are some interesting differences by sex.

Perhaps this paper’s most important contribution has been to the general question of modelling correlations between spouses’ behaviours. We show that both probit and bivariate probit equations, without controls for individual heterogeneity, reveal a positive correlation between partners’ smoking participation: this is consistent with both matching and bargaining. Controlling for fixed effects allows us to distinguish between opposing interpretations. In our preferred specification, a bivariate probit with random effects, partners’ behaviours are statistically independent: all of the correlation in smoking status works through the correlation in individual fixed effects. Further, we find very little evidence to support social learning in terms of smoking status. As such, we believe that the correlation in the raw data reflects matching on the marriage market, rather than bargaining or learning within the couple.

Our results allow us to contribute directly to the public policy debate. Given the matching of partners’ preferences for smoking, but only weak evidence of spillovers in cigarette consumption between partners during marriage, it seems essential to target both partners in order to reduce household smoking. Interventions targeting only the female partner (for instance during pregnancy) would not appear to be effective in reducing male smoking.

Appendix A: Descriptive Statistics

Sample 1 (N = 63530)	
<u>DEPENDENT VARIABLE</u>	
Smokes at *t*	27.4%
<u>LAGGED VARIABLES</u>	
Smokes at *t-1*	27.9%
Smokes at *t-2*	27.2% (N=51182)
Quit between *t-1* and *t* (full sample / sub-sample of smokers)	2.3%
Partner quits between *t-2* and *t-1* (full sample / sub-sample of smokers)	2.4% (N=31433)
Partner smokes at *t-1*	26.6% (N=39506)
Partner smokes at *t-2*	25.9% (N=31433)
<u>SOCIO-DEMOGRAPHIC VARIABLES</u>	
Age	45.6 (18.3)
Male	46.7%
Log yearly real income	8.716 (1.098)
Household Size = 2: ***reference***	33.9%
Household Size = 3	19.6%
Household Size = 4	19.8%
Household Size = 5	7.9%
Household Size = 6	2.7%
Has at least one child	29.7%
New child between *t-1* and *t*	3.4%
Married: ***reference***	56.6%
Living together	8.2%
Widowed	8.9%
Divorced or separated	5.7%
Separated	1.7%
Never married	19.9%
Manager (permanent)	10.3%
Supervisor (permanent)	8.3%
No responsibilities (permanent)	25.0%
No responsibilities (temporary): ***reference***	7.2%
Self-employed	3.9%
Unemployed	20.1%
Retired	9.3%
Mother-at-home	4.2%
School or training	3.9%
Labour force status: not defined	10.3%
Education ≥ A-level	63.0%
Education < A-level: ***reference***	37.0%
<u>HEALTH VARIABLES</u>	
Health status is good	67.2%
Partner's health status is good (if has a partner)	68.7% (N=39544)

Appendix B – Additional results

Note: in all regressions, standard errors in parentheses, adjusted for clustering on households. *=significant at the 10% level, **=significant at the 5% level, ***=significant at the 1% level, N.S.=insignificant at the 10% level. LnL_0 = log-likelihood for the constant-only model. Other household size and marital status dummies were not significant.

Table B1: Simple probit models.

Specification	1: Probit	2: Bivariate Probit	
Sample	Sample 1	Sample 2	
Equation	*Pooled*	*Male*	*Female*
Past participation: $Y_{i,t-1}$	3.181*** (0.019)	3.081*** (0.035)	3.331*** (0.039)
Partner's past participation $Y_{-i,t-1}$	No	0.328*** (0.037)	0.271*** (0.041)
Age/10	-0.008 (0.037)	0.109 (0.096)	0.069 (0.086)
$Age^2/100$	-0.007* (0.004)	-0.018* (0.010)	-0.009 (0.009)
Sex	0.069*** (0.020)	No	No
Log (real income)	-0.054*** (0.010)	-0.077*** (0.025)	-0.083*** (0.020)
Has at least one child at home	0.037 (0.029)	0.002 (0.057)	0.084 (0.059)
Newborn child between t-1 and t	-0.043 (0.063)	-0.026 (0.085)	0.047 (0.111)
Education ≥A-level	-0.151*** (0.021)	-0.148*** (0.038)	-0.122*** (0.044)
Employed & manager	-0.146* (0.054)	0.392* (0.226)	0.336*** (0.125)
Employed & supervisor	-0.096* (0.055)	0.314* (0.175)	0.205 (0.136)
Employed & no responsibilities & permanent job.	-0.100** (0.047)	0.133 (0.135)	0.228* (0.126)
Self-employed	-0.098* (0.056)	0.430* (0.224)	0.580** (0.258)
Unemployed	0.030 (0.065)	0.053 (0.120)	-1.080** (0.526)
Retired	-0.152*** (0.057)	0.226 (0.160)	-1.048** (0.447)
Mother/father at home	-0.148*** (0.055)	-0.291 (0.192)	0.058 (0.158)
At school or enrolled in government training	-0.148** (0.070)	-1.116** (0.527)	-2.716** (1.347)
Labour Force Status: not defined	0.069 (0.061)	0.297* (0.160)	-0.336 (0.220)
Married	-0.182*** (0.032)	Reference	Reference
Living together	-0.025 (0.042)	0.062 (0.060)	0.245*** (0.065)
Separated	0.188*** (0.070)	No	No
Never married	Reference	No	No
Household size=3	0.019 (0.028)	0.097* (0.052)	0.087 (0.057)
Household size=5	0.087** (0.042)	0.175** (0.079)	0.195** (0.078)

Year = 1992	Reference	Reference	Reference
Year = 1993	0.027 (0.039)	0.041 (0.069)	0.082 (0.076)
Year = 1994	0.100** (0.039)	0.114* (0.069)	0.091 (0.074)
Year = 1995	0.196*** (0.038)	0.168** (0.069)	0.301*** (0.071)
Year = 1996	0.108*** (0.038)	0.188*** (0.067)	0.122* (0.072)
Year = 1997	0.113*** (0.038)	0.080 (0.066)	0.116 (0.074)
Year = 1998	0.106*** (0.037)	0.109* (0.065)	0.174** (0.070)
Year = 1999	0.159*** (0.038)	0.189*** (0.067)	0.149** (0.074)
Controls region	Yes	Yes	Yes .
Mills Ratio (selection into stable couples)	No	1.333*** (0.480)	2.518** (1.151)
Partner's Mills ratio	No	0.102 (0.094)	0.132 (0.130)
Constant	-1.067*** (0.121)	-2.324*** (0.493)	-3.070*** (0.772)
Rho	No	0.480*** (0.033)	
N	63530	19307	
Wald Chi-2 (#df)	29841 (50)	16669 (96)	
LnL (LnL_0)	-11290 (-37320)	-5852 (-21707)	

Table B2: Instrumental regression for selection bias.

Selection	**Sample 1 into sample 2**		**Sample 1 into sample 3**	
Equation	*Male*	*Female*	*Male*	*Female*
Unemployment rate	0.015 (0.013)	-0.009 (0.015)	0.008 (0.013)	-0.015 (0.014)
Partner unemployment rate	-0.021 (0.016)	0.011 (0.012)	-0.005 (0.015)	0.011 (0.012)
Manager (permanent)	0.781*** (0.047)	0.091** (0.045)	0.638*** (0.048)	0.054 (0.044)
Supervisor (permanent)	0.528*** (0.048)	0.143*** (0.044)	0.407*** (0.049)	0.029 (0.043)
No responsibilities (permanent)	0.346*** (0.043)	0.143*** (0.038)	0.337*** (0.045)	0.086** (0.038)
Self-employed	0.791*** (0.048)	0.389*** (0.054)	0.562*** (0.048)	0.174*** (0.051)
Unemployed	-0.037 (0.051)	-0.673*** (0.058)	-0.090* (0.054)	-0.729*** (0.063)
Retired	0.414*** (0.045)	-0.596*** (0.039)	0.418*** (0.047)	-0.412*** (0.039)
Mother-at-home	0.239** (0.116)	0.202*** (0.040)	0.107 (0.118)	0.015 (0.039)
School or training	-1.373*** (0.066)	-1.588*** (0.058)	-1.320*** (0.083)	-1.458*** (0.067)
Labour force status: not defined	0.401*** (0.054)	-0.256*** (0.053)	0.325*** (0.055)	-0.219*** (0.054)
Education = A-level	-0.124*** (0.021)	0.011 (0.018)	-0.094*** (0.020)	0.024 (0.018)
Education>A – level	0.048** (0.021)	0.068*** (0.020)	0.080*** (0.019)	0.129*** (0.020)
Controls for years	Yes	Yes	Yes	Yes

and regions				
N	29692	33838	29692	33838
LnL (LnL_0)	-17261 (-18833)	-20926 (-23011)	-19249 (-20333)	-21282 (-22547)

Table B3: Coefficients of conditioning variables** ($Y_{h,1}$,$X_{h,2}$) **controlling for initial conditions.

Specification	**5: Dynamic Random-effect probit regressions by gender**		**6: Dynamic bivariate random-effect probit**					
Point of support n°	Irrelevant		1,2		2,1		2,2	
Equation	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
$Y_{.O}$	3.006*** (0.253)	4.557*** (0.487)	1.052 (0.860)	-6.523*** (0.983)	-5.155*** (0.749)	0.523 (0.926)	-4.329*** (0.756)	-5.394*** (0.817)
Partner's Y_0	0.424** (0.177)	0.061 (0.201)	No		No		No	
Newborn child between t-1 and t	-0.284 (0.302)	0.158 (0.277)	0.726 (1.057)		0.999 (1.662)		1.237 (1.335)	
(Male) Mills ratio	1.040*** (0.368)	-0.248 (0.602)	-0.055 (1.468)		-0.054 (1.724)		-2.128 (1.647)	
Partner’s or female Mills ratio	0.618* (0.370)	0.221 (0.381)	1.890 (1.401)		-0.193 (1.434)		0.514 (1.406)	
Has at least one child at home	-0.008 (0.174)	0.410* (0.217)	0.013 (0.749)		0.622 (0.681)		-0.212 (0.664)	
Initial income	-0.035 (0.099)	-0.154** (0.078)	-0.435 (0.521)	0.719 (0.356)	-0.366 (0.479)	0.432 (0.368)	0.124 (0.494)	0.775** (0.337)
Initial age/10	-0.624 (1.579)	-1.388 (2.115)	0.580 (3.476)	-1.338 (3.257)	2.529 (3.547)	-2.767 (3.456)	2.018 (3.290)	-2.296 (3.263)
Initial age^2/100	0.013 (0.063)	0.139* (0.080)	-0.117 (0.354)	0.185 (0.358)	-0.216 (0.360)	0.294 (0.364)	-0.160 (0.327)	0.230 (0.347)
Initial education ≥ A-level	0.238 (0.302)	-0.195 (0.356)	0.208 (0.682)	0.367 (0.707)	0.045 (0.662)	0.561 (0.699)	0.235 (0.610)	0.495 (0.661)
Initial marital status: live together	-0.475 (0.319)	-0.153 (0.392)	-1.859 (1.225)		0.067 (1.345)		-0.021 (1.002)	
Initial household size=5	0.354* (0.214)	0.119 (0.256)	-1.004 (0.854)		-1.334* (0.793)		-1.143 (0.826)	
Constant	No	No	0.670 (6.684)		0.758 (6.216)		-0.571 (6.276)	

Note: In specification 6, point number 2 (cf. Table 6 above) is the reference for the estimation of the multinomial logit probabilities conditional on ($Y_{h,1}$,$X_{h,2}$) (see equation **(5)**).

References

Blank, R.M. (2002), “Can equity and efficiency complement each other?”. *Labour Economics*, **9**, 451-468.

Bolin, K., Jacobson, L. and Lindgren, B. (2001), “The family as the health producer – when spouses are Nash-bargainers”, *Journal of Health Economics*, 20, 349-362.

Bolin, K., Jacobson, L. and Lindgren, B. (2002), “The family as the health producer – when spouses act strategically”, *Journal of Health Economics*, 21, 475-495.

Celeux, G., Chauveau, D. and Diebolt, J. (1995), “On Stochastic Versions of the EM algorithm”, Working Paper n°2514, Paris: Institut National de Recherche en Informatique et en Automatique, 1-22.

Chiappori, P.A. (1992). “Rational Household Labour Supply”. *Econometrica*, **56**, 63-89.

Clark, A.E., and Etilé, F. (2002). “Do Health Changes Affect Smoking? Evidence from British Panel Data”. *Journal of Health Economics*, **21**, 533-562.

Currie, J. and Madrian, B. (1999), “Health, Health Insurance and the Labour Market”, *Handbook of Labor Economics* (Eds: Ashenfelter, O. and Card, D.). Amsterdam: Elsevier.

Contoyannis P, and Jones AM. (2001), “Socio-economic status, health and lifestyle”, working paper York Seminar in Health Econometrics, www1.york.ac.uk/res/herc/papers.html.

Dempster, A.P., Laird N.M. and Rubin, P.B. (1977), “Maximum-Likelihood from incomplete data via the E-M algorithm”, *Journal of the Royal Statistical Society B*, 39, 1-38.

Deb, P. and Trivedi P.K. (1997), “Demand for Medical Care by the Elderly: A Finite Mixture Approach”, Journal of Applied Econometrics, 12, 313-336.

Farrell, L. and Shields, M. (2002). "Investigating the economic and demographic determinants of sporting participation in England". *Journal of the Royal Statistical Society A*, **165**, 335-348.

Jedidi, K., Jagpal, H.S. and DeSarbo, W.S. (1997), “Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity”, *Marketing Science*, 16(1), 39-59.

Keane, M. (1992), “A Note on Identification in the Multinomial Probit Model”, *Journal of Business and Economic Statistics*, 10(2), 193-200.

Latkin, C., Mandell, W., Oziemkowska, M., Celentano, D., Vlahov, D., Ensminger, M., & Knowlton, A. (1995). "Using social network analysis to study patterns of drug use among urban drug users at high risk for HIV/AIDS". *Drug and Alcohol Dependence*, **38**, 1-9.

Leonard, K., & Mudar, P. (2003). "Peer and Partner Drinking and the Transition to Marriage: A Longitudinal Examination of Selection and Influence Processes". *Psychology of Addictive Behaviors,* **17**, 115-125.

Levine, P.B., Gustafson, T.A., and Velenchik, A.D. (1997), "More Bad News for Smokers? The Effects of Cigarette Smoking on Wages". *Industrial and Labor Relations Review*, **50**, 493-509.

Suranovic, S.M., R.S. Goldfarb and T.C. Leonard (1999), "An Economic Theory of Cigarette Addiction", *Journal of Health Economics*, 18, 1-29.

Van Ours, J. (2002), "A Pint a Day Raises a Man's Pay; But Smoking Blows that Gain Away", IZA Discussion Paper No.473.

Vella, F. (1993), "A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors", *International Economic Review*, 34 (2), 441-457.

Wedel, M. , DeSarbo, W.S., Bult, J.R. and Ramaswamy, V. (1993), "A Latent Class Poisson Regression Model for Heterogeneous Count Data", *Journal of Applied Econometrics*, 8, 397-411.

Wilson, S.E. (2001), "The Health Capital of Families: An Investigation of the Inter-Spousal Correlation in Health Status", Brigham Young University, mimeo.

Wooldridge, J.M. (2002a), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: The MIT Press.

Wooldridge, J.M. (2002b), "Simple solutions to the initial conditions problem in dynamic non-linear panel data models with unobserved heterogeneity", IFS/UCL CEMFI Working Paper 18/02, 1-44.