Volume Title: Research Findings in the Economics of Aging

Volume Author/Editor: David A. Wise, editor

Volume Publisher: The University of Chicago Press

Volume ISBN: 0-226-90306-0

Volume URL: http://www.nber.org/books/wise08-1

Conference Dates: May 10-13, 2007

Publication Date: February 2010

Chapter Title: Comment on "Social Interactions and Smoking"

Chapter Author: Arie Kapteyn

Chapter URL: http://www.nber.org/chapters/c8197

Chapter pages in book: (141 - 144)

neighborhood on disadvantaged youths. NBER Working Paper no. 3705. Cambridge, MA: National Bureau of Economic Research, May.
Christakis, N. A., and J. H. Fowler. 2007. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine* 357 (4): 370–79.
Crane, J. 1991. The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *American Journal of Sociology* 96 (5): 1226–59.
DeMarzo, P. M., D. Vayanos, and J. Zweibel. 2003. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics* 118 (3): 909–68.
Ellison, G., and D. Fudenberg. 1993. Rules of thumb for social learning. *Journal of Political Economy* 101 (4): 612–43.
Evans, W. N., M. C. Farrelly, and E. B. Montgomery. 1999. Do workplace smoking bans reduce smoking? *American Economic Review* 89 (5): 729–47.
George, L., and J. Waldfogel. 2003. Who affects whom in daily newspaper markets? *Journal of Political Economy* 111 (4): 765–84.
———. 2006. The *New York Times* and the market for local newspapers. *American Economic Review* 96 (1): 435–47.
Glaeser, E., B. Sacerdote, and J. Scheinkman. 1996. Crime and social interactions. *Quarterly Journal of Economics* 111 (2): 507–48.
———. 2003. The social multiplier. *Journal of the European Economics Association* 1 (2): 345–53.
Kuziemko, I. 2006. Is having babies contagious? Fertility peer effects between adult siblings. Princeton University. Unpublished Manuscript.
Manski, C. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60 (3): 531–42.
Sacerdote, B. 2001. Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics* 116 (2): 681–704.
Topa, G. 2001. Social interactions, local spillovers and unemployment. *Review of Economic Studies* 68 (2): 261–95.

Comment Arie Kapteyn

It is gratifying to see that economics is catching up quickly with other social sciences (particularly sociology and social psychology) by incorporating social interactions into models of behavior. One contribution economics may make is to bring rigor to the field and to characterize in particular what is and what is not identifiable in the models that we consider (as in Manski [1993]). Smoking is clearly an example where we would expect social interactions to be important, but also one where social interactions are hard to distinguish from other reasons for observing clusters of smokers or nonsmokers. The most obvious problem is that smokers (or nonsmokers) may flock together. So if we see that smokers often have friends or spouses who smoke, this may point to social interactions, but it may also simply indicate correlated preferences.

Arie Kapteyn is a senior economist at RAND Labor and Population.

In their chapter, Cutler and Glaeser implement an independent variable (IV) strategy, where the existence of workplace smoking bans is used as an instrument. They also consider various other pieces of evidence for the existence of social interactions, to which I will return later.

To set the stage, I find it useful to write down a simple linear model that is akin to theirs, but with slightly more detail. Thus, consider the following structure:

$$X_1 = a_1 + bX_2 + \varepsilon_1 \tag{1}$$
$$X_2 = a_2 + bX_1 + \varepsilon_2$$

Here, X_1 and X_2 are indicators for smoking by spouses 1 and 2, respectively; a_1, a_2, and b are parameters; ε_1 and ε_2 are random i.i.d. error terms, with variances σ_1^2 and σ_2^2 and covariance σ_{12}. I will assume that $0 < b < 1$. In principle, the parameters a_1 and a_2 can be made functions of other explanatory variables and I will return to that later. For now we note that the system (1) implies the following reduced form

$$X_1 = \frac{1}{1 - b^2}[a_1 + a_2 b + b\varepsilon_2 + \varepsilon_1] \tag{2}$$
$$X_2 = \frac{1}{1 - b^2}[a_2 + a_1 b + b\varepsilon_1 + \varepsilon_2].$$

Thus, if we regress X_1 on X_2, for instance, then the regression coefficient in the population would be equal to:

$$\frac{\text{cov}(X_1, X_2)}{\text{var}(X_2)} = \frac{(1 + b^2)\sigma_{12} + b(\sigma_1^2 + \sigma_2^2)}{b^2\sigma_1^2 + 2b\sigma_{12} + \sigma_2^2}. \tag{3}$$

For $\sigma_1^2 = \sigma_2^2 = \sigma^2$ this simplifies to

$$\frac{\text{cov}(X_1, X_2)}{\text{var}(X_2)} = \frac{(1 + b^2)\sigma_{12} + 2b\sigma^2}{(1 + b^2)\sigma^2 + 2b\sigma_{12}}. \tag{4}$$

And finally, if we assume $\sigma_{12} = 0$ then the regression coefficient is equal to $2b/(1 + b^2)$, which is the result given in the chapter.

It is easy to see that for $\sigma_{12} \geq 0$ the regression of X_1 on X_2 will lead to an overestimate of b, no matter if we apply (3), (4), or the result obtained in the chapter (i.e., with $\sigma_{12} = 0$). It seems very unlikely that σ_{12} would be negative. Hence, we would expect a consistent estimation method for b to yield a smaller value than ordinary least squares (OLS). Yet the chapter finds the opposite: the IV estimate of b is roughly double the OLS estimate.

Cutler and Glaeser offer a number of possible explanations for this finding. I would like to suggest one additional possibility: assortative mating. Suppose that spouses search each other out partly on the basis of smoking preferences. This would suggest that an individual who does not smoke may be more likely to select a spouse who is subject to a smoking ban. This, then, is likely to introduce a correlation between the smoking ban variable and the error term in this individual's smoking equation and thereby invalidating this variable as an instrumental variable. This possibility also suggests a simple, though perhaps crude, test. Given that the prevalence of smoking bans has increased over time, one would expect the correlation of the instrument with the error term to be less for couples who have been together longer. The reasoning for this is that for a couple that got married when there were no smoking bans, this cannot have affected their marriage decision. If this suspicion is correct, the IV estimate of b should be higher for recently weds than for couples that married a long time ago. As a first approximation, one can use age as a proxy for the duration of the marriage.

There is a more general question as to why one needs to use the smoking ban variable as an instrumental variable at all. Let us consider the IV estimation a little closer. In particular, assume that $a_1 = \alpha_1 + \delta_1 z_1$ and $a_2 = \alpha_2 + \delta_2 z_2$, where z_1 and z_2 are exogenous variables that can be used as instruments. Clearly, 2SLS will yield consistent estimates of the parameters in the model, including b. The model used in the chapter includes a large number of controls, which are assumed exogenous, or at least predetermined. In the logic of simultaneous equations, all of these can be used as instruments. There does not seem to be a need, therefore, to use workplace smoking bans.

An interesting second approach to the analysis of social interactions is to consider "social multipliers" and "excess variance" across groups. This is easily illustrated in the framework previously introduced. Let us first consider social multipliers. Suppose that in (1) both a_1 and a_2 are increased by an amount Δa. Ignoring social interaction, both X_1 and X_2 would increase by the same amount Δa. Taking into account social interactions, that is, using (2), we find that both X_1 and X_2 increase by $\Delta a/(1 - b)$, which is larger. The social multiplier is equal to $1/(1 - b)$. Although I illustrate the social multiplier in the model for two spouses, the same idea applies to other groups in which social interactions may take place, like metropolitan areas or states. It is found that the effect of smoking bans on smoking goes up quite strongly when moving from individual level data to metropolitan level data and then to state level data. Qualitatively similar increases in the social multiplier are found for log-income, but not for years of education. Within the context of the model that finding is problematic, since the simple set-up used in the chapter (and in the previous equations) would imply identical social multipliers for any exogenous variable (any variable on which a_1 and a_2 would depend).

I will also illustrate the notion of excess variance in the context of the spouses model. For simplicity, consider the case $\sigma_1^2 = \sigma_2^2 = \sigma^2$. We then have for $b = 0$ that the conditional variance of "mean smoking" in the household is equal to

(5) $$\mathrm{Var}\left(\frac{X_1 + X_2}{2} \mid a_1, a_2\right) = \frac{1}{2}\{\sigma^2 + \sigma_{12}\}.$$

For $0 < b < 1$ we have that

(6) $$\mathrm{Var}\left(\frac{X_1 + X_2}{2} \mid a_1, a_2\right) = \frac{1}{2(1 - b)^2}\{\sigma^2 + \sigma_{12}\},$$

which is clearly larger. Again, the same idea can be applied to larger groups than just couples. The data suggest a substantial amount of excess variance. It would have been interesting to see the same idea applied to couples, as in this example. The formulas also suggest a cautionary note (acknowledged in the chapter): if one were to incorrectly assume that the errors are uncorrelated then the presence of σ_{12} in (5) would suggest excess variance, where only correlated effects are to blame for the larger than expected variance.

A final exercise in the chapter involves a simple dynamic model of adoption of smoking and nonsmoking. It turns out that the dynamic model provides a poor fit of the aggregate data. It is not clear why that is, and I would hesitate at this moment to take that as evidence against social interactions.

So where does this leave us? The chapter presents an elegant framework suggesting various patterns in the data if social interactions are present. By and large, the predicted patterns are indeed found in the data and I am inclined to interpret this as solid evidence of social interactions in smoking. One might argue that social interactions in smoking are so obvious, that empirical evidence to support it is hardly necessary. Apart from the obvious rejoinder that obtaining a quantitative estimate of the extent of interactions in smoking is important by itself, perhaps the most important contribution of the chapter is to show how the analytic framework used helps to interpret and understand data. At the same time, the patterns in the data are not all consistent with the model predictions. The largest marginal value of future work may therefore be found in adaptations of the analytic framework, to be able to accommodate richer patterns in the data. The chapter provides an excellent illustration of the value of a powerful analytic framework for understanding the extent and nature of social interactions.

Reference

Manski, C. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60 (3): 531–42.