TI 2015-093/V
Tinbergen Institute Discussion Paper

Smoking Initiation: Peers and Personality

Chih-Sheng Hsieh[1]
Hans van Kippersluis[2]

[1] The Chinese University of Hong Kong, Hong Kong, PR China;
[2] Erasmus School of Economics, Erasmus University Rotterdam, and Tinbergen Institute, the Netherlands.

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Smoking Initiation: Peers and Personality*

Chih-Sheng Hsieh[1] and Hans van Kippersluis[2,3]

[1]Department of Economics, Chinese University of Hong Kong, Hong Kong
[2]Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands
[3]Tinbergen Institute, The Netherlands

August 3, 2015

Abstract

Social interactions are generally thought to play an important role in smoking initiation among adolescents. In this paper we exploit detailed friendship nominations in the US Add Health data, and extend the Spatial Autoregressive Model (SAR) model to deal with (i) endogenous peer selection, and (ii) unobserved contextual effects, in order to identify the endogenous peer effect. We show that peer effects in the uptake of smoking are predominantly affecting individuals who are emotionally unstable. That is, individuals with 'weaker' personalities are more vulnerable to peer pressure. This finding not only helps understanding heterogeneity in peer effects, but additionally provides a promising mechanism through which personality affects later life health and socioeconomic outcomes.

Keywords: Smoking, Peer effects, Personality, SAR model, Bayesian MCMC
JEL Codes : C11, C21, I12

*Van Kippersluis thanks the Netherlands Organization of Scientific Research (NWO Veni grant 016.145.082) and the National Institute on Aging of the National Institutes of Health (R01AG037398) for financial support.The authors are grateful to seminar participants at Erasmus University Rotterdam and the National Graduate Institute for Policy Studies (GRIPS) Tokyo for helpful comments.

1 Introduction

While smoking rates have fallen over past decades, recently this trend has stalled (DHHS, 2012), and smoking continues to be the leading preventable cause of death, killing nearly 6 million people each year (Mokdad et al., 2004; Danaei et al., 2009; OECD, 2013). Reliably identifying the causal factors underlying youth smoking initiation is vital to develop effective smoking prevention programs (Heckman et al., 2008). The economics literature has traditionally focused on price, taxation, and addiction as determinants of smoking (Chaloupka and Warner, 2000; DeCicca et al., 2002), yet in recent years considerably more attention is paid to social interactions in smoking and other unhealthy behaviors (DeCicca et al., 2008; Cawley and Ruhm, 2011). This is not surprising as social interactions and peer effects are not just often-cited determinants of smoking initiation, but – when present – additionally capable of generating social multiplier effects of policy interventions (Cutler and Glaeser, 2010; Fletcher, 2010; Cawley and Ruhm, 2011).

This paper is – to the best of our knowledge – the first to investigate whether peer influences are moderated by personality. In particular, we intend to answer the question: Are individuals with stronger personalities less vulnerable to peer pressure in the uptake of smoking? The paper contributes to two distinct lines of thriving literatures.

First, we contribute to the literature on the effects of personality on health behavior and health. It is strongly established that personality traits such as conscientiousness and emotional stability are linked to healthy behavior and health (Borghans et al., 2008; Almlund et al., 2011). In fact, improving personality traits is one of the key mechanisms through which early-childhood interventions have long-lasting effects on life outcomes (Heckman et al., 2013; Campbell et al., 2014). Nonetheless, the reason for the relationship between personality and health is poorly understood (Almlund et al., 2011; Young and Beaujean, 2011). Here, we investigate whether adolescents who are more conscientious and emotionally stable are less susceptible to peer influences, and better able to resist pressure from bad role models. If true, this could provide an important mechanism through which personality affects later life health.

Second, we make two contributions to the literature on the identification and interpretation of peer effects. While the importance of peer effects in smoking is now widely recognized (Chaloupka and Warner, 2000; Heckman et al., 2008; Cawley and Ruhm, 2011), implicitly homogenous effects are typically assumed (see section 2.1 for a review). This implies that we know strikingly little about which adolescents are most likely to join in versus avoid the deviant behavior that is present to some degree in almost all adolescent peer groups (Allen et al., 2012). Our first contribution to the peer effects literature is to improve understanding of heterogeneity in social interactions with respect to personality. The second contribution, which we explain in more detail below, is to introduce a Spatial Autoregressive Model (SAR) model that can simultaneously deal with (i) endogenous selection of friends, and (ii) unobserved contextual effects, in the identification of endogenous peer effects. This methodological extension overcomes the problem of disentangling the endogenous peer effect from unobserved contextual effects (see e.g., Fruehwirth, 2014), while at the same time addressing the problem of endogenous friendship formation (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2015).

Identifying peer effects is notoriously plagued with two major identification problems (Manski, 1993; Moffitt, 2000; Graham, 2015): (i) the reflection problem, and (ii) correlated effects. The reflection problem arises because the peers' observed outcome is the result of the peers' background (Bramoullé et al., 2009; Sacerdote, 2011), and hence it is difficult to distinguish between endogenous effects (the individual's behavior is directly affected by peers' behavior) and contextual effects (the individual's behavior is affected by the characteristics of his/her peers). The second problem, correlated effects, is due to selection (e.g., parents choose schools for their children; students select friends on basis of same gender, race, etc.) or due to sharing common environments (e.g., same teachers). Hence, it is difficult to separate peer effects from spurious correlations in behavior due to common characteristics and environments.

While the use of randomization in identifying peer effects is gaining popularity (e.g., Sacerdote, 2001; Zimmerman, 2003; Eisenberg et al., 2014), and has recently been vociferously advocated (Angrist, 2014), randomization has two fundamental limitations specific to the peer effects literature. First, while randomization is the ideal approach to tackle correlated effects, it does not solve the reflection problem. Indeed, Sacerdote (2001) and Carrell et al. (2013) used randomly assigned roommates in colleges, yet could not distinguish between endogenous and contextual peer effects. Second, it is extremely difficult, if not impossible, to exogenously manipulate an individual's peer group. After all, if you are randomly assigned a roommate in college that you do not like, then you are unlikely to spend time with the roommate, and the peer effects in such settings may be very different from the peer effect in naturally occurring settings (Card and Giuliano, 2013). Indeed, Carrell et al. (2013) report that randomly assigned Air Force Academy students segregated into homogeneous subgroups, which illustrates the sheer difficulty of randomly manipulating peer groups.

In contrast to randomization, the SAR model is able to tackle the reflection problem by exploiting information of friendship networks to separate endogenous effects and contextual effects (Bramoullé et al., 2009; Lee et al., 2010; Lin, 2010). The intuition is that since peer groups are not completely overlapping, one can use the characteristics of the non-overlapping friends of your friends as instrumental variables for the outcome of your friends. This approach however has two main limitations. First, relying on friendship nominations aggravates the problem of correlated effects. After all, a selected group of nominated friends is highly likely to share common characteristics and environments. To account for similar environments, we include peer group fixed effects. To account for individual correlated effects, we explicitly model the friendship formation using observed and unobserved (latent) factors influencing both the selection of friends and the outcome variable following Hsieh and Lee (2015).[1] Second, while the SAR model is technically able to separate endogenous peer effects from contextual effects, the endogenous peer effect will still be biased in case of unobserved contextual effects (e.g., Fruehwirth, 2014). Our methodological contribution is to account for unobserved (latent) contextual effects. We use placebo tests to gauge the potential of the approach, and present evidence that our selection-corrected SAR (SC-SAR) model with latent contextual effects is able to deal with some of the most notorious and persistent problems in identifying the endogenous peer effect.

Our SC-SAR estimates are based upon the Add-Health data, which has three main advantages. First, Add-Health provides detailed friendship nominations that enable not just solving the reflection problem, but additionally identifying the most relevant peer group. Second, the data contains personality measures for conscientiousness and emotional stability, both of which have been linked to health behaviors (Hampson et al., 2007, 2010), and allow to establishing heterogeneity in peer effects with respect to personality. Third, the Add-Health data interviews high-school students in grades 7-12 (i.e. between age 12 and 18). Since more than 80% of adult smokers begin smoking by 18 years of age (DHHS, 2012), the age span of the Add-Health data is the most relevant one in terms of smoking prevention efforts.

Our results provide strong evidence that peer effects in smoking are moderated by personality. Individuals with 'weaker' personalities in terms of emotional stability have larger peer effects compared to 'stronger' personalities. The social interactions between individuals who are both emotionally unstable are particularly vulnerable to the adoption of unhealthy habits. While it seems extremely difficult to manipulate the composition of peer groups on basis of personality, the results do suggest that interventions aimed at groups of emotionally unstable individuals have the largest scope in reducing the uptake of smoking and other unhealthy behaviors in adolescence.

The findings are also suggestive of an important mechanism through which personality affects later life outcomes. We find that emotional stability, which is associated with the skills of self-control and resisting temptation from peers (Costa and McCrae, 1992), is important to

[1] An alternative is to use the whole classroom as the relevant peer group. This could take away worries about endogenous network formation, and the variation in group sizes can still identify peer effects (Lee, 2007; Boucher et al., 2014). The drawback is however that not all classmates are one's peers, and therefore we prefer to focus on friendship nominations.

defy smoking initiation in social interactions among adolescents. Since we find similar patterns for the prevalence of getting drunk, it seems plausible that the skills of resisting temptations and standing up against group pressure are productive more generally in maintaining a healthy lifestyle and perhaps even becoming socioeconomically successful. Our results therefore provide a promising mechanism in the strong association between personality characteristics and later-life outcomes that is so far poorly understood.

This paper is organized as follows. In section 2 we discuss the literature on peer effects in smoking, and the literature on the relationship between personality and smoking. Section 3 discusses the data, and 4 presents the empirical model used to identify peer effects. In section 5 we discuss our results, after which we present robustness checks in section 6. Section 7 summarizes and discusses the implications of the results.

2 Related literature

In a comprehensive review of the social science literature, Conrad et al. (1992) report that the most important predictors of smoking initiation are socioeconomic background, social bonding variables, peer effects, and a range of non-cognitive skills. In this section we focus on the latter two, and discuss the literature on peer effects in smoking (section 2.1)[2] and the literature on the relationship between personality (non-cognitive skills) and smoking (section 2.2).

2.1 Peer effects in smoking

Glaeser and Scheinkman (2003) and Cutler and Glaeser (2010) describe various mechanisms that could produce peer effects in smoking. First, peer effects could include what they term "learning", which may have both positive and negative consequences. When your peers smoke, information becomes available about the benefits and costs of smoking and you may act on this. Second, they discuss stigma. When many peers around you smoke, this tends to reduce the negative social stigma that is normally associated with smoking. Third, there may be taste-related interactions, due to a desire for conformity and imitation. In simple terms, it is more pleasurable to do something together. Finally, Cutler and Glaeser (2010) note that the supply side plays a role, e.g., healthy alternatives to cigarettes (e.g., fruit) may be less available in certain neighborhoods.

The empirical identification of peer effects is challenging. First, one should distinguish social effects from correlated effects (selection). Someone's peer group tends to be a group

[2]See Sacerdote (2011) for a review on the literature of peer effects in education, and Cawley and Ruhm (2011) for a review of the literature on peer effects in wider health behaviors.

of individuals with similar characteristics and preferences, and so the correlation in outcomes such as smoking could simply be driven by similar preferences. Second, one should distinguish between endogenous social effects and exogenous social effects, commonly known as the reflection problem. Given a dependence of the peers outcome on the peers characteristics it is hard to distinguish between the two.

In the past two decades, many scholars in economics have attempted to estimate the endogenous peer effect in smoking.[3] In most of the early attempts (Gaviria and Raphael, 2001; Powell et al., 2005; Lundborg, 2006; Clark and Lohéac, 2007; Kooreman, 2007), the reflection problem is tackled by assuming contextual effects are absent.[4] In this case, (a subset of) peers' characteristics can serve as instrumental variables (IVs) for the endogenous peer outcome. Moreover, these studies typically used the whole classroom as the relevant peer group, such that correlated effects are minimized when class fixed effects are taken into account. Most of these studies estimate relatively large endogenous peer effects in smoking.

The next generation of studies has used specific IVs to identify the endogenous peer effects, whilst allowing for the influence of contextual effects. An early attempt was Norton et al. (1998), who used neighborhood characteristics as IVs for the endogenous peer effects. Later examples include Eisenberg (2004), who used a friend moving away or graduating as a shock to one's peer group, Fletcher (2010), who used the proportion of classmates of which a household member smokes as instrument for the group average smoking, Cutler and Glaeser (2010) who exploit workplace smoking bans as exogenous shocks in peer's (spousal) smoking behavior, and Argys and Rees (2008), who exploit birth- and kindergarten start dates as exogenous variation in the age of one's peers, and find that females with older peers are more likely to smoke – consistent with endogenous peer effects.

In recent years, scholars have either used random assignment of college roommates (Eisenberg et al., 2014), or a more structural approach that combined functional form assumptions with exclusion restrictions (Soetevent and Kooreman, 2007; Krauth, 2007; Card and Giuliano, 2013) to identify peer effects in smoking. With these increasingly convincing identification strategies, the resulting endogenous peer effects become gradually smaller, yet generally survive even in the most convincing designs. Hence, our reading of the literature is that peer effects in smoking seem well-established.

While the effect on the average individual seems well-established, the literature has hardly investigated heterogeneity in peer effects. Given that it is difficult to prevent adolescents from

[3]See Christakis and Fowler (2008) for evidence from the epidemiological literature, and Cohen-Cole and Fletcher (2008a,b); Fowler and Christakis (2008); Lyons (2011); VanderWeele et al. (2012) for methodological discussions of these findings.

[4]Kawaguchi (2004) used the individual's perception of peer behavior to overcome the reflection problem.

affiliating with peers that may exert negative influences, knowledge on mechanisms and which individuals are particularly susceptible to peer influences in smoking are critical for prevention efforts (Brechwald and Prinstein, 2011). Since peer influence is contingent on openness to influence/susceptibility (Brown et al., 2008), it seems particularly relevant to investigate the moderating role of personality.

2.2 Personality and smoking

The most widely-accepted taxonomy of personality (also known as non-cognitive skills) is the so-called "Big Five" (acronym OCEAN, Digman, 1990; Matthews et al., 2003). The five factors can be described as

1. Openness to experience ("the degree to which a person needs intellectual stimulation, change, and variety")
2. Conscientiousness ("the degree to which a person is willing to comply with conventional rules, norms, and standards")
3. Extraversion ("the degree to which a person needs attention and social interaction")
4. Agreeableness ("the degree to which a person needs pleasant and harmonious relations with others")
5. Emotional stability (or Neuroticism, "the degree to which an individual experiences the world as threatening and beyond his/her control")

While the association between personality and economic outcomes, including health, has been studied extensively in other disciplines (see e.g., Deary et al., 2010, for an overview of the psychological literature), in economics personality was for a long time understudied. Interest dates back at least to Bowles and Gintis (1976), but only recently became very popular mainly due to the work by James Heckman and co-authors (Heckman, 2000; Heckman et al., 2006). In particular, Heckman et al. (2006) suggests that personality is at least equally important as cognitive ability in determining adult's outcomes including health behaviors, and Heckman et al. (2013) suggests that influential pre-school programs were mainly effective in improving individual's earnings, health, and other socioeconomic outcomes by boosting personality traits.

There are only few studies in economics specifically studying personality traits and health behaviors. Fletcher et al. (2009) use Add Health data to show that individuals with low self control (mainly related to conscientiousness and emotional stability) are less responsive to cigarette taxes, consistent with behavioral economic models of cue-triggered addiction and self-control (Bernheim and Rangel, 2004; Gul and Pesendorfer, 2004). Chiteji (2010) uses the

US Panel Study of Income Dynamics (PSID) and finds that future orientation and self-efficacy (related to emotional stability) are associated with better health behavior. Cobb-Clark et al. (2014) use the Australian HILDA data and find that an internal locus of control (also related to emotional stability, whether you think life's outcomes are under our control) is related to better health behavior including reduced smoking. Mendolia and Walker (2014) use the Longitudinal Study of Young People in England and find that individuals with external locus of control, low self-esteem, and low levels of work ethics, are more likely to engage in risky health behaviors including smoking.

These studies suggest that there is an association between certain personality traits and risky health behaviors including smoking. Indeed, in comprehensive reviews of the psychology and economics literature, Borghans et al. (2008) and Almlund et al. (2011) conclude that especially conscientiousness and, to a slightly lesser extent, emotional stability are most important in determining later life economic and social outcomes, including health and smoking.

Despite a growing number of studies on personality and health behavior, the mechanisms are unexplored (Almlund et al., 2011). It is not known *how* personality affects health behavior and health outcomes. We hypothesize that the susceptibility to peer influences is one the mechanisms through which personality affects health behaviors. Since the effect of peer influence is known to be moderated by the "openness to influence", but also by the "salience of influencers" (Brown et al., 2008), it seems plausible that the personality of both the individual and his/her peers plays a role. Therefore, we will investigate heterogeneity in peer effects stratified by the personality of the individual and his/her peers, to test the hypothesis that personality is a key moderator of peer influence in smoking.

3 Data and descriptive statistics

Our study is based on the Add Health survey,[5] which is a longitudinal study on a nationally representative sample covering adolescents in grade 7 through 12 (average age from 12 to 17) from 132 schools. With the purpose of understanding how social environments and behaviors in adolescence are linked to health and achievement outcomes in young adulthood, the Add Health data contains detailed information about respondents' demographic backgrounds, academic performance, health related behaviors, psychological and physical well-being. Most

[5]This is a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). No direct support was received from grant P01-HD31921 for this analysis.

uniquely, the Add Health asked each respondent to nominate their male and female friends so that researchers can use the information to construct students' friendship networks.

Four waves of surveys were conducted from 1994 to 2008. In Wave I, a total of approximately 90,000 students were sampled and surveyed at school, and a subset of 20,745 students participated in the in-home survey. The in-home survey data contains more detailed questions on family background than the in-school survey data, and includes information on individual's personality characteristics. In the following waves, all surveys are conducted at home, tracking subsets of the total sample. We only use the Wave I in-home data for its advantage on data coverage. We focus on small- and mid-size schools that have less than three hundred students,[6] and we remove observations with missing values on basic demographic information. Eventually, we obtain a final sample with 9,748 students in 118 schools for our analysis.

Figure 1 depicts the distribution of the number of friendship nominations that is observed in our sample and Table 1 shows the average number of nominated friends stratified by the personality measures of emotional stability and conscientiousness. While the average number of nominated friends is 2.76 in the Add Health in-school sample, the average number of observed friends in the in-home survey is on average slightly more than 1. Hence we observe only a subset of the full friendship network among the in-home survey respondents, a phenomenon known as the missing link problem. Reassuringly, in section 6 we show that restricting to the so-called saturated sample for which we observe the full friendship network is not affecting our main results. Table 1 further shows that for emotional stability the average number of nominations along the diagonal is slightly larger than off the diagonal. This suggests that there is some homophily in terms of personality, something we will explore in more detail below.

We construct the main dependent variable of the paper, smoking, as a dummy variable. If a student reported he/she smoked at least once a month during each of the past twelve months, then we code him/her as a smoker (smoke dummy equal to one). Otherwise, we code him/her as a non-smoker (smoke dummy equal to zero).

The Add Health survey allows constructing three out of the big five personality characteristics during adolescence: emotional stability, conscientiousness, and extraversion.[7] We follow Young and Beaujean (2011) to measure the three personality dimensions by selecting 13 items

[6]We do this for a computational reason since the computation time required increases exponentially with network size.

[7]The timing of the personality measures is very similar to other surveys like the British Cohort Study (age 10, see Conti et al., 2010), National Child Development Study (age 7, 11, and 16, see Conti and Hansman, 2013), National Longitudinal Study of Youth 79 (age 14 to 21, see Heckman et al., 2006), Longitudinal Study of Young People in England (age 14, see Mendolia and Walker, 2014), and the Terman data (age 12, see Savelyev, 2014).

from the survey according to the Lexical approach and exploratory factor analysis. The details of these 13 items are in Table 2. The responses of items are ranging from 1 (strongly disagree) to 5 (strongly agree). We conduct a factor analysis on these items and identify one main factor for each personality measure, which explains more than 90% of variation in the corresponding items. We report the rotated factor loadings for each item and the regression coefficients for predicting factor scores (Thomson, 1951). The predicted scores have a zero mean, and the sign and the magnitude reflect individuals' personalities.

Borghans et al. (2008) suggest that particularly emotional stability and conscientiousness are important in determining smoking. For this reason, we will explore heterogeneity in peer effects along those two dimensions. We do allow extraversion to influence smoking decisions, and the nomination of friends, but we will not investigate heterogeneity in the peer effect with respect to extraversion. This is because for extraversion, other than for conscientiousness and emotional stability, there is no obvious taxonomy of 'weak' and 'strong'. Moreover, we found evidence that extraversion is potentially affected by peers (see section 6.1), such that the subgroups defined by extraversion are endogenously determined and subject to change depending on the composition of the peer group.

In the model specification we additionally include a wide array of demographic and socioeconomic characteristics that determine the individual's smoking decision ('own effects'), and also the smoking decisions of his/her friends ('contextual effects'). Most variables are relatively standard and are listed in Table 3. The variables of low parent control (e.g., "do your parents let you make your own decisions?") and maternal care (e.g., "How much do you think your mother cares about you?") are constructed from Add Health Wave I in-home survey following Driscoll et al. (2008) and Shakya et al. (2012) by taking average responses from seven and four survey questions, respectively.

Table 3 provides the summary statistics of variables used in this study. Based on the whole sample, 22.4% of students are identified as smokers. There are slightly more girls (53.4%) than boys (46.6%) in our sample. In terms of race, White (54%) is the majority in the sample, followed by Black (22.8%) and Asian (11%). 94% of students report that they have received information on the health consequence of smoking (school taught) in class. There are 64.3% of students having at least one parent previously or currently smoking (smoke parent) at home. The average of low parent control is 0.741 (the value 1 represents weak control and the value 0 represents strong control), and the average of maternal care is 4.55 (the value 5 represents high warmth and the value 1 represents low warmth).

4 Methodology

4.1 SAR model

The traditional workhorse model for studying social interactions is the linear-in-means model. However, the linear-in-means model suffers from the reflection problem (Manski, 1993), which prevents researchers from distinguishing between endogenous and contextual peer effects. The reflection problem can be solved by utilizing information of detailed friendship links among individuals, summarized by a spatial weight matrix in the Spatial Autoregressive (SAR) model. In the SAR model, both the endogenous and contextual effects are identified as long as individuals' friends are not perfectly overlapping (Bramoullé et al., 2009; Lin, 2010; Lee et al., 2010).

Most of the existing SAR model applications (Bramoullé et al. (2009); Lin (2010); Lee et al. (2010); Hsieh and Lee (2015); Fortin and Yazbeck (2015); among others) focus on a homogenous endogenous peer effect, which can be regarded as the average of heterogeneous peer effects.[8] While the average endogenous peer effect in smoking initiation is certainly of interest, our main objective is to study the moderating role of personality. In other words, we intend to investigate whether peer effects are stronger among individuals with 'weaker' personalities. Therefore, we extend the conventional SAR model to capture heterogeneous endogenous peer influences from friends with different personalities.

Our model considers an environment where students are placed in schools $g \in \{1, \cdots, G\}$. In school g, student i's smoking behavior is represented by the variable $y_{i,g}$ and his/her personality is represented by a R-dimensional row vector $s_{i,g}$. The other individual exogenous characteristics are represented by a K-dimensional row vector $x_{i,g}$. The vector $Y_g(m_g \times 1)$, matrix $S_g(m_g \times R)$, and matrix $X_g(m_g \times K)$ summarize smoking variables, personalities, and characteristics of m_g students in school g, respectively. The friendship network in group g is represented by a $m_g \times m_g$ spatial weight matrix W_g. Each element of W_g, $w_{ij,g}$, is a binary indicator which equals one if individual i sends a friendship nomination to individual j, and zero otherwise. Since friendship nominations are directional without guaranteed reciprocality, W_g is not necessarily symmetric.

The heterogeneous network interaction equation for student i's smoking moderated by the

[8] Some recent exceptions include Card and Giuliano (2013), who study heterogeneity with respect to gender, and Lin and Weinberg (2014), who study heterogeneity with respect to reciprocated and unreciprocated friends.

r^{th} personality measure is specified as

$$\begin{aligned}
y_{i,g} &= \lambda_{11} I(s_{ir,g} < \overline{S}_{r,g}) \sum_{j\neq i} w_{ij,g} I(s_{jr,g} < \overline{S}_{r,g}) y_{j,g} \\
&+ \lambda_{12} I(s_{ir,g} < \overline{S}_{r,g}) \sum_{j\neq i} w_{ij,g} I(s_{jr,g} \geq \overline{S}_{r,g}) y_{j,g} \\
&+ \lambda_{21} I(s_{ir,g} \geq \overline{S}_{r,g}) \sum_{j\neq i} w_{ij,g} I(s_{jr,g} < \overline{S}_{r,g}) y_{j,g} \\
&+ \lambda_{22} I(s_{ir,g} \geq \overline{S}_{r,g}) \sum_{j\neq i} w_{ij,g} I(s_{jr,g} \geq \overline{S}_{r,g}) y_{j,g} \\
&+ x_{i,g}\beta_1 + \sum_{j\neq i} w_{ij,g} x_{i,g}\beta_2 + s_{i,g}\beta_3 + \sum_{j\neq i} w_{ij,g} s_{i,g}\beta_4 + \alpha_g + \epsilon_{i,g} \qquad (1)
\end{aligned}$$

for $i = 1, \cdots, m_g$, where $I(A)$ denotes an indicator function which equals one if A is satisfied and equals zero, otherwise. $\overline{S}_{r,g}$ denotes the average of the r^{th} personality measure in group g.

The innovation of Eq. (1) compared to the conventional SAR model, is to allow for differential peer effects according to individuals' own and friends' personalities. To be specific, the coefficient λ_{11} captures the endogenous peer effect for a pair of individuals that both have weak personalities. Coefficients λ_{12} and λ_{21} capture endogenous peer effects for the cases that one individual has a strong, but the other one has a weak personality. The coefficient λ_{22} considers the case that both individuals have strong personalities. Whether an individual's personality is regarded as strong or weak is determined by whether their personality score is below or above the sample average.

The coefficients β_1 and β_3 capture the individual ('own') effect of exogenous characteristics x and personality s, respectively. β_2 and β_4 reflect the contextual effects from exogenous characteristics and personalities, respectively.[9] The term α_g represents the group fixed effect, which plays a key role in capturing the environmental correlated effects shared by all members in the same group, e.g., teacher quality, classroom facility, etc. The error term $\epsilon_{i,g}$ is assumed normally distributed with a zero mean and a variance equal to σ_ϵ^2.

For the ease of presentation, the vector expression of Eq. (1) is,

$$Y_g = \lambda_{11} W_{11,g} Y_g + \cdots + \lambda_{22} W_{22,g} Y_g + X_g\beta_1 + W_g X_g \beta_2 + S_g\beta_3 + W_g S_g \beta_4 + \ell_g \alpha_g + \epsilon_g, \qquad (2)$$

for $g = 1, \cdots, G$. The spatial weight matrix W_g is now divided into 2×2 blocks, where each block, W_g^{pq}, $p, q = 1, 2$, represents the subnetwork between individuals in the personality

[9] Although it is straightforward to generalize Eq. (1) with heterogeneous contextual effects, in order to focus on the endogenous effect as well as maintain model parsimony, we leave contextual effects to be homogenous in this paper.

subgroups 1 and 2. $W_{pq,g}$ is a $m_g \times m_g$ matrix with the corresponding $(p,q)^{\text{th}}$ block equal to W_g^{pq} and 0 elsewhere, X_g and S_g are matrices of individuals characteristics and personalities, respectively, and ℓ_g is a $m_g \times 1$ vector of ones.

One choice regarding the SAR model specification is whether to use the raw spatial weight matrix or to row-normalize it. In the raw case, every friend receives a weight of one, while the row-normalizing case ensures that the sum of each row of the spatial weight matrix equals one. For example, if an individual nominates four friends, they all receive a weight of one-fourth. In Liu et al. (2014), they interpret the SAR model as the "local average" model if the matrix is row-normalized, with network participants obtaining higher marginal utilities by conforming to the social norm of their reference groups. In the other case where the raw matrix is used, they interpret the SAR model as the "local aggregate" model where network participants obtain higher marginal utilities from having more active friends. In our study with heterogeneous peer effects from four blocks of the spatial weight matrix, we choose to use the raw weight matrix for two reasons. First, when normalizing, it is unclear whether the four blocks of the matrix should be normalized individually or jointly by the aggregated row-sums from all blocks. Second, in the case of a row-normalized matrix, the endogenous peer effects obtained from different blocks are insensitive to the number of friendship links in each block. This is unlikely to be true, and additionally restricts the multiplier effects to be the same within personality types irrespective of the number of friends. In contrast, the un-normalized model allows having different multiplier effects for each individual on basis of the different number of friends in each personality subgroup.

4.2 SC-SAR model with unobserved individual heterogeneity

SC-SAR model The SAR model is fully capable of solving the reflection problem. However, the issue of correlated effects cannot be adequately solved by the conventional SAR model. One can include network fixed effects to account for common environments among individuals within the network, but one cannot rule out that there are individual correlated effects. Unobserved individual characteristics that are correlated to smoking may also affect the selection of friends. For example, an individual's attitude toward freshness and excitement, which is an unobserved personal trait, may not only affect the smoking decision, but also the friendship choices. As a result, the peers' outcome will – indirectly through the selection of friends – be influenced by the same characteristics that influence your own outcome. In terms of equation (2), the matrices $W_{pq,g}$, $p,q = 1,2$, and the outcome vector Y_g are both influenced by some unobserved individual traits, so $W_{pq,g}$, $p,q = 1,2$ is endogenous, and the estimates of the endogenous peer effects will be biased.

To overcome this issue, Hsieh and Lee (2015) introduced the selection corrected-SAR (SC-

SAR) model. Effectively, the SC-SAR model introduces an additional equation in which the spatial weight matrix W_g is endogenously determined, and allows observed and unobserved (latent) characteristics to influence both the friendship link formation and the individual's outcome. The latent variables are called $z_{i,g}$ and are assumed to be multidimensional (with a total of $\bar{d}$ dimensions) to accommodate the unknown number of underlying individual correlated effects. The outcome equation of the SC-SAR model can be written down as

$$Y_g = \lambda_{11} W_{11,g} Y_g + \cdots + \lambda_{22} W_{22,g} Y_g + X_g \beta_1 + W_g X_g \beta_2 + S_g \beta_3 + W_g S_g \beta_4 + Z_g \delta_1 + \ell_g \alpha_g + u_g, \tag{3}$$

where the error term u_g is uncorrelated with all regressors in the equation, and $Z_g = (z'_{1,g}, \cdots, z'_{m_g,g})'$.[10] Essentially, compared to the SAR model of Eq.(2), the extra term $Z_g\delta_1$ in Eq.(3) represents a control function to handle the endogeneity of W_g (Navarro, 2008).

The link formation equation of the SC-SAR model endogenously models the individual elements $w_{ij,g}$ of the spatial weight matrix W_g, and is specified as:

$$P(w_{ij,g}|c_{ij,g}, s_{ig}, s_{j,g}, z_{i,g}, z_{j,g}) = \left(\frac{\exp(\psi_{ij,g})}{1+\exp(\psi_{ij,g})}\right)^{w_{ij,g}} \left(\frac{1}{1+\exp(\psi_{ij,g})}\right)^{1-w_{ij,g}},$$
$$\psi_{ij,g} = c_{ij,g}\gamma_0 + \gamma_1|s_{i1,g} - s_{j1,g}| + \cdots + \gamma_R|s_{iR,g} - s_{jR,g}| + \eta_1|z_{i1,g} - z_{j1,g}| + \cdots + \eta_{\bar{d}}|z_{i\bar{d},g} - z_{j\bar{d},g}|. \tag{4}$$

Hence, the probability to form a link between individual i and individual j in group g, $P(w_{ij,g})$ in Eq. (4), is determined by a latent index $\psi_{ij,g}$. This latent index in turn is determined by the difference in characteristics between individuals i and j, where the more dissimilar individuals are, the less likely they are to become friends. We include differences in personality between individuals to capture the homophily effect from personalities. In the same spirit, the difference in the latent variables z reflects homophily in terms of unobserved characteristics between individuals i and j. $c_{ij,g}$ represents the $\bar{q}$-dimensional dyad-specific variables between individuals i and j, e.g., whether individual i and j are same gender, age, and race for example. These dyad-specific variables are the exclusion restrictions to identify the network interactions, as they can be naturally excluded from the outcome equation.

[10] Hsieh and Lee (2015) assume that the error term of the outcome equation $\epsilon_{i,g}$ and the latent variables $z_{i,g}$ follow an i.i.d. multivariate normal distribution,

$$(\epsilon_{i,g}, z_{i,g})' \sim i.i.d.\ \mathcal{N}_{\bar{d}+1}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon z} \\ \sigma_{z\epsilon} & \Sigma_z \end{pmatrix}\right).$$

According to the covariance structure between $\epsilon_{i,g}$ and $z_{i,g}$, we have $\epsilon_{i,g} = z_{i,g}\Sigma_z^{-1}\sigma_{z\epsilon} + u_{i,g}$, where $u_{i,g} \sim \mathcal{N}(0, \sigma_u^2)$ and $\sigma_u^2 = \sigma_\epsilon^2 - \sigma_{\epsilon z}\Sigma_z^{-1}\sigma_{z\epsilon}$. In terms of identification, this joint normal distribution assumption is not necessary, and we only require the following moment conditions: $(\epsilon_{i,g}, z_{i,g})$ are i.i.d., $\mathrm{E}(\epsilon_{i,g}) = 0$, $\mathrm{E}(z_{i,g}) = 0$, $\mathrm{Var}(\epsilon_{i,g}) = \sigma_\epsilon^2$, $\mathrm{Var}(z_{i,g}) = \Sigma_z$, $\mathrm{E}(\epsilon_{i,g}, z_{i,g}) = \sigma_{\epsilon z}$, and $\mathrm{E}(\epsilon_{i,g}|z_{i,g}) = z_{i,g}\Sigma_z^{-1}\sigma_{z\epsilon}$. Under this assumption, δ_1 in Eq. (3) is restricted to $\delta_1 = \Sigma_z^{-1}\sigma_{z\epsilon}$.

Unobserved contextual effects While the SC-SAR model introduced by Hsieh and Lee (2015) goes a long way in dealing with the endogeneity of the spatial weight matrix W_g, it still does not fully account for the potential endogeneity of the peer's outcome. As argued by Fruehwirth (2014), the peer outcome is likely to reflect unobserved contextual effects and this will contaminate the true endogenous peer effects. In simple terms, if the friends smoking decision is determined partly by, say, intelligence, and we do not control for the contextual effect of friend's intelligence, then this will be absorbed into the endogenous peer effect of smoking.

To estimate a clean version of the endogenous peer effects, we add the contextual latent variables $W_g Z_g$ into the outcome equation to account for possibly omitted contextual effects. Thus, the extended outcome equation is written as,

$$\begin{aligned} Y_g = {} & \lambda_{11} W_{11,g} Y_g + \cdots + \lambda_{22} W_{22,g} Y_g + X_g \beta_1 + W_g X_g \beta_2 + S_g \beta_3 + W_g S_g \beta_4 \\ & + Z_g \delta_1 + W_g Z_g \delta_2 + \ell_g \alpha_g + u_g, \end{aligned} \tag{5}$$

where $u_g \sim i.i.d.\ \mathcal{N}_{m_g}(0, \sigma_u^2 I_{m_g})$.

The extended outcome model of Eq. (5), combined with the link formation model of Eq. (4) forms the extended SC-SAR model to account for (i) the reflection problem through the use of non-overlapping friendship nominations in the SAR model, (ii) network-level correlated effects through the use of network fixed effects, (iii) the endogeneity of the spatial weight matrix through the use of dyad-specific variables as exclusion restrictions and latent variables in both equations of the SC-SAR model, and (iv) unobserved contextual effects through our extension of the SC-SAR model. The extended SC-SAR model will be the main model we use to study heterogeneous peer effects on smoking.

We follow Hsieh and Lee (2015) to use a Bayesian approach to estimate this extended SC-SAR model, which is effective in handling estimation of models with latent variables (Zeger and Karim, 1991). A full discussion of the identification and estimation of the SC-SAR model can be found in supplementary appendices A and B.

4.3 Placebo tests

While it is impossible to test directly whether our extended SC-SAR has solved all endogeneity issues completely, simulation results suggest that a SC-SAR model performs well in eliminating sources of correlated effects (Hsieh and Lee, 2015). Here, we use a placebo test to present complementary evidence that our extended SC-SAR model is able to deal with some of the most notorious sources of correlated effects.

The idea is as follows. We pick an outcome measure for which, realistically, the endogenous peer effect is zero. An example is father's education: it is highly unlikely that the educational level of the father of your peers is going to affect your own father's educational level. Hence, we expect no endogenous peer effect. Nonetheless, it is perceivable that your own father's educational level and the father's educational level of your peers are correlated, since you may select your peers on basis of social background, of which father's education is one proxy. Hence, in model specifications where individual correlated effects are not sufficiently accounted for, one is likely to find a spurious endogenous peer effect of father's education.

Indeed, in the regular SAR model (see the first three columns of Table 4) the endogenous peer effect is estimated to be positive and statistically significant for father's education. This strongly suggests that the regular SAR model does not fully take into account individual correlated effects, even when group fixed effects, observed individual characteristics, and contextual effects, are controlled for (cf. columns 1-3 of Table 4).

The idea of our extended SC-SAR model is to account for *unobserved* factors that correlate with both the outcome and the selection of peers. When adding dimensions of unobserved factors, we expect the extended SC-SAR model to approach the true endogenous peer effect of zero. It can be seen that correlated effects can be persistent, since even in the SC-SAR model with a one-dimensional latent factor (as in Goldsmith-Pinkham and Imbens, 2013), there is evidence for an endogenous peer effect of father's education. When increasing the dimension of the latent factors (moving from column 4 to 8 in Table 4), the effect becomes statistically insignificant. In line with the evidence from simulations of Hsieh and Lee (2015), it takes multi-dimensional latent factors to overcome the issue of correlated effects. This suggests that, in contrast to the standard SAR model, a higher-dimensional SC-SAR model is capable of dealing with some of the most notorious and persistent sources of correlated effects.

5 Results

5.1 The peer effects on smoking: SAR model

We first present the baseline estimate for the peer effect on smoking from the SAR model in Table 5. When a homogenous peer effect is considered, the estimated endogenous effect equals 0.0922, which implies that when one of individual's friends changes from a non-smoker to a smoker, the individual increases his/her chance of being a smoker by 9.22 percentage points. The total peer effect is roughly the same when we multiply this estimate with the number of peers, since the average number of peers is one in this sample (Table 1). While this estimate of the SAR model is subject to potential bias due to individual correlated effects, the estimate is very close to the result obtained in Card and Giuliano (2013), and to the range reported in the literature varying from around 0.05 (Clark and Lohéac, 2007; Fletcher, 2010) to around

0.15 (Gaviria and Raphael, 2001; Powell et al., 2005; Lundborg, 2006; Krauth, 2007).

Adding personality measures as control variables (in column 2) does not fully account for individual correlated effects, yet may alleviate part of the omitted variable problem, and the endogenous peer effect decreases to 0.0842. In line with the evidence discussed in section 2.2, students with stronger personalities tend to smoke less. The three parent-related variables, parent smoke, low parent control, and strong maternal care, all have significant effects on their children's smoking behaviors and the signs of effects are in line with our expectation. Black and Asian students are less likely to be smokers compared to their white counterparts. Most of the contextual effects are non-significant (not shown), yet it seems that peer's extraversion has a negative effect on individual's smoking.

When studying heterogeneous peer effects (columns 3 to 4), it is found that the peer effect from "weaker" personalities to "weaker" personalities is strongest for both personality measures, albeit more strikingly so for emotional stability than for conscientiousness. Two individuals who are both emotionally unstable affect each other way more in smoking, compared with two individuals of which at least one is emotionally stable. This holds to a lesser extent for conscientiousness. These baseline findings are suggestive of important heterogeneity in peer effects, yet may suffer from unobserved variables affecting both the outcome and friendship formation. Therefore we turn to the extended SC-SAR model.

5.2 The peer effects on smoking: extended SC-SAR model

Estimation results of the extended SC-SAR model are reported in Table 6. On basis of theoretical and empirical criterions, we focus on the cases with latent variables in four and five dimensions and leave other cases (with lower dimensions) available upon request. The theoretical criterion we employ is the AICM (Akaike's information criterion - Monte Carlo) proposed by Raftery et al. (2007).[11] The empirical criterion is based on the idea that the estimates from the SC-SAR model should at some point stabilize after increasing the dimension of the latent variables. From Table 6, we conclude that the SC-SAR(4) model is the preferred model on basis of both criterions.

We start with discussing the estimates of the link formation model in the middle panel of Table 6. The first rows show that there is strong homophily in terms of grade, sex, and race: individuals in the same grade, and of the same sex and race tend to nominate each other as friends. Individuals tend to hang out with other individuals who are in the same

[11] AICM is an estimate of the conventional AIC, which is not directly obtained from the posterior simulation as the maximum loglikelihood value may not be available. Given that the distribution of the loglikelihoods from each posterior draws is approximately a gamma distribution, we can obtain an estimate of AICM as well as its standard error. Same as the conventional AIC, the model with a lower AICM value is favored.

grade, and are similar in terms of sex and race. This is reassuring, since these variables help identifying the endogenous friendship formation, and are naturally excluded from the outcome equation.[12] The link formation results further help to understand how personality affects friendship formation. The result shows that students are less likely to hang out together if they have different levels of extraversion. Differences in emotional stability also affect friendship formations, where the coefficients are on the margin of being significant. However, the difference in conscientiousness levels does not seem to matter.

The top panel of Table 6 shows the endogenous peer effects in smoking. For the homogenous endogenous peer effect on smoking, the estimate equals 0.0374, suggesting that on average, when a friend starts smoking, you are 3.74 percentage points more likely to start smoking. With a baseline smoking prevalence of 22 percent, this is equivalent to an effect size of 17 percent. The SC-SAR estimate is less than half of the estimate obtained from the SAR model, and lower than most of the peer effects estimated in the literature. Hsieh and Lee (2015) reported a similar percentage of bias correction in the SC-SAR model when studying student's academic performance as the dependent variable. Taken together, this suggests that the conventional SAR model erroneously assumes that the spatial weight matrix is exogenously given, and overestimates endogenous peer effects substantially.

The heterogenous peer effects are presented in the remaining columns of Table 6. It can be observed that although the values of estimates also become smaller compared to Table 5, the same pattern across peers of different personalities remains. Two individuals who are both emotionally unstable influence each other way more in terms of smoking than two individuals of which at least one is emotionally stable. This difference is statistically significant, as the highest posterior density intervals are not overlapping, and cannot be fully accounted for by the difference in smoking prevalence across personality subgroups (18 vs. 26 percent). Emotional stability is thus a very important moderator of peer effects in smoking. Interestingly, for conscientiousness this is not the case. While the weak-to-weak interactions produce slightly larger point estimates compared with the other interactions, these differences are not statistically significant. In fact, moving across rows suggest that the heterogeneity in peer effects with respect to conscientiousness is very modest, and all interactions are close to the average (homogenous) peer effect.

In Figure 2 we plot the distribution of the social multiplier effects based on the result in column 3 of Table 6. The social multiplier is the predicted total impact of an individual starting smoking, taking into account both the direct 'own' effect and the indirect effect

[12] The individual's own grade, race, and gender, and the average grade, race, and gender of his/her friends may still affect the smoking decision, but we assume that the dyad specific pairs (e.g., individual i and j share the same sex and race) do not affect individual's i smoking decision.

running through the impact on his/her peers, and hence is always larger than or equal to 1.[13] The figure shows the social multipliers separately for groups of strong and weak emotional stabilities. In line with the point estimates, one can see that the emotionally unstable students generally experience higher multiplier effects than the emotionally stable students.

6 Robustness checks

6.1 Endogenous personality

When stratifying the sample on basis of personality, one may be worried that personality is endogenous. This could be either because one's personality is affected by the peer group, or since personality is correlated to unobserved variables that also affect the outcome. There is a good deal of evidence that genetic effects explain a large, up to 2/3, fraction of the variance in personality (Bouchard Jr., 1994; Bouchard and Loehlin, 2001). As a result, conscientiousness and emotional stability vary very little between ages 10 and 20 (Roberts et al., 2006; Roberts and Mroczek, 2008; Cobb-Clark and Schurer, 2012). This suggests that personality is relatively stable in high school, and not severely influenced by peers, and justifies our treatment of personality as being exogenous in the main analysis.

Nonetheless, in this section we will investigate the robustness of our results by treating personality as being endogenously determined. We will do so by introducing an additional equation in which personality is endogenously determined, and we allow the unobserved latent variables to additionally influence personality, apart from their influence on friendship decisions and smoking. Ideally one would like to model the decisions regarding personalities and network links as a general simultaneous equation system, in which personality affects friendship decisions, and friends in turn may affect each other's personalities. However, no exogenous instrumental variables for either personality or network links are available, and so the simultaneous equation system is not identified. Instead, we consider two alternative restrictions on the simultaneous equations that permit identification.

In the first approach, endogenous personalities are allowed to affect friendship decisions, but we restrict the endogenous peer effect (i.e., the effect operating through network links) on personality to be zero. This seems a reasonable assumption given the large genetic component of conscientiousness and emotional stability. Accordingly, individual's r^{th} personality is modeled by a simple linear regression without an endogenous peer effect or contextual effects,

$$s_{ir,g} = x_{i,g}\phi_{1r} + z_{i,g}\tau_{1r} + \kappa_{r,g} + v_{ir,g}, \quad v_{ir,g} \sim \mathcal{N}(0, \sigma^2_{v_r}), \tag{6}$$

[13] The social multipliers are calculated by the formula $(I_{m_g} - \lambda_{11}W_{11,g} - \cdots - \lambda_{22}W_{22,g})^{-1}\ell_g$, $g = 1, \cdots, G$.

where $\kappa_{r,g}$ stands for the group correlated effect for personality, and recall that $z_{i,g}$ captures the latent variables that additionally influence the friendship and smoking decisions.

In the second approach, we allow for endogenous peer effects in personality, yet assume that the effect of personality on friendship formation is zero. The corresponding equation for endogenous personality is specified as

$$s_{ir,g} = \rho_r \sum_{j \neq i} w_{ij,g} s_{jr,g} \\ + x_{i,g}\phi_{1r} + \sum_{j \neq i} w_{ij,g} x_{j,g} \phi_{2r} + z_{i,g}\tau_{1r} + \sum_{j \neq i} w_{ij,g} z_{j,g} \tau_{2r} + \kappa_{r,g} + v_{ir,g}, \quad v_{ir,g} \sim \mathcal{N}(0, \sigma^2_{v_r}), \tag{7}$$

where personality is potentially influenced by friendship networks though endogenous and contextual peer effects. We estimate equations (6) and (7) for both personality measures, along with the SC-SAR extension of smoking outcome in Eq. (5) and the link formation model in Eq. (4).

The results based on equation (6) are in Table 7. In order to compare with Table 6 where personalities are treated exogenously, we again report the cases with latent variables in four and five dimensions. Most important observation is that the endogenous peer effects vary very little between the two tables, and if anything the endogenous peer effects are larger when the potential endogeneity of personality is taken into account. Also, our main finding that the social interactions between two individuals with weaker personalities are most important holds up in the setting with endogenous personality.

The results based on equation (7) in Table 8 reassuringly show that emotional stability and conscientiousness are not affected by peers. Note that extraversion is endogenously affected by peers, which was our reason for not stratifying the sample on basis of extraversion. When treating personality as being endogenously determined, as in Table 7, the estimates of endogenous peer effects from the SC-SAR model if anything become higher after endogenizing personalities, and the interactions between weaker personalities produce the largest peer effects. Overall, we conclude that the possible endogeneity of personality is not driving our results.

6.2 Full network based on saturated sample

While we require the more extensive list of variables from the Add Health in-home survey to comprehensively study whether peer effects in smoking are moderated by personality, the main concern is the missing links problem (Chandrasekhar and Lewis, 2011; Liu, 2013). That is, not all nominated friends from the in-school survey were selected for the in-home interview, and the social interactions matrix is not complete when only focusing on the in-home respondents. This could potentially result in biased estimates of the endogenous peer effect.

One way to check the robustness of our results to the missing link problem is to restrict our sample to the so-called "saturated" schools in the Add Health survey. Our sample contains 13 schools (with a total of 705 students) in which all students in school were selected for the in-home survey. For the group of students in these schools, we do observe the full friendship network, and the average number of friendship nominations made in these 13 schools is 3.45, which is larger than the values reported in Table 1.

Table 9 presents the results. While the homogenous endogenous peer effect is statistically insignificant, our main result that students who are both emotionally unstable are most severely affecting each other in smoking, is strongly confirmed in this saturated sample. This corroborates that the average peer effect masks considerable heterogeneity, and shows that possible missing network links are not affecting our main findings.

6.3 Other health behaviors

Network interaction process based on binary variables may involve the issue of multiple equilibria (Brock and Durlauf, 2001; Krauth, 2006; Soetevent and Kooreman, 2007; Nakajima, 2007). In our main analysis we neglect this issue, such that results are not contingent on additional model assumptions and structures. As a robustness check, we construct a continuous variable of smoking intensity by looking at how frequent students smoke in a week (smoking frequency). The estimation results for smoking frequency are reported in Table 10. Compared to the binary smoking results in Table 6, the magnitude of the endogenous peer effect on smoking frequency is found to be higher in absolute terms, yet somewhat smaller relatively with an effect size of around 8 percent. Our main finding that the emotionally unstable face larger peer effects in terms of smoking still holds.

The final robustness check is to study how specific results are for the case of smoking. Smoking is known to (i) severely affect health, (ii) be related to personality, and (iii) be vulnerable to peer pressure, and therefore represents an excellent outcome of our study. Nonetheless, if our results are specific to smoking, this would limit the moderating role of personality in general social interactions. Therefore, we estimated the same set of models for another health behavior that is known to affect health and to be vulnerable to peer pressure: the prevalence of getting drunk.

In Table 11 we present the homogenous and heterogeneous peer effects on getting drunk. The average peer effects are slightly smaller for getting drunk than for smoking, in particular given the higher prevalence of getting drunk vs. smoking (31 percent vs. 22 percent). Nonetheless, we observe a very similar pattern when stratifying the sample with respect to personality. The interactions between two individuals who are both emotionally unstable are most vulnerable to peer effects, while the other interactions are closer to the average peer effect. This suggests that the vulnerability of social interactions among emotionally unstable individuals is not specific to smoking, but a general finding that exists across a wider set of unhealthy behaviors.[14]

7 Discussion

We used friendship nominations from the Add Health data to study peer effects in smoking; in particular, whether peer effects are stronger among individuals with weaker personalities. We extended the SAR model to correct for the endogeneity of the spatial weight matrix, and allowed for unobserved contextual effects. Our results suggest that the conventional SAR model overestimates endogenous peer effects, yet that our extended SC-SAR model provides a promising method for studying homogenous and heterogeneous peer effects. Our main conclusion is that an individual's personality plays a very important role in social interactions, and it does so along two main dimensions.

First, the average peer effect masks considerable heterogeneity in responses with respect to personality. While the average peer effect in smoking is non-negligible, it is particularly pronounced within certain subgroups of students. We show that the individual personality trait of emotional stability is an important moderator of peer effects. Interactions between individuals who are both emotionally unstable are significantly more vulnerable to peer effects than interactions in which at least one of the students is emotionally stable. Despite being an important driver of smoking decisions, we find no evidence that conscientiousness plays a similar role in social interactions.

Second, apart from moderating the vulnerability to peer effects, personality also plays a role in friendship formation: individuals with similar personalities are more likely to hang out with each other. This is consistent with studies in social psychology (Cohen and Prinstein, 2006; Allen et al., 2012) that suggest that while generally people value the opinion of high-status or popular peers, not all adolescents desire identification with high-status peers. Instead, they will adopt a local set of norms that may be more salient to these adolescent's identity development.

These two findings suggest a potentially harmful cycle in which emotionally unstable individuals are more likely to smoke in the first place, are likely to hang out with other emotionally unstable individuals, and in turn lack the skills to stand up against the peer pressure of ini-

[14] We also studied the frequency of exercise, but could not reject a zero average effect. This suggests that peer effects are more important for unhealthy behaviors than for healthy behaviors.

tiating/continuing smoking. This pattern is not restricted just to smoking, but also holds for the prevalence of getting drunk. Hence, the benefits of a strong personality in social interactions provide a promising mechanism through which personality affects health behavior, and potentially even socioeconomic life outcomes.

In terms of policy implications, policy makers and teachers should at least realize the dominant role that personality plays in social interactions. On the positive side, our findings suggest multiple options for breaking the cycle between personality and peer pressure. One difficult but rewarding option could simply be to train emotional stability in early childhood programs, and evidence shows this can be effective at young ages (Heckman, 2000). Second, one could try to target emotionally unstable students in high school and teach them skills to stand up against peer pressure. Finally, one could try to mix students more on basis of personality to avoid groups of students who are all emotionally unstable. In many cases subgroups in school are formed on the basis of cognitive skills, but there is no reason why that could not be done on basis of non-cognitive skills. We should acknowledge however that our effects are estimated on basis of the current classroom composition, and people are likely to respond to the new situation (Graham et al., 2010; Carrell et al., 2013; Fruehwirth, 2014).

References

Allen, Joseph P., Joanna Chango, David Szwedo, Megan Schad, Emily Marston. 2012. Predictors of susceptibility to peer influence regarding substance use in adolescence. *Child Development* **83**(1) 337–350.

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, Tim Kautz. 2011. Personality psychology and economics. *Handbook of the economics of education* **4**(1).

Angrist, Joshua D. 2014. The perils of peer effects. *Labour Economics* **30** 98–108.

Argys, Laura M, Daniel I Rees. 2008. Searching for peer group effects: A test of the contagion hypothesis. *The Review of Economics and Statistics* **90**(3) 442–458.

Bernheim, B Douglas, Antonio Rangel. 2004. Addiction and cue-triggered decision processes. *American Economic Review* **94**(5) 1558–1590.

Borghans, Lex, Angela Lee Duckworth, James J. Heckman, Bas ter Weel. 2008. The economics and psychology of personality traits. *Journal of Human Resources* **43**(4) 972–1059.

Bouchard, Jr., Thomas J., John C. Loehlin. 2001. Genes, evolution, and personality. *Behavior Genetics* **31**(3) 243–273.

Bouchard Jr., Thomas J. 1994. Genes, environment, and personality. *Science* **264**(5166) 1700–1701.

Boucher, Vincent, Yann Bramoullé, Habiba Djebbari, Bernard Fortin. 2014. Do peers affect student achievement? evidence from canada using group size variation. *Journal of Applied Econometrics* **29**(1) 91–109.

Bowles, Samuel, Herbert Gintis. 1976. *Schooling in capitalist America*, vol. 57. New York: Basic Books.

Bramoullé, Yann, Habiba Djebbari, Bernard Fortin. 2009. Identification of peer effects through social networks. *Journal of Econometrics* **150**(1) 41 – 55.

Brechwald, Whitney A., Mitchell J. Prinstein. 2011. Beyond homophily: A decade of advances in understanding peer influence processes. *Journal of Research on Adolescence* **21**(1) 166–179.

Brock, W. A., S. N. Durlauf. 2001. Discrete choice with social interactions. *The Review of Economic Studies* **68(2)**.

Brown, B Bradford, Jeremy P Bakken, Suzanne W Ameringer, Shelly D Mahon. 2008. A comprehensive conceptualization of the peer influence process in adolescence. M.J. Prinstein, K.A. Dodge, eds., *Understanding peer influence in children and adolescents*. The Guilford Press, 17–44.

Campbell, Frances, Gabriella Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, Yi Pan. 2014. Early childhood investments substantially boost adult health. *Science* **343**(6178) 1478–1485.

Card, David, Laura Giuliano. 2013. Peer effects and multiple equilibria in the risky behavior of friends. *Review of Economics and Statistics* **95**(4) 1130–1149.

Carrell, Scott E., Bruce I. Sacerdote, James E. West. 2013. From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica* **81**(3) 855–882.

Cawley, John, Christopher J. Ruhm. 2011. The economics of risky health behaviors. Thomas G. Mcguire Mark V. Pauly, Pedro P. Barros, eds., *Handbook of Health Economics*, *Handbook of Health Economics*, vol. 2. Elsevier, 95 – 199.

Chaloupka, Frank J., Kenneth E. Warner. 2000. The economics of smoking. Anthony J. Culyer, Joseph P. Newhouse, eds., *Chapter 29*, *Handbook of Health Economics*, vol. 1, Part B. Elsevier, 1539 – 1627.

Chandrasekhar, Arun, Randall Lewis. 2011. Econometrics of sampled networks. *Unpublished manuscript, MIT.[422]* .

Chiteji, Ngina. 2010. Time preference, noncognitive skills and well being across the life course: Do noncognitive skills encourage healthy behavior? *American Economic Review* **100**(2) 200–204.

Christakis, Nicholas A., James H. Fowler. 2008. The collective dynamics of smoking in a large social network. *New England Journal of Medicine* **358**(21) 2249–2258, pMID: 18499567.

Clark, Andrew E., Youenn Lohéac. 2007. "it wasn't me, it was them!" social influence in risky behavior by adolescents. *Journal of Health Economics* **26**(4) 763 – 784.

Cobb-Clark, Deborah A., Sonja C. Kassenboehmer, Stefanie Schurer. 2014. Healthy habits: The connection between diet, exercise, and locus of control. *Journal of Economic Behavior & Organization* **98**(0) 1 – 28.

Cobb-Clark, Deborah A., Stefanie Schurer. 2012. The stability of big-five personality traits. *Economics Letters* **115**(1) 11 – 15.

Cohen, Geoffrey L., Mitchell J. Prinstein. 2006. Peer contagion of aggression and health risk behavior among adolescent males: An experimental investigation of effects on public conduct and private attitudes. *Child Development* **77**(4) 967–983.

Cohen-Cole, Ethan, Jason M Fletcher. 2008a. Detecting implausible social network effects in acne, height, and headaches: longitudinal analysis. *BMJ* **337**.

Cohen-Cole, Ethan, Jason M. Fletcher. 2008b. Is obesity contagious? social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics* **27**(5) 1382 – 1387.

Conrad, Karen M, Brian R Flay, David Hill. 1992. Why children start smoking cigarettes: predictors of onset. *British Journal of Addiction* **87**(12) 1711–1724.

Conti, Gabriella, Christopher Hansman. 2013. Personality and the education–health gradient: A note on "understanding differences in health behaviors by education". *Journal of Health Economics* **32**(2) 480 – 485.

Conti, Gabriella, James Heckman, Sergio Urzua. 2010. The education-health gradient. *American Economic Review* **100**(2) 234–38.

Costa, Paul T, Robert R McCrae. 1992. Normal personality assessment in clinical practice: The neo personality inventory. *Psychological assessment* **4**(1) 5.

Cutler, David M, Edward L Glaeser. 2010. Social interactions and smoking. *Research Findings in the Economics of Aging*. University of Chicago Press, 123–141.

Danaei, Goodarz, Eric L Ding, Dariush Mozaffarian, Ben Taylor, Jürgen Rehm, Christopher JL Murray, Majid Ezzati. 2009. The preventable causes of death in the united states: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS medicine* **6**(4) e1000058.

Deary, Ian J., Alexander Weiss, G. David Batty. 2010. Intelligence and personality as predictors of illness and death: How researchers in differential psychology and chronic disease epidemiology are collaborating to understand and address health inequalities. *Psychological Science in the Public Interest* **11**(2) 53–79.

DeCicca, Philip, Donald Kenkel, Alan Mathios. 2002. Putting out the fires: Will higher taxes reduce the onset of youth smoking? *Journal of Political Economy* **110**(1) 144–169.

DeCicca, Philip, Donald Kenkel, Alan Mathios, Yoon-Jeong Shin, Jae-Young Lim. 2008. Youth smoking, cigarette prices, and anti-smoking sentiment. *Health Economics* **17**(6) 733–749.

DHHS. 2012. Preventing tobacco use among youth and young adults: A report of the surgeon general. *Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health* **2**.

Digman, John M. 1990. Personality structure: Emergence of the five-factor model. *Annual review of psychology* **41**(1) 417–440.

Driscoll, Anne K, Stephen T Russell, Lisa J Crockett. 2008. Parenting styles and youth well-being across immigrant generations. *Journal of Family Issues* **29**(2) 185–209.

Eisenberg, Daniel. 2004. Peer effects for adolescent substance use: Do they really exist. *UC-Berkeley School of Public Health working paper* .

Eisenberg, Daniel, Ezra Golberstein, Janis L. Whitlock. 2014. Peer effects on risky behaviors: New evidence from college roommate assignments. *Journal of Health Economics* **33**(0) 126 – 138.

Fletcher, Jason M. 2010. Social interactions and smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects. *Health Economics* **19**(4) 466–484.

Fletcher, Jason M, Partha Deb, Jody L Sindelar. 2009. Tobacco use, taxation and self control in adolescence. Tech. rep., National Bureau of Economic Research Working Paper 15130.

Fortin, B., M. Yazbeck. 2015. Peer effects, fast food consumption and adolescent weight gain. *Journal of Health Economics* **forthcoming**.

Fowler, J.H., N.A. Christakis. 2008. Estimating peer effects on health in social networks: A response to cohen-cole and fletcher; and trogdon, nonnemaker, and pais. *Journal of Health Economics* **27**(5) 1400 – 1405.

Fruehwirth, Jane Cooley. 2014. Can achievement peer effect estimates inform policy? a view from inside the black box. *Review of Economics and Statistics* **96**(3) 514–523.

Gaviria, Alejandro, Steven Raphael. 2001. School-based peer effects and juvenile behavior. *Review of Economics and Statistics* **83**(2) 257–268.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, A. F. M. Smith, eds., *BAYESIAN STATISTICS*, vol. 4. Oxford University Press.

Glaeser, Edward L., Jose A. Scheinkman. 2003. Nonmarket interactions. M. Dewatripont, L. P. Hansen, S.J. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications*. Cambridge University Press, 339–369.

Goldsmith-Pinkham, Paul, Guido W. Imbens. 2013. Social networks and the identification of peer effects. *Journal of Business & Economic Statistics* **31**(3) 253–264.

Graham, Bryan S. 2015. Methods of identification in social networks. *Annual Review of Economics* **forthcoming**.

Graham, Bryan S, Guido W Imbens, Geert Ridder. 2010. Measuring the effects of segregation in the presence of social spillovers: a nonparametric approach. Tech. rep., National Bureau of Economic Research Working Paper w16499.

Gul, Faruk, Wolfgang Pesendorfer. 2004. Self-control and the theory of consumption. *Econometrica* **72**(1) 119–158.

Hampson, Sarah E, Lewis R Goldberg, Thomas M Vogt, Joan P Dubanoski. 2007. Mechanisms by which childhood personality traits influence adult health status: educational attainment and healthy behaviors. *Health psychology* **26**(1) 121.

Hampson, Sarah E., Elizabeth Tildesley, Judy A. Andrews, Koen Luyckx, Daniel K. Mroczek. 2010. The relation of change in hostility and sociability during childhood to substance use in mid adolescence. *Journal of Research in Personality* **44**(1) 103 – 114.

Handcock, Raftery A. E., M. S., J. M. Tantrum. 2007. Model based clustering for social networks. *Journal of the Royal Statistical Society: Series A* **170(2)**.

Heckman, James, Rodrigo Pinto, Peter Savelyev. 2013. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review* **103**(6) 2052–86.

Heckman, James J. 2000. Policies to foster human capital. *Research in economics* **54**(1) 3–56.

Heckman, James J., Fredrick Flyer, Colleen Loughlin. 2008. An assessment of causal inference in smoking initiation research and a framework for future research. *Economic Inquiry* **46**(1) 37–44.

Heckman, James J., Jora Stixrud, Sergio Urzua. 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics* **24**(3) pp. 411–482.

Hoff, Raftery A. E., P. D., M. S. Handcock. 2002. Latent space approaches to social network analysis. *Journal of the american Statistical association* **97(460)**.

Hsieh, Chih-Sheng, Lung-Fei Lee. 2015. A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics* .

Ichimura, Hidehiko. 1993. Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics* **58**(1) 71–120.

Kawaguchi, Daiji. 2004. Peer effects on substance use among american teenagers. *Journal of Population Economics* **17**(2) 351–367.

Kooreman, Peter. 2007. Time, money, peers, and parents; some data and theories on teenage behavior. *Journal of Population Economics* **20**(1) 9–33.

Krauth, B. V. 2006. Simulation-based estimation of peer effects. *Journal of Econometrics* **133(1)**.

Krauth, Brian V. 2007. Peer and selection effects on youth smoking in california. *Journal of Business & Economic Statistics* **25**(3) 288–298.

Lee, Lung-Fei. 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* **140**(2) 333 – 374.

Lee, Lung-Fei, Xiaodong Liu, Xu Lin. 2010. Specification and estimation of social interaction models with network structures. *Econometrics Journal* **13**(2) 145–176.

Lin, Xu. 2010. Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics* **28**(4) 825–860.

Lin, Xu, Bruce Weinberg. 2014. Unrequited friendship? how reciprocity mediates adolescent peer effects. *Regional Science and Urban Economics* **48** 144–153.

Liu, Xiaodong. 2013. Estimation of a local-aggregate network model with sampled networks. *Economics Letters* **118**(1) 243–246.

Liu, Xiaodong, Eleonora Patacchini, Yves Zenou. 2014. Endogenous peer effects: local aggregate or local average? *Journal of Economic Behavior & Organization* **103** 39–59.

Lundborg, Petter. 2006. Having the wrong friends? peer effects in adolescent substance use. *Journal of Health Economics* **25**(2) 214 – 233.

Lyons, Russell. 2011. The spread of evidence-poor medicine via flawed social-network analysis. *Statistics, Politics, and Policy* **2**(1).

Manski, Charles F. 1993. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* **60**(3) 531–542.

Matthews, Gerald, Ian J Deary, Martha C Whiteman. 2003. *Personality traits*. Cambridge University Press.

Mendolia, Silvia, Ian Walker. 2014. The effect of noncognitive traits on health behaviours in adolescense. *Health Economics* **23**(9) 1146–1158.

Moffitt, Robert A. 2000. Policy interventions, low-level equilibria and social interactions. *Social Dynamics*. MIT Press, 45–82.

Mokdad, A.H., J.S. Marks, D.F. Stroup, J.L. Gerberding. 2004. Actual causes of death in the united states, 2000. *JAMA* **291**(10) 1238–1245.

Nakajima, R. 2007. Measuring peer effects on youth smoking behaviour. *The Review of Economic Studies* **74(3)**.

Navarro, Salvador. 2008. Control functions. *The new Palgrave dictionary of economics* .

Norton, Edward C., Richard C. Lindrooth, Susan T. Ennett. 1998. Controlling for the endogeneity of peer substance use on adolescent alcohol and tobacco use. *Health Economics* **7**(5) 439–453.

OECD. 2013. Health at a glance 2013: Oecd indicators. *Organization of Economic Cooperation and Development* .

Powell, Lisa M., John A. Tauras, Hana Ross. 2005. The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior. *Journal of Health Economics* **24**(5) 950 – 968.

Raftery, A. E., S. M. Lewis. 1992. How many iterations in the gibbs sampler? J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, A. F. M. Smith, eds., *BAYESIAN STATISTICS*, vol. 4. Oxford University Press.

Raftery, Adrian E., Michael A. Newton, Jaya M. Satagopan, Pavel N. Krivitsky. 2007. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, M. West, eds., *BAYESIAN STATISTICS*, vol. 8. Oxford University Press, 1 – 45.

Roberts, Brent W, Daniel Mroczek. 2008. Personality trait change in adulthood. *Current directions in psychological science* **17**(1) 31–35.

Roberts, Brent W, Kate E Walton, Wolfgang Viechtbauer. 2006. Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological bulletin* **132**(1) 1.

Sacerdote, Bruce. 2001. Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics* **116**(2) 681–704.

Sacerdote, Bruce. 2011. Peer effects in education: How might they work, how big are they and how much do we know thus far? *Handbook of the Economics of Education* **3** 249–277.

Savelyev, Peter A. 2014. Socioemotional skills, education, and longevity of high-ability individuals. Tech. rep., Available at SSRN: http://ssrn.com/abstract=1715942.

Shakya, Holly B, Nicholas A Christakis, James H Fowler. 2012. Parental influence on substance use in adolescent social networks. *Archives of pediatrics & adolescent medicine* **166**(12) 1132–1139.

Soetevent, Adriaan R., Peter Kooreman. 2007. A discrete-choice model with social interactions: with an application to high school teen behavior. *Journal of Applied Econometrics* **22**(3) 599–624.

Thomson, Sir Godfrey Hilton. 1951. *The factorial analysis of human ability*. University of London Press London.

VanderWeele, Tyler J, Elizabeth L Ogburn, Eric J Tchetgen Tchetgen. 2012. Why and when "flawed" social network analyses still yield valid tests of no contagion. *Statistics, Politics, and Policy* **3**(1).

Young, J Kenneth, A Alexander Beaujean. 2011. Measuring personality in wave i of the national longitudinal study of adolescent health. *Frontiers in psychology* **2**.

Zeger, Scott L, M Rezaul Karim. 1991. Generalized linear models with random effects; a gibbs sampling approach. *Journal of the American statistical association* **86**(413) 79–86.

Zimmerman, David J. 2003. Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics* **85**(1) 9–23.

Tables and Figures

Table 1: Average number of nominated friends within and across personality groups

	Emotional Stability			Conscientiousness		
	Weak	Strong	Total	Weak	Strong	Total
Weak	0.5265	0.4977	1.0242	0.4606	0.5405	1.0011
Strong	0.4997	0.5501	1.0498	0.4711	0.5953	1.0664

Note: The above statistics are based on our network samples from the Add Health Wave I in-home survey.

Table 2: Add Health Survey Items and Factor Analysis for Personalities

Item identifier	Description	Factor loadings			Regression Coef.		
		ES	C	E	ES	C	E
H1PF30	You have a lot of good qualities	0.6207			0.3164		
H1PF32	You have a lot to be proud of	0.6602			0.3930		
H1PF33	You like yourself just the way you are	0.4145			0.0150		
H1PF34	You feel like you are doing everything just about right	0.3502			-0.0371		
H1PF35	You feel socially accepted	0.4191			0.0408		
H1PF36	You feel wanted and loved	0.4903			0.1137		
H1PF18	You get as many facts about the problem as possible when you have problems to be solved		0.6300			0.2778	
H1PF19	You think of as many different ways to approach a problem as possible when you are attempting to find a solution		0.6700			0.3195	
H1PF20	You generally use a systematic method for judging and comparing alternatives When making decisions		0.6245			0.2741	
H1PF21	You usually try to analyze what went right and what went wrong after carrying out a solution to a problem		0.5680			0.2295	
S62B*	I feel close to people at school			0.7014			0.3543
S62E*	I feel like I am a part of this school			0.7175			0.3786
S62O*	I feel socially accepted			0.6245			0.2696

Note: These 13 items are selected by the Lexical approach and the exploratory factor analysis according to Young and Beaujean (2011). We conduct a factor analysis on these items and identify one main factor for each personality measure, which explains more than 90% of variation. We report the factor loadings for each item after rotation and the regression coefficients for predicting factor scores. ES: emotional stability; C:conscientiousness; E: extroversion. $*$ denotes that data source is Wave I in-school survey.

Table 3: Summary Statistics for the Whole Sample and by Personality Measures

	Whole Sample				Emotional Stability				conscientiousness			
					strong		weak		strong		weak	
	average	s.d.	min	max	average	s.d.	average	s.d.	average	s.d.	average	s.d.
smoke	0.224	0.417	0.000	1.000	0.184	0.388	0.261	0.439	0.205	0.404	0.247	0.431
smoke frequency	0.904	2.162	0.000	7.000	0.721	1.955	1.070	2.321	0.832	2.095	0.989	2.236
drunk	0.311	0.463	0.000	1.000	0.272	0.445	0.346	0.476	0.296	0.457	0.327	0.469
emotional stability	0.015	0.715	-4.033	1.021	0.636	0.285	-0.547	0.485	0.146	0.670	-0.140	0.735
conscientiousness	-0.032	0.831	-3.790	1.565	0.188	0.877	-0.232	0.733	0.536	0.499	-0.701	0.619
extraversion	0.009	0.840	-2.385	1.260	0.193	0.806	-0.157	0.835	0.082	0.838	-0.077	0.833
male	0.466	0.499	0.000	1.000	0.517	0.500	0.420	0.494	0.475	0.499	0.456	0.498
white	0.541	0.498	0.000	1.000	0.538	0.499	0.544	0.498	0.530	0.499	0.554	0.497
black	0.228	0.419	0.000	1.000	0.258	0.438	0.200	0.400	0.239	0.427	0.215	0.411
asian	0.110	0.313	0.000	1.000	0.100	0.301	0.119	0.324	0.107	0.309	0.115	0.319
hisp	0.064	0.244	0.000	1.000	0.048	0.214	0.078	0.267	0.068	0.253	0.058	0.233
other race	0.057	0.232	0.000	1.000	0.055	0.229	0.059	0.235	0.056	0.230	0.059	0.235
school taught	0.934	0.248	0.000	1.000	0.942	0.233	0.927	0.261	0.940	0.238	0.928	0.259
smoke parent	0.643	0.479	0.000	1.000	0.632	0.482	0.652	0.476	0.634	0.482	0.654	0.476
prof	0.275	0.447	0.000	1.000	0.299	0.458	0.253	0.435	0.278	0.448	0.272	0.445
home	0.134	0.341	0.000	1.000	0.124	0.330	0.143	0.350	0.138	0.345	0.130	0.337
nonprof	0.427	0.495	0.000	1.000	0.416	0.493	0.437	0.496	0.428	0.495	0.426	0.495
low parent control	0.741	0.217	0.000	1.000	0.738	0.217	0.744	0.217	0.741	0.219	0.742	0.215
maternal care	4.550	0.526	1.000	5.000	4.627	0.485	4.481	0.552	4.596	0.500	4.497	0.550
sample size	9728				4619		5109		5258		4470	

Note: Strong (weak) personality refers to individuals' personality index which is above (below) the mean. "school taught" means the consequence of smoking is taught in school. "smoke parent" means either resident father or mother has ever smoked at home. "prof" means resident mother works as a professional (response 1 to 3 in Add Health survey item H1rm4), "home" indicates resident mother does not work (response 16 in H1rm4). "nonprof" indicates resident mother works in other categories (responses 4 to 14 in H1rm4). The omitted group for resident mother's occupation is the response 15 (other jobs) in H1rm4. "low parent control" reflects the degree to which your parents let you make your own decisions and is constructed by the average of items from H1WP1 to H1WP7. "maternal care" reflects how much you think your mother cares about you and is constructed by the average of items from h1wp9 to h1wp12.

Table 4: Placebo Test for Examining Peer effects on Father's Education

	SAR model			SCSAR model				
				SCSAR(1)	SCSAR(2)	SCSAR(3)	SCSAR(4)	SCSAR(5)
Endogenous effect								
	0.0419*	0.0187*	0.0160*	0.0160*	0.0152*	0.0139	0.0118	0.0046
	(0.0069)	(0.0047)	(0.0073)	(0.0073)	(0.0074)	(0.0074)	(0.0075)	(0.0090)
Own effect								
low parent control	−0.0180	−0.0243	−0.0229	−0.0227	−0.0241	−0.0232	−0.0227	−0.0230
	(0.0153)	(0.0169)	(0.0170)	(0.0170)	(0.0167)	(0.0169)	(0.0167)	(0.0167)
maternal care	0.0117*	0.0300*	0.0299*	0.0300*	0.0293*	0.0292*	0.0295*	0.0291*
	(0.0031)	(0.0067)	(0.0068)	(0.0065)	(0.0069)	(0.0067)	(0.0067)	(0.0068)
male	0.0178*	0.0211*	0.0205*	0.0206*	0.0206*	0.0204*	0.0204*	0.0202*
	(0.0073)	(0.0071)	(0.0073)	(0.0073)	(0.0074)	(0.0073)	(0.0072)	(0.0072)
black	−0.0816*	−0.0621*	−0.0596*	−0.0591*	−0.0600*	−0.0597*	−0.0597*	−0.0599*
	(0.0102)	(0.0116)	(0.0121)	(0.0121)	(0.0123)	(0.0120)	(0.0124)	(0.0121)
hisp	−0.0937*	−0.0947*	−0.0915*	−0.0908*	−0.0911*	−0.0914*	−0.0912*	−0.0916*
	(0.0122)	(0.0135)	(0.0135)	(0.0137)	(0.0135)	(0.0136)	(0.0134)	(0.0134)
asian	0.0199	0.0209	0.0176	0.0184	0.0177	0.0169	0.0177	0.0171
	(0.0158)	(0.0170)	(0.0174)	(0.0176)	(0.0175)	(0.0175)	(0.0172)	(0.0174)
other race	−0.0373*	−0.0385*	−0.0364*	−0.0361*	−0.0374*	−0.0363*	−0.0372*	−0.0360*
	(0.0160)	(0.0160)	(0.0162)	(0.0160)	(0.0166)	(0.0162)	(0.0159)	(0.0163)
prof	0.6090*	0.5854*	0.5846*	0.5846*	0.5839*	0.5844*	0.5848*	0.5839*
	(0.0110)	(0.0112)	(0.0112)	(0.0110)	(0.0112)	(0.0110)	(0.0113)	(0.0114)
home	0.1786*	0.1955*	0.1965*	0.1959*	0.1954*	0.1960*	0.1955*	0.1964*
	(0.0205)	(0.0206)	(0.0205)	(0.0207)	(0.0209)	(0.0204)	(0.0208)	(0.0207)
nonprof	0.2791*	0.2820*	0.2825*	0.2824*	0.2823*	0.2829*	0.2827*	0.2823*
	(0.0081)	(0.0083)	(0.0082)	(0.0082)	(0.0081)	(0.0082)	(0.0083)	(0.0081)
mom edu	0.2419*	0.2211*	0.2207*	0.2205*	0.2209*	0.2208*	0.2198*	0.2194*
	(0.0076)	(0.0077)	(0.0077)	(0.0078)	(0.0076)	(0.0076)	(0.0076)	(0.0078)
σ_ϵ^2	0.1446	0.1413	0.1412	0.1412	0.1408	0.1409	0.1402	0.1367
Contextual effect	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effect	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: We report the posterior mean of each parameter and the standard deviation in the parenthesis. The asterisk indicates that its 95% highest posterior density range does not cover zero. The MCMC sampling is running for 250,000 iterations with the first 50,000 iterations dropped for burn-in. All cases pass the convergence diagnostics provided by Geweke (1992) and Raftery and Lewis (1992).

Table 5: Peer Effects on Smoking – SAR models with both Networks and Personalities assumed Exogenous

	Homogeneous		Emotional Stability	Conscientiousness
Endogenous effect				
	0.0922*	0.0842*		
	(0.0074)	(0.0077)		
weak-to-weak			0.1119*	0.0914*
			(0.0127)	(0.0156)
strong-to-weak			0.0614*	0.0866*
			(0.0188)	(0.0175)
weak-to-strong			0.0802*	0.0807*
			(0.0175)	(0.0176)
strong-to-strong			0.0700*	0.0802*
			(0.0171)	(0.0137)
Own effect				
emotional stability		−0.0268*	−0.0256*	−0.0276*
		(0.0064)	(0.0066)	(0.0063)
conscientiousness		−0.0170*	−0.0170*	−0.0161*
		(0.0051)	(0.0051)	(0.0053)
extraversion		−0.0476*	−0.0477*	−0.0475*
		(0.0051)	(0.0051)	(0.0051)
parent smoke	0.0760*	0.0721*	0.0727*	0.0724*
	(0.0087)	(0.0086)	(0.0085)	(0.0085)
low parent control	0.1127*	0.1142*	0.1164*	0.1157*
	(0.0198)	(0.0197)	(0.0194)	(0.0195)
maternal care	−0.0555*	−0.0366*	−0.0349*	−0.0348*
	(0.0077)	(0.0079)	(0.0076)	(0.0077)
school taught	−0.0118	−0.0039	−0.0023	−0.0026
	(0.0167)	(0.0166)	(0.0165)	(0.0166)
male	−0.0031	0.0067	0.0068	0.0069
	(0.0083)	(0.0083)	(0.0083)	(0.0083)
black	−0.1203*	−0.1213*	−0.1210*	−0.1208*
	(0.0139)	(0.0139)	(0.0138)	(0.0138)
hisp	−0.007	−0.0110	−0.0105	−0.0108
	(0.0155)	(0.0155)	(0.0155)	(0.0155)
asian	−0.0565*	−0.0695*	−0.0692*	−0.0692*
	(0.0197)	(0.0195)	(0.0195)	(0.0195)
other race	0.0384	0.0306	0.0312	0.0310
	(0.0183)	(0.0180)	(0.0181)	(0.0181)
prof	0.0124	0.0155	0.0158	0.0162
	(0.0129)	(0.0128)	(0.0128)	(0.0128)
home	−0.0042	−0.0056	−0.0049	−0.0047
	(0.0152)	(0.0152)	(0.0151)	(0.0151)
nonprof	0.0291*	0.0284*	0.0286*	0.0290*
	(0.0119)	(0.0118)	(0.0119)	(0.0119)
Contextual effect	Yes	Yes	Yes	Yes
Group fixed effect	Yes	Yes	Yes	Yes
σ_ϵ^2	0.1577	0.1551	0.1550	0.1551

Note: We report the posterior mean of each parameter and the standard deviation in the parenthesis. The value with asterisk means its 95% highest posterior density range does not cover zero. The MCMC sampling is running for 50,000 iterations with the first 5,000 iterations dropped for burn-in. All cases pass the convergence diagnostics provided by Geweke (1992) and Raftery and Lewis (1992). In the heterogenous peer effect case, A-to-B denotes the peer effect that B receives from A.

Table 6: Peer Effects on Smoking – SCSAR Models with Endogenous Networks and Exogenous Personality

	Homogeneous		Emotional Stability		Conscientiousness	
	SCSAR(4)	SCSAR(5)	SCSAR(4)	SCSAR(5)	SCSAR(4)	SCSAR(5)
Endogenous Effect						
	0.0374*	0.0328*				
	(0.0098)	(0.0113)				
weak-to-weak			0.0668*	0.0586*	0.0442*	0.0439*
			(0.0147)	(0.0148)	(0.0174)	(0.0173)
strong-to-weak			0.0247	0.0128	0.0382*	0.0355
			(0.0191)	(0.0196)	(0.0181)	(0.0181)
weak-to-strong			0.0316	0.0226	0.0357*	0.0352
			(0.0186)	(0.0181)	(0.0180)	(0.0182)
strong-to-strong			0.0360*	0.0204	0.0322*	0.0333*
			(0.0183)	(0.0177)	(0.0152)	(0.0159)
Own and Contextual effect	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
σ_u^2	0.1341	0.1315	0.1344	0.1310	0.1343	0.1324
Link formation						
constant	1.0340*	2.0975*	1.0395*	2.1011*	1.0333*	2.0972*
	(0.0789)	(0.0939)	(0.0799)	(0.0916)	(0.0788)	(0.0939)
grade	2.7500*	2.8303*	2.7533*	2.8290*	2.7553*	2.8331*
	(0.0380)	(0.0408)	(0.0392)	(0.0418)	(0.0390)	(0.0415)
sex	0.3511*	0.3591*	0.3498*	0.3565*	0.3506*	0.3588*
	(0.0342)	(0.0338)	(0.0348)	(0.0353)	(0.0335)	(0.0345)
race	1.1213*	1.1911*	1.1216*	1.1863*	1.1190*	1.1961*
	(0.0415)	(0.0432)	(0.0413)	(0.0409)	(0.0403)	(0.0445)
emotional stability	−0.0563	−0.0606	−0.0582	−0.0603*	−0.0583	−0.0591
	(0.0291)	(0.0303)	(0.0291)	(0.0297)	(0.0301)	(0.0310)
conscientiousness	−0.041	−0.0399	−0.0400	−0.0381	−0.0397	−0.0411
	(0.0255)	(0.0269)	(0.0255)	(0.0274)	(0.0256)	(0.0258)
extraversion	−0.2636*	−0.2734*	−0.2654*	−0.2732*	−0.2626*	−0.2707*
	(0.0277)	(0.0297)	(0.0269)	(0.0287)	(0.0274)	(0.0267)
δ_1	−3.4773*	−3.3539*	−3.5842*	−3.3043*	−3.4545*	−3.1596*
	(0.1152)	(0.3992)	(0.1443)	(0.2731)	(0.1417)	(0.1472)
δ_2	−3.2812*	−2.7606*	−3.2175*	−2.7642*	−3.2042*	−2.8837*
	(0.0909)	(0.1306)	(0.1220)	(0.1064)	(0.1027)	(0.1213)
δ_3	−3.0676*	−2.5773*	−2.9778*	−2.5906*	−3.0365*	−2.6648*
	(0.1213)	(0.1195)	(0.1015)	(0.1235)	(0.0933)	(0.1084)
δ_4	−2.7491*	−2.4316*	−2.8179*	−2.4266*	−2.8567*	−2.4600*
	(0.1542)	(0.1126)	(0.0926)	(0.0961)	(0.1115)	(0.1255)
δ_5		−2.2726*		−2.2867*		−2.2102*
		(0.1364)		(0.1055)		(0.1331)
AICM	116060	129340	118330	129220	121290	132450
se(AICM)	3749	5029	3887	4630	4063	4821

Note: We report the posterior mean of each parameter and the standard deviation in the parenthesis. The asterisk indicates that its 95% highest posterior density range does not cover zero. The MCMC sampling is running for 250,000 iterations with the first 100,000 iterations dropped for burn-in. All cases pass the convergence diagnostics provided by Geweke (1992) and Raftery and Lewis (1992). In the heterogenous peer effect case, A-to-B denotes the peer effect that B receives from A.

Table 7: Peer Effects on Smoking – SCSAR Models with both Endogenous Networks and Personality (Specification based on Eq.(6))

	Homogeneous		Emotional Stability		Conscientiousness	
	SCSAR(4)	SCSAR(5)	SCSAR(4)	SCSAR(5)	SCSAR(4)	SCSAR(5)
Smoking						
Endogenous Effect						
	0.0689*	0.0645*				
	(0.0086)	(0.0087)				
weak-to-weak			0.0720*	0.0682*	0.0482*	0.0462*
			(0.0154)	(0.0141)	(0.0171)	(0.0174)
strong-to-weak			0.0214	0.0187	0.0406*	0.0384*
			(0.0204)	(0.0195)	(0.0172)	(0.0172)
weak-to-strong			0.0389*	0.0321	0.0439*	0.0394*
			(0.0209)	(0.0187)	(0.0185)	(0.0180)
strong-to-strong			0.0324	0.0294	0.0394*	0.0356*
			(0.0189)	(0.0172)	(0.0145)	(0.0146)
Own & Contextual effects	Yes	Yes	Yes	Yes	Yes	Yes
σ_u^2	0.1406	0.1419	0.1345	0.1341	0.1346	0.1348
Emotional Stability						
maternal care	0.2505*	0.2517*	0.2459*	0.2485*	0.2478*	0.2482*
	(0.0130)	(0.0125)	(0.0131)	(0.0127)	(0.0131)	(0.0133)
male	0.1840*	0.1833*	0.1803*	0.1837*	0.1822*	0.1812*
	(0.0141)	(0.0135)	(0.0136)	(0.0135)	(0.0139)	(0.0137)
black	0.0839*	0.0894*	0.0877*	0.0834*	0.0884*	0.0844*
	(0.0230)	(0.0220)	(0.0223)	(0.0226)	(0.0233)	(0.0226)
hisp	−0.0664*	−0.0635*	−0.0651*	−0.0675*	−0.0655*	−0.0638*
	(0.0261)	(0.0263)	(0.0259)	(0.0260)	(0.0260)	(0.0267)
asian	−0.2047*	−0.2123*	−0.2024*	−0.2108*	−0.1994*	−0.2134*
	(0.0326)	(0.0331)	(0.0345)	(0.0323)	(0.0332)	(0.0328)
other race	−0.0393	−0.0530	−0.0432	−0.0439	−0.0410	−0.0465
	(0.0308)	(0.0307)	(0.0310)	(0.0298)	(0.0304)	(0.0319)
σ_v^2	0.3591	0.2830	0.3560	0.2396	0.3574	0.2585
Conscientiousness						
maternal care	0.1618*	0.1635*	0.1594*	0.1641*	0.1582*	0.1636*
	(0.0154)	(0.0153)	(0.0161)	(0.0157)	(0.0157)	(0.0153)
male	0.0416*	0.0402*	0.0386*	0.0410*	0.0395*	0.0384*
	(0.0164)	(0.0169)	(0.0165)	(0.0165)	(0.0160)	(0.0162)
black	0.0578*	0.0643*	0.0657*	0.0597*	0.0610*	0.0601*
	(0.0274)	(0.0271)	(0.0276)	(0.0275)	(0.0276)	(0.0277)
hisp	−0.0600	−0.0577	−0.0558	−0.0597	−0.0595	−0.0607
	(0.0314)	(0.0304)	(0.0315)	(0.0325)	(0.0308)	(0.0311)
asian	0.0465	0.0393	0.0484	0.0388	0.0495	0.0371
	(0.0394)	(0.0389)	(0.0410)	(0.0392)	(0.0399)	(0.0400)
other race	−0.0296	−0.0394	−0.0293	−0.0317	−0.0301	−0.0335
	(0.0369)	(0.0371)	(0.0370)	(0.0372)	(0.0369)	(0.0368)
σ_v^2	0.5749	0.5405	0.5574	0.5452	0.5543	0.5196
Extraversion						
maternal care	0.1813*	0.1855*	0.1779*	0.1790*	0.1775*	0.1783*
	(0.0160)	(0.0159)	(0.0156)	(0.0154)	(0.0158)	(0.0154)
male	0.0689*	0.0687*	0.0653*	0.0678*	0.0676*	0.0655*
	(0.0166)	(0.0167)	(0.0167)	(0.0171)	(0.0168)	(0.0165)
black	−0.1066*	−0.1040*	−0.1041*	−0.1047*	−0.1015*	−0.1057*
	(0.0275)	(0.0271)	(0.0271)	(0.0274)	(0.0286)	(0.0272)
hisp	−0.0225	−0.0205	−0.0217	−0.0237	−0.0192	−0.0206
	(0.0309)	(0.0320)	(0.0323)	(0.0316)	(0.0312)	(0.0319)
asian	−0.1592*	−0.1689*	−0.1541*	−0.1630*	−0.1512*	−0.1655*
	(0.0398)	(0.0406)	(0.0400)	(0.0390)	(0.0405)	(0.0406)
other race	−0.1346*	−0.1482*	−0.1360*	−0.1380*	−0.1362*	−0.1366*
	(0.0373)	(0.0377)	(0.0366)	(0.0365)	(0.0358)	(0.0372)

Table – Continued

σ_v^2	0.5491	0.5537	0.5330	0.5314	0.5282	0.5243
Link formation						
constant	0.4312*	1.0996*	0.4476*	1.1780*	0.4521*	1.2231*
	(0.0821)	(0.0834)	(0.0757)	(0.0848)	(0.0863)	(0.0873)
grade	2.7328*	2.7652*	2.7336*	2.7750*	2.7256*	2.7864*
	(0.0387)	(0.0426)	(0.0359)	(0.0406)	(0.0385)	(0.0420)
sex	0.3625*	0.3665*	0.3662*	0.3626*	0.3636*	0.3564*
	(0.0329)	(0.0316)	(0.0334)	(0.0344)	(0.0337)	(0.0340)
race	1.1036*	1.1603*	1.1133*	1.1466*	1.1231*	1.1462*
	(0.0381)	(0.0386)	(0.0405)	(0.0418)	(0.0404)	(0.0415)
emotional stability	0.1828*	0.1713*	0.2051*	0.2259*	0.1981*	0.2230*
	(0.0340)	(0.0331)	(0.0350)	(0.0489)	(0.0363)	(0.0479)
conscientiousness	0.0816*	0.0564	0.1103*	0.0331	0.1154*	0.0753*
	(0.0279)	(0.0271)	(0.0274)	(0.0313)	(0.0296)	(0.0334)
extraversion	−0.0829*	−0.1735*	−0.0567	−0.1164*	−0.0482	−0.1030*
	(0.0318)	(0.0265)	(0.0330)	(0.0334)	(0.0331)	(0.0368)
δ_1	−4.2009*	−3.3621*	−4.6450*	−3.5463*	−4.5735*	−3.5853*
	(0.1764)	(0.0799)	(0.2689)	(0.1425)	(0.1547)	(0.1767)
δ_2	−2.9791*	−3.1517*	−2.6994*	−3.2085*	−2.7141*	−3.1837*
	(0.0865)	(0.0900)	(0.0714)	(0.1006)	(0.0569)	(0.1406)
δ_3	−2.5722*	−2.9442*	−2.6704*	−2.9966*	−2.6942*	−2.9729*
	(0.1128)	(0.0843)	(0.0721)	(0.0966)	(0.0552)	(0.1260)
δ_4	−2.5441*	−2.6709*	−2.6318*	−2.7680*	−2.6548*	−2.7775*
	(0.1144)	(0.0934)	(0.0730)	(0.0919)	(0.0616)	(0.1281)
δ_5		−0.5845*		−0.5872*		−0.6926*
		(0.0493)		(0.0719)		(0.0866)
AICM	191220	310990	195390	300200	181760	283000
se(AICM)	4470	11923	5908	14049	4908	12654

Note: We report the posterior mean of each parameter and the standard deviation in the parenthesis. The asterisk indicates that its 95% highest posterior density range does not cover zero. The MCMC sampling is running for 250,000 iterations with the first 100,000 iterations dropped for burn-in. All cases pass the convergence diagnostics provided by Geweke (1992) and Raftery and Lewis (1992). Group fixed effects are included in all equations for smoking and personalities. In equations for personalities, we also include *weak parent control*, *prof*, *home*, *other job* as explanatory variables but their coefficients are not significant and therefore not reported in this table.

Table 8: Peer Effects on Smoking – SCSAR Models with both Endogenous Networks and Personality (Specification based on Eq.(7))

	Homogeneous		Emotional Stability		Conscientiousness	
	SCSAR(4)	SCSAR(5)	SCSAR(4)	SCSAR(5)	SCSAR(4)	SCSAR(5)
Smoking						
Endogenous Effect						
	0.0514*	0.0433*				
	(0.0088)	(0.0098)				
weak-to-weak			0.0833*	0.0732*	0.0628*	0.0477*
			(0.0137)	(0.0140)	(0.0165)	(0.0171)
strong-to-weak			0.0289	0.0156	0.0535*	0.0408*
			(0.0196)	(0.0198)	(0.0175)	(0.0182)
weak-to-strong			0.0529*	0.0454*	0.0545*	0.0403*
			(0.0181)	(0.0179)	(0.0177)	(0.0187)
strong-to-strong			0.0450*	0.0328	0.0520*	0.0403*
			(0.0178)	(0.0180)	(0.0149)	(0.0161)
σ_u^2	0.1381	0.1328	0.1390	0.1349	0.1389	0.1323
Emotional Stability						
Endogenous Effect						
	−0.0239	−0.0205	−0.0251*	−0.0161	−0.0221	−0.0230
	(0.0128)	(0.0134)	(0.0108)	(0.0138)	(0.0118)	(0.0142)
σ_v^2	0.3076	0.2795	0.3067	0.2762	0.3111	0.2676
Conscientiousness						
Endogenous Effect						
	−0.0135	−0.0101	−0.0077	−0.0133	−0.0071	−0.0063
	(0.0094)	(0.0116)	(0.0095)	(0.0111)	(0.0111)	(0.0106)
σ_v^2	0.5265	0.5276	0.5320	0.5238	0.5298	0.5330
Extraversion						
Endogenous Effect						
	0.0315*	0.0266*	0.0286*	0.0273*	0.0330*	0.0271*
	(0.0092)	(0.0106)	(0.0091)	(0.0101)	(0.0088)	(0.0105)
σ_v^2	0.5038	0.4910	0.5145	0.4947	0.5006	0.4932
Link formation						
constant	0.4892*	1.4371*	0.5058*	1.3709*	0.5099*	1.3661*
	(0.0776)	(0.0878)	(0.0759)	(0.1028)	(0.0708)	(0.1110)
grade	2.7074*	2.7905*	2.7040*	2.7811*	2.7036*	2.7819*
	(0.0382)	(0.0396)	(0.0369)	(0.0408)	(0.0378)	(0.0411)
sex	0.3510*	0.3536*	0.3541*	0.3519*	0.3540*	0.3500*
	(0.0322)	(0.0350)	(0.0321)	(0.0348)	(0.0326)	(0.0337)
race	1.0909*	1.1496*	1.0923*	1.1465*	1.0901*	1.1468*
	(0.0419)	(0.0421)	(0.0412)	(0.0424)	(0.0413)	(0.0429)
δ_1	−4.0791*	−3.6291*	−4.1871*	−3.5618*	−4.0638*	−3.4320*
	(0.1539)	(0.2158)	(0.1725)	(0.1956)	(0.1951)	(0.1277)
δ_2	−3.8302*	−3.1710*	−3.6956*	−3.2083*	−3.7240*	−3.1881*
	(0.1244)	(0.1315)	(0.1252)	(0.1233)	(0.1147)	(0.0929)
δ_3	−3.4747*	−2.9346*	−3.5022*	−2.9786*	−3.5276*	−3.0348*
	(0.1443)	(0.1030)	(0.1338)	(0.0980)	(0.1284)	(0.0896)
δ_4	−1.2659*	−2.7293*	−1.2913*	−2.7609*	−1.3238*	−2.8207*
	(0.0942)	(0.1174)	(0.0813)	(0.1310)	(0.0813)	(0.1180)
δ_5		−0.8252*		−0.7083*		0.7103*
		(0.0800)		(0.1060)		(0.1287)
AICM	195310	240050	192690	256840	179850	241750
se(AICM)	4943	7826	4768	8837	3998	7942

Note: We report the posterior mean of each parameter and the standard deviation in the parenthesis. The asterisk indicates that its 95% highest posterior density range does not cover zero. The MCMC sampling is running for 250,000 iterations with the first 100,000 iterations dropped for burn-in. All cases pass the convergence diagnostics provided by Geweke (1992) and Raftery and Lewis (1992). Own and contextual (observed and latent) variables and group fixed effects are included in all equations for smoking and personalities.

Table 9: Peer Effects on Smoking – results based on Add Health Saturated sample

	Homogeneous		Emotional Stability		Conscientiousness	
	SCSAR(4)	SCSAR(5)	SCSAR(4)	SCSAR(5)	SCSAR(4)	SCSAR(5)
Smoking						
Endogenous effect	0.0182	0.0106				
	(0.0197)	(0.0197)				
weak-to-weak			0.0923^{*}	0.0916^{*}	0.0014	-0.0042
			(0.0267)	(0.0281)	(0.0386)	(0.0395)
strong-to-weak			-0.0525	-0.0482	-0.0101	-0.0140
			(0.0392)	(0.0379)	(0.0331)	(0.0339)
weak-to-strong			-0.0048	-0.0043	0.0403	0.0337
			(0.0361)	(0.0363)	(0.0371)	(0.0377)
strong-to-strong			0.0196	0.0166	0.0463	0.0454
			(0.0371)	(0.0364)	(0.0275)	(0.0282)
Own & Contextual effects	Yes	Yes	Yes	Yes	Yes	Yes
Endogenous link formation	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
σ_u^2	0.1012	0.0970	0.0996	0.0956	0.1010	0.0978
AICM	15776	17604	16132	17122	15452	16729
se(AICM)	340	468	363	440	321	417

Note: We report the posterior mean of each parameter and the standard deviation in the parenthesis. The asterisk indicates that its 95% highest posterior density range does not cover zero. The MCMC sampling is running for 250,000 iterations with the first 100,000 iterations dropped for burn-in. All cases pass the convergence diagnostics provided by Geweke (1992) and Raftery and Lewis (1992). The saturated sample contains 13 schools and 705 students.

Table 10: Peer Effects on Smoking Frequency

	Homogeneous		Emotional Stability		Conscientiousness	
	SCSAR(4)	SCSAR(5)	SCSAR(4)	SCSAR(5)	SCSAR(4)	SCSAR(5)
Smoking frequency						
Endogenous effect	0.0731*	0.0666*				
	(0.0088)	(0.0099)				
weak-to-weak			0.1000*	0.0952*	0.0738*	0.0707*
			(0.0152)	(0.0153)	(0.0168)	(0.0169)
strong-to-weak			0.0969*	0.0943*	0.0757*	0.0686*
			(0.0194)	(0.0203)	(0.0187)	(0.0183)
weak-to-strong			0.0436*	0.0393*	0.0745*	0.0748*
			(0.0171)	(0.0170)	(0.0199)	(0.0196)
strong-to-strong			0.0411*	0.0377*	0.0687*	0.0625*
			(0.0158)	(0.0165)	(0.0142)	(0.0144)
Own & Contextual effects	Yes	Yes	Yes	Yes	Yes	Yes
Endogenous link formation	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
σ_u^2	3.6027	3.5359	3.5977	3.5505	3.5872	3.5443
AICM	156510	159580	151350	155520	148710	158470
se(AICM)	4258	4526	3949	4262	3788	4459

Note: We report the posterior mean of each parameter and the standard deviation in the parenthesis. The asterisk indicates that its 95% highest posterior density range does not cover zero. The MCMC sampling is running for 250,000 iterations with the first 100,000 iterations dropped for burn-in. All cases pass the convergence diagnostics provided by Geweke (1992) and Raftery and Lewis (1992). The dependent variable of smoking frequency is a continuous variable which takes value between 0 and 7 days per week on which the student smokes.

Table 11: Peer Effects on Getting Drunk

	Homogeneous		Emotional Stability		Conscientiousness	
	SCSAR(4)	SCSAR(5)	SCSAR(4)	SCSAR(5)	SCSAR(4)	SCSAR(5)
Getting drunk						
Endogenous effect	0.0409*	0.0309*				
	(0.0136)	(0.0107)				
weak-to-weak			0.0553*	0.0423*	0.0556*	0.0379*
			(0.0167)	(0.0176)	(0.0181)	(0.0174)
strong-to-weak			0.0254	0.0125	0.0486*	0.0262
			(0.0183)	(0.0206)	(0.0180)	(0.0171)
weak-to-strong			0.0421	0.0284	0.0269	0.0073
			(0.0185)	(0.0198)	(0.0187)	(0.0181)
strong-to-strong			0.0291	0.0147	0.0516*	0.0326*
			(0.0179)	(0.0192)	(0.0152)	(0.0149)
Own effect						
emotional stability	−0.0343*	−0.0341*	−0.0327*	−0.0330*	−0.0336*	−0.0348*
	(0.0070)	(0.0070)	(0.0073)	(0.0075)	(0.0070)	(0.0071)
conscientiousness	−0.0139*	−0.0134*	−0.0137*	−0.0138*	−0.0122*	−0.0116*
	(0.0057)	(0.0056)	(0.0057)	(0.0057)	(0.0059)	(0.0060)
extraversion	−0.0248*	−0.0257*	−0.0252*	−0.0246*	−0.0249*	−0.0253*
	(0.0056)	(0.0056)	(0.0058)	(0.0056)	(0.0057)	(0.0056)
Other own & contextual effects	Yes	Yes	Yes	Yes	Yes	Yes
Endogenous link formation	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
σ_u^2	0.1568	0.1535	0.1568	0.1513	0.1581	0.1505
AICM	132700	143330	143640	172650	129460	136200
se(AICM)	4670	5358	5335	7180	4470	4981

Note: We report the posterior mean of each parameter and the standard deviation in the parenthesis. The asterisk indicates that its 95% highest posterior density range does not cover zero. The MCMC sampling is running for 250,000 iterations with the first 100,000 iterations dropped for burn-in. All cases pass the convergence diagnostics provided by Geweke (1992) and Raftery and Lewis (1992). The dependent variable is a binary variable indicating whether student has been drunk over the past twelve months.

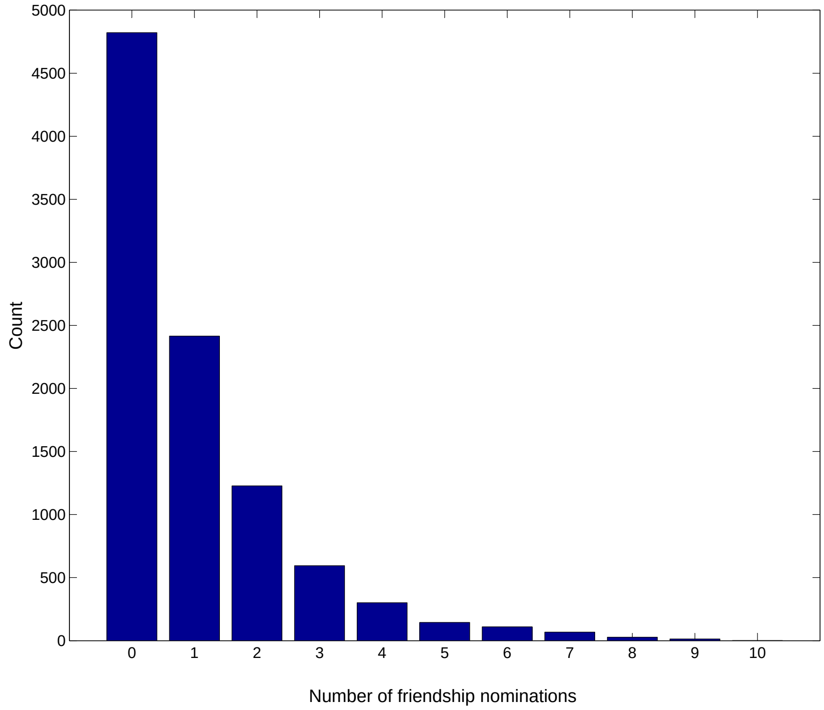

Figure 1: Distribution of friendship nominations based on our network samples from the Add Health Wave I in-home survey

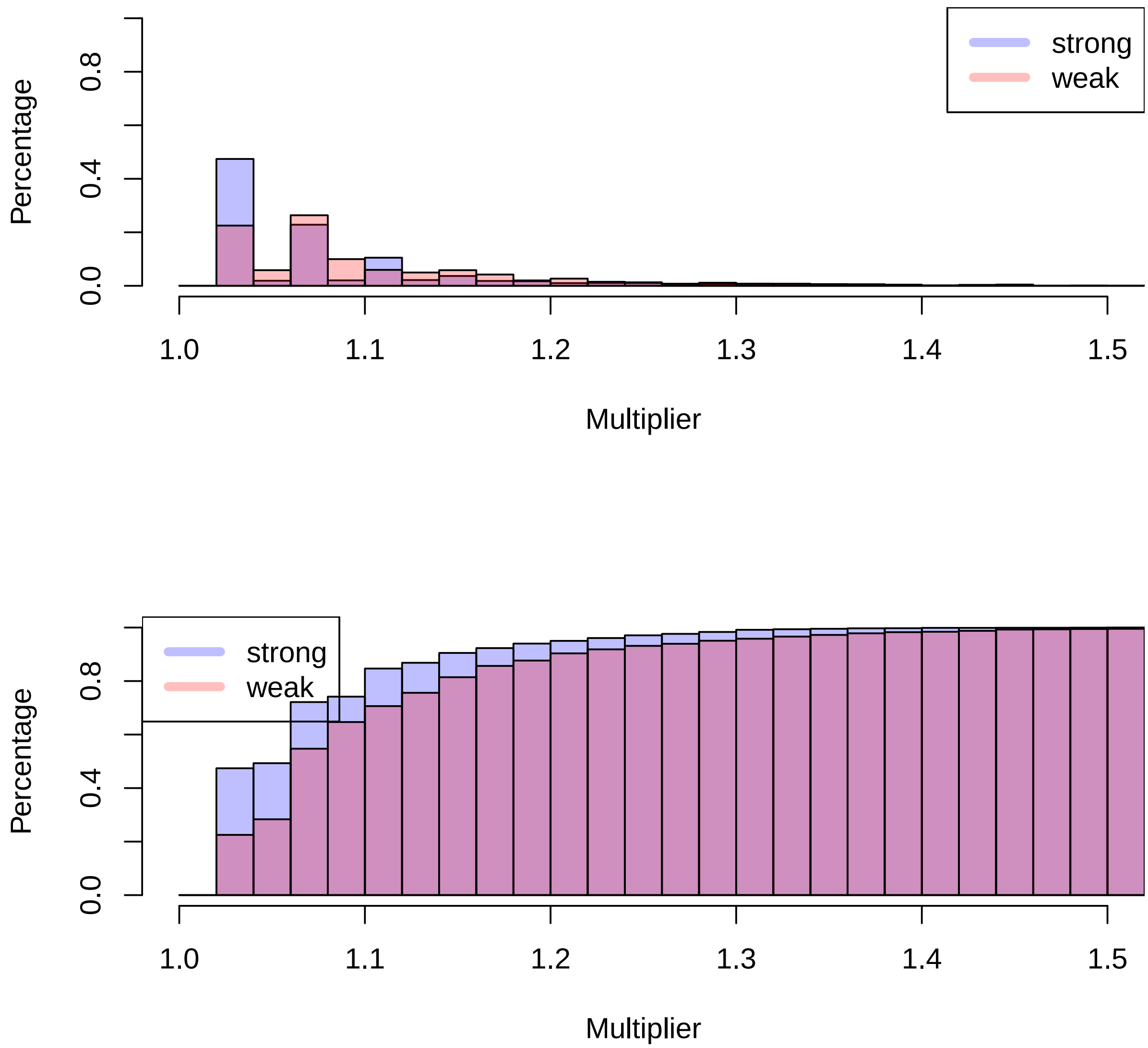

Figure 2: Distributions of multiplier effects on smoking from weak and strong emotional stability groups. Students without friendship links are excluded.

A Identification of the extended SC-SAR model

In this appendix we will discuss and justify the parametric assumptions on the individual latent variables $z_{i,g}$ that we impose when estimating the extended SC-SAR model. Apart from the model specifications of the link formation in Eq. (4) and the outcome in Eq. (5), we need the following assumptions: (1) the variance of $z_{i,g}$ is normalized to one; (2) if $z_{i,g}$ is multidimensional, different dimensions should be independent of each other; (3) $z_{i,g}$ follows a known distribution, which is assumed to be normal; (4) the magnitude of the homophily coefficients of $z_{i,g}$ in Eq. (4) follow a descending order, i.e., $|\eta_1| \geq |\eta_2| \geq \cdots \geq |\eta_{\bar{d}}|$.

Let us start with the least amount of parametric assumptions. Assume in the link formation of Equation (4), we only have the specification of $\psi_{ij,g}$ and no distributional assumption on the error term. This link formation equation, which is similar to a standard dichotomous choice model, can be motivated from the behavior of utility maximization. For each individual i, he/she chooses $w_{ij,g} = 1$ if $v_{ij,g}(w_{ij,g} = 1) - v_{ij,g}(w_{ij,g} = 0) > 0$ and $w_{ij,g} = 0$ otherwise, where $v_{ij,g}$ stands for i's utility function from the link ij. We can express the above utility difference as

$$v_{ij,g}(w_{ij,g} = 1) - v_{ij,g}(w_{ij,g} = 0) = \mu_{ij,g}(C_g, S_g, \gamma) + \xi_{ij,g}(Z_g, \eta), \tag{A.1}$$

where the deterministic component $\mu_{ij,g}(C_g, S_g, \gamma)$ contains $c_{ij,g}\gamma_0 + \sum_{r=1}^{R} \gamma_r |s_{ir,g} - s_{jr,g}|$, the observed exogenous dyad-specific variables $c_{ij,g}$ and personalities $s_{ir,g}$. The error component $\xi_{ij,g}(Z_g, \eta)$ contains $\sum_{d=1}^{\bar{d}} \eta_d |z_{id,g} - z_{jd,g}|$ to capture homophily on basis of the latent variables $z_{i,g}$, and $\zeta_{ij,g}$ as a pure i.i.d. disturbance.

By the semiparametric identification results in Ichimura (1993), the dichotomous choice model for the network link implies the following single index equation,

$$\mathrm{E}(w_{ij,g}|C_g, S_g) = P(w_{ij,g} = 1|C_g, S_g) = 1 - F_{\xi_{ij,g}}(-\mu_{ij,g}), \tag{A.2}$$

where $F_{\xi_{ij,g}}(\cdot)$ is the unknown (nonparametric) distribution function of $\xi_{ij,g}$. The identification results in Ichimura (1993) show that the parameters in the linear index $\mu_{ij,g}$ are identified up to a scale and therefore a normalization on one parameter is needed.[15] As the index function is identified, the distribution function $F_{\xi_{ij,g}}$ can also be identified (estimated) from the data by a nonparametric kernel regression with the index $\mu_{ij,g}$ as the regressor. Since the disturbances are continuously distributed, by assuming that $\mu_{ij,g}$ can take on values which cover the support of the probability density $f_{\xi_{ij,g}}$, the moments of $\xi_{ij,g}$ can also be estimated from the data.

We can study the identification constraints on the coefficients η_d, $d = 1, \cdots, \bar{d}$ from the central moments of $\xi_{ij,g}$. We temporarily suppress the group subscript in the following discussion for simplicity. First, we consider a case where Z is one dimensional, i.e.,

[15] With a parametric assumption on $\xi_{ij,g}$, this normalization is not needed.

$\xi_{ij}(Z,\eta) = \eta_1|z_{i1} - z_{j1}| + \zeta_{ij}$. The variance of ξ_{ij} equals $\eta_1^2\text{Var}(|z_{i1} - z_{j1}|) + \sigma_\zeta^2$, where σ_ζ^2 is the variance of ζ_{ij}. The variance σ_ζ^2 is typically normalized to one because of the arbitrary scaling problem in discrete choice models. However, in order to identify η_1, we also need to normalize the variance of unobservable Z to one so that $\text{Var}(|z_{11} - z_{21}|)$ is a known value. This normalization is required for every dimension of Z that we consider.

Next, we consider the case where Z has two dimensions, i.e., $\xi_{ij}(Z,\eta) = \eta_1|z_{i1} - z_{j1}| + \eta_2|z_{i2} - z_{j2}| + \zeta_{ij}$. The variance of ξ_{ij} equals $(\eta_1^2 + \eta_2^2)\text{Var}(|z_{i1} - z_{j1}|) + \sigma_\zeta^2$. When Z is multidimensional, we need an independence assumption between its different dimensions, otherwise, unknown correlations between the different dimensions will make our attempt to identify the coefficients η_d from the central moments of ξ_{ij} impossible. However, even with an independence across dimensions of Z, we still cannot separately identify η_1 and η_2 from $\text{Var}(\xi_{ij})$. Thus, we need to check other identification conditions from higher order central moments of ξ_{ij}.

From the third-order central moment, we can obtain the value of $(\eta_1^3 + \eta_2^3)t + \text{E}[(\zeta_{ij} - \text{E}(\zeta_{ij}))^3]$, where $t = \text{E}[(|z_{i1} - z_{j1}| - \text{E}(|z_{i1} - z_{j1}|))^3]$. To identify η_1 and η_2, we need to know the values of t and $\text{E}[(\zeta_{ij} - \text{E}(\zeta_{ij}))^3]$. Thus, we should assume the unobservable Z comes from a known distribution, e.g., a normal distribution, and ζ_{ij} comes from a known distribution, e.g., a logistic distribution. So far from the second- and the third-order central moments we obtain two polynomial equations involving η_1 and η_2, however, they are still not sufficient to pin down η_1 and η_2. More polynomial equations on η_1 and η_2 can be developed from the fourth and higher order central moments of ξ_{ij}. Eventually, the system of these polynomial equations can be used to solve for the values of η_1 and η_2. The only remaining problem is that the values of η_1 and η_2 can be arbitrarily switched. To avoid this problem, we require $|\eta_1| \geq |\eta_2|$. When Z has $\bar{d}$ dimensions, we will then require $|\eta_1| \geq |\eta_2| \geq \cdots \geq |\eta_{\bar{d}}|$. The implication of this constraint is that z_{i1} ($z_{i\bar{d}}$) represents the dimension of Z which has the greatest (smallest) influence on friendship formation. Using the same strategy, we can identify the coefficients η_d for Z in three or higher dimensions.

We next discuss the identification of the outcome interactions in Eq. (5). First, the group fixed effect in the extended SC-SAR model can be eliminated by a difference approach. The variables Y_g, W_gY_g, X_g, W_gX_g, Z_g, W_gZ_g, and u_g are transformed to $Y_g^* = T_gY_g$, $(W_{pq,g}Y_g)^* = T_g(W_{pq,g}Y_g)$, $X_g^* = T_gX_g$, $(W_gX_g)^* = T_g(W_gX_g)$, $Z_g^* = T_gZ_g$, $(W_gZ_g)^* = T_g(W_gZ_g)$, and $u_g^* = T_gu_g$, where T_g is a $(m_g - 1) \times m_g$ matrix

$$T_g = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & -1 & 1 & \\ & & & \ddots & \ddots \end{pmatrix}.$$

After transformation, we obtain

$$
\begin{aligned}
Y_g^* = \lambda_{11}(W_{11,g}Y_g)^* + \cdots + \lambda_{22}(W_{22,g}Y_g)^* + X_g^*\beta_1 + (W_gX_g)^*\beta_2 \\
+ Z_g^*\delta_1 + (W_gZ_g)^*\delta_2 + u_g^*, \quad g = 1, \cdots, G.
\end{aligned} \tag{A.3}
$$

Taking the expectation of Eq. (A.3) conditional on W_g, we have

$$
\begin{aligned}
\mathrm{E}(Y_g^*|W_g) = \lambda_{11}\mathrm{E}((W_{11,g}Y_g)^*|W_g) + \cdots + \lambda_{22}\mathrm{E}((W_{22,g}Y_g)^*|W_g) \\
+ \mathrm{E}(X_g^*|W_g)\beta_1 + \mathrm{E}((W_gX_g)^*|W_g)\beta_2 + \mathrm{E}(Z_g^*|W_g)\delta_1 + \mathrm{E}((W_gZ_g)^*|W_g)\delta_2,
\end{aligned} \tag{A.4}
$$

for $g = 1, \cdots, G$. Note that all terms in expectation in Eq. (A.4) can be identified from the data. In particular, we can identify $\mathrm{E}(Z_g^*|W_g) = \int_z Z_g^* P(Z_g|W_g)dZ_g = \int_z Z_g^* \frac{P(Z_g)P(W_g|Z_g)}{P(W_g)} dZ_g$ given that the parameters in $P(W_g|Z_g)$ are identified (estimated) from the link formation model. Similarly, we can identify $\mathrm{E}((W_gZ_g)^*|W_g) = \int_z (W_gZ_g)^* P(Z_g|W_g)dZ_g$.
Let $\mathbb{T} = [\mathrm{E}((W_{11}Y)^*|W), \cdots, \mathrm{E}((W_{22}Y)^*|W),\ \mathrm{E}(X^*|W),\ \mathrm{E}((WX)^*|W),\ \mathrm{E}(Z^*|W),\ \mathrm{E}((WZ)^*|W)]$ without the subscript g denote observations stacked across groups. The condition that $\mathbb{T}'\mathbb{T}$ has a full column rank will identify the parameters in Eq. (A.4).

B Estimation details of the SC-SAR model

We estimate the unknown parameters in the extended SC-SAR model through the joint likelihood function of Y_g and W_g, which is given by,

$$
P(Y_g, W_g|X_g, S_g, C_g, \theta, \alpha_g) = \int_{Z_g} P(Y_g|W_g, X_g, S_g, Z_g, \theta, \alpha_g) \cdot P(W_g|C_g, S_g, Z_g, \theta) \cdot f(Z_g|\mu_{z,g})dZ_g, \tag{B.1}
$$

where $\theta = (\gamma', \eta', \lambda', \beta', \delta', \sigma_u^2)$. Notice that in the link formation of Eq. (4), each network link is assumed to be independent conditioning on the observed variables C_g, S_g, and the latent variables Z_g. Therefore, the likelihood function of the network in Eq.(B.1) can be simplified as

$$
P(W_g|C_g, S_g, Z_g, \theta) = \prod_{i}^{m_g} \prod_{j \neq i}^{m_g} P(w_{ij,g}|C_g, S_g, Z_g, \theta). \tag{B.2}
$$

With the likelihood function of Eq.(B.1), we use a Bayesian estimation approach. There are two reasons why we choose the Bayesian approach instead of the classical approach. First, we require some identification constraints on the parameters of the SC-SAR model in the presence of latent variables (see discussion in Appendix A), which generally increases the difficulty of a classical numerical optimization. Using a Bayesian MCMC rejection sampling method such as the Metropolis-Hastings algorithm, the draws that violate the constraint will be easily recognized and rejected. Second, the Bayesian MCMC is effective in handling estimation of models with latent variables (Zeger and Karim, 1991; Hoff and Handcock, 2002; Handcock and Tantrum, 2007). During the posterior simulation, the unobserved correlated effects of the SC-SAR model, including latent variables $\{Z_g\}_{g=1}^G$, their prior means, $\{\mu_{z,g}\}_{g=1}^G$, and groups fixed effects $\{\alpha_g\}_{g=1}^G$, are simulated from the conditional posterior distributions along with the other model parameters to simplify the evaluation of the likelihood function.

We specify the prior distributions of θ, unobserved characteristics $\{z_{i,g}\}$, and group effects $\{\alpha_g\}$ as follows:

$$z_{i,g} \sim \mathcal{N}_{\bar{d}}(\mu_{z,g}, I_{\bar{d}}), \quad i = 1, \cdots, m_g; \; g = 1, \cdots, G, \tag{B.3}$$

$$\mu_{z,g} \sim \mathcal{N}_{\bar{d}}(0, \xi^2 I_{\bar{d}}), \quad g = 1, \cdots, G, \tag{B.4}$$

$$\omega = (\gamma', \eta') \sim \mathcal{N}_{\bar{q}+R+\bar{d}}(\omega_0, \Omega_0) \text{ on the support } O_1, \tag{B.5}$$

$$\lambda = (\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}) \sim U_4(O_2), \tag{B.6}$$

$$\beta = (\beta_1', \beta_2') \sim \mathcal{N}_{2k}(\beta_0, B_0), \tag{B.7}$$

$$\sigma_u^2 \sim \mathcal{IG}(\frac{\kappa_0}{2}, \frac{\nu_0}{2}), \tag{B.8}$$

$$\delta = (\delta_1', \delta_2') \sim \mathcal{N}_{2\bar{d}}(\delta_0, \Delta_0), \tag{B.9}$$

$$\alpha_g \sim \mathcal{N}(\alpha_0, A_0), \quad g = 1, \cdots, G, \tag{B.10}$$

where $\mathcal{IG}$ represents an inverse Gamma distribution. The coefficients γ and η in the function $\psi_{ij,g}$ of Eq. (4) are grouped into ω with the support on O_1 where the identification constraint $|\eta_1| \geq |\eta_2| \geq \cdots \geq |\eta_{\bar{d}}|$ is held. For the endogenous effect λ, we employ a multivariate uniform distribution with a restricted parameter space O_2.[16] The other priors are the commonly used conjugate (uninformative) priors in the Bayesian literature. We choose hyperparameters $\xi^2 = 2$, $\omega_0 = 0$, $\Omega_0 = 100$, $\beta_0 = 0$, $B_0 = 100$, $\kappa_0 = 2.2$, $\nu_0 = 0.1$, $\delta_0 = 0$, $\Delta_0 = 100$, $\alpha_0 = 0$, $A_0 = 100$ to ensure that the prior densities are relatively flat over the range of the data.

Here we list the set of derived conditional posterior distributions that serve as input to the Gibbs sampler:

(i-1) $P(z_{i,g}|Y_g, W_g, \theta, \alpha_g, Z_{-i,g})$, $i = 1, \cdots, m_g$, $g = 1, \cdots, G$.

[16] The restricted parameter space O_2 reflects the stationarity condition required by the outcome equation of the SC-SAR model, which is the matrix $S_g = I_{m_g} - \lambda_{11} W_{11,g} - \cdots - \lambda_{22} W_{22,g}$, $g = 1, \cdots, G$, is invertible, i.e., $det(S_g) > 0$, $g = 1, \cdots, G$, where $det(\cdot)$ stands for the determinant. With an invertible S_g, the outcome vector Y_g is guaranteed not to explode. Due to the restriction imposed on the support of prior distributions, we reject Metropolis-Hastings candidate values of λ which violate this stationarity condition during the posterior simulation.

By applying Bayes' theorem,

$$P(z_{i,g}|Y_g, W_g, \theta, \alpha_g, Z_{-i,g}) \propto \phi_{\bar{d}}(z_{i,g}; \mu_{z,g}, I_{\bar{d}}) \cdot P(Y_g, W_g|\theta, \alpha_g, Z_g), \tag{B.11}$$

where $\phi_{\bar{d}}(.; \nu_{z,g}, I_{\bar{d}})$ is the multivariate normal density function. We simulate $z_{i,g}$ from $P(z_{i,g}|Y_g, W_g, \theta, \alpha_g, Z_{-i,g})$ using the Metropolis-Hastings (M-H) algorithm.

(i-2) $P(\mu_{z,g}|Y_g, W_g, \theta, \alpha_g, Z_g)$, $g = 1, \cdots, G$.
By applying Bayes' theorem,

$$P(\mu_{z,g}|Y_g, W_g, \theta, \alpha_g, Z_g) \propto \mathcal{N}_{\bar{d}}\left(\frac{m_g \bar{Z}_g}{m_g + 1/\xi^2}, \frac{1}{m_g + 1/\xi^2}\right), \tag{B.12}$$

where $\bar{Z}_g = \frac{1}{m_g}\sum_{i=1}^{m_g} z_{i,g}$.

(ii) $P(\omega|\{W_g\}, \{Z_g\})$.
By applying Bayes' theorem, we have

$$P(\omega|\{W_g\}, \{Z_g\}) \propto \phi_{\bar{q}+R+\bar{d}}(\omega; \omega_0, \Omega_0) \cdot \prod_{g=1}^{G} P(W_g|Z_g, \phi) \cdot I(\omega \in O_1), \tag{B.13}$$

where $I(A)$ is an indicator function with $I(A) = 1$ if A is true and zero otherwise. We simulate ω from $P(\omega|\{W_g\}, \{Z_g\})$ using the M-H algorithm.

(iii) $P(\lambda|\{Y_g\}, \{W_g\}, \beta, \sigma_u^2, \delta, \{\alpha_g\}, \{Z_g\})$.
By applying Bayes' theorem, we have

$$P(\lambda|\{Y_g\}, \{W_g\}, \beta, \sigma_u^2, \delta, \{\alpha_g\}, \{Z_g\}) \propto \prod_{g=1}^{G} P(Y_g|W_g, \lambda, \beta, \sigma_u^2, \delta, \alpha_g, Z_g) \cdot I(\lambda \in O_2). \tag{B.14}$$

We simulate λ from $P(\lambda|\{Y_g\}, \{W_g\}, \beta, \sigma_u^2, \delta, \{\alpha_g\}, \{Z_g\})$ using the M-H algorithm.

(iv) $P(\beta|\{Y_g\}, \{W_g\}, \lambda, \sigma_u^2, \delta, \{\alpha_g\}, \{Z_g\})$.
By applying Bayes' theorem, we have

$$\begin{aligned} &P(\beta|\{Y_g\}, \{W_g\}, \lambda, \sigma_u^2, \delta, \{\alpha_g\}, \{Z_g\}) \\ &\quad \propto \phi_{2k}(\beta; \beta_0, B_0) \cdot \prod_{g=1}^{G} P(Y_g|W_g, \lambda, \beta, \sigma_u^2, \delta, \alpha_g, Z_g). \end{aligned}$$

Since both $\phi_{2k}(\beta;\beta_0,B_0)$ and $P(Y_g|W_g,\lambda,\beta,\sigma_u^2,\delta,\alpha_g,Z_g)$ are normal density functions, we simplify the expression to

$$
\begin{aligned}
& P(\beta|\{Y_g\},\{W_g\},\lambda,\sigma_u^2,\delta,\{\alpha_g\},\{Z_g\}) \propto \mathscr{N}_{2k}\left(\beta;\hat{\beta},\mathbf{B}\right) \\
& \hat{\beta} = \mathbf{B}\left(B_0^{-1}\beta_0 + \sum_{g=1}^{G}\mathbf{X}_g'(\sigma_u^2 I_{m_g})^{-1}(S_g Y_g - \mathbf{Z}_g\delta - l_g\alpha_g)\right) \\
& \mathbf{B} = \left(B_0^{-1} + \sum_{g=1}^{G}\mathbf{X}_g'(\sigma_u^2 I_{m_g})^{-1}\mathbf{X}_g\right)^{-1},
\end{aligned}
\tag{B.15}
$$

where $\mathbf{X}_g = (X_g, W_g X_g)$, $\mathbf{Z}_g = (Z_g, W_g Z_g)$, and $S_g = (I_{m_g} - \lambda_{11}W_{11,g} - \cdots - \lambda_{22}W_{22,g})$.

(v) $P(\sigma_u^2|\{Y_g\},\{W_g\},\lambda,\beta,\delta,\{\alpha_g\},\{Z_g\})$.

By applying Bayes' theorem, we have

$$
\begin{aligned}
& P(\sigma_u^2|\{Y_g\},\{W_g\},\lambda,\beta,\delta,\{\alpha_g\},\{Z_g\}) \\
& \quad \propto \mathscr{IG}\left(\sigma_u^2;\frac{\kappa_0}{2},\frac{\nu_0}{2}\right)\prod_{g=1}^{G}P(Y_g|W_g,\lambda,\beta,\sigma_u^2,\delta,\alpha_g,Z_g) \\
& \quad \propto \mathscr{IG}\left(\sigma_u^2;\frac{\kappa_0+\sum_{g=1}^{G}m_g}{2},\frac{\nu_0+\sum_{g=1}^{G}u_g'u_g}{2}\right),
\end{aligned}
\tag{B.16}
$$

where $u_g = S_g Y_g - \mathbf{X}_g\beta - \mathbf{Z}_g\delta - l_g\alpha_g$.

(vi) $P(\delta|\{Y_g\},\{W_g\},\lambda,\beta,\sigma_u^2,\{\alpha_g\},\{Z_g\})$.

By applying Bayes' theorem, we have

$$
\begin{aligned}
& P(\delta|\{Y_g\},\{W_g\},\lambda,\beta,\sigma_u^2,\{\alpha_g\},\{Z_g\}) \\
& \quad \propto \mathscr{N}_{2\bar{d}}(\delta;\delta_0,\Delta_0)\prod_{g=1}^{G}P(Y_g|W_g,\lambda,\beta,\sigma_u^2,\delta,\alpha_g,Z_g),
\end{aligned}
\tag{B.17}
$$

Similar to (v), we can further obtain

$$
\begin{aligned}
& P(\delta|\{Y_g\},\{W_g\},\lambda,\beta,\sigma_u^2,\{\alpha_g\},\{Z_g\}) \propto \phi_{2\bar{d}}(\delta;\hat{\delta},\mathbf{D}), \\
& \hat{\delta} = \mathbf{D}\left(\Delta_0^{-1}\delta_0 + \sum_{g=1}^{G}\mathbf{Z}_g'(\sigma_u^2 I_{m_g})^{-1}(S_g Y_g - \mathbf{X}_g\beta - l_g\alpha_g)\right) \\
& \mathbf{D} = \left(\Delta_0^{-1} + \sum_{g=1}^{G}\mathbf{Z}_g'(\sigma_u^2 I_{m_g})^{-1}\mathbf{Z}_g\right)^{-1},
\end{aligned}
\tag{B.18}
$$

(vii) $P(\alpha_g|Y_g,W_g,\lambda,\beta,\sigma_u^2,\delta,Z_g)$, $g=1,\cdots,G$.

By applying Bayes' theorem, we have

$$
P(\alpha_g|Y_g,W_g,\lambda,\beta,\sigma_u^2,\delta,Z_g) \propto \phi(\alpha_g;\alpha_0,A_0)\cdot P(Y_g|W_g,\lambda,\beta,\sigma_u^2,\delta,\alpha_g,Z_g). \tag{B.19}
$$

Similar to (v), we can further obtain

$$
\begin{aligned}
&P(\alpha_g|Y_g, W_g, \lambda, \beta, \sigma_u^2, \delta, Z_g) \propto \mathscr{N}(\alpha_g; \hat{\alpha}_g, R_g), \\
&\hat{\alpha}_g = R_g\Big(A_0^{-1}\alpha_0 + l_g'(\sigma_u^2 I_{m_g})^{-1}(S_g Y_g - \mathbf{X}_g\beta - \mathbf{Z}_g\delta)\Big), \\
&R_g = \Big(A_0^{-1} + l_g(\sigma_u^2 I_{m_g})^{-1} l_g'\Big)^{-1}.
\end{aligned} \tag{B.20}
$$