DEPARTMENT OF ECONOMICS AND FINANCE WORKING PAPER SERIES • *April 2011*

Impact of Education on Lifestyles: What Do Longitudinal Data Show?

Shah Danyal, Bichaka Fayissa*, and Jong-Sung Lee
Middle Tennessee State University

Abstract

This essay investigates the effect of education on different lifestyle variables using NLSY79 panels for 1992, 1994, and 1998. The lifestyle variables are smoking, drinking, marijuana use, and cocaine use. The analysis addresses the joint determination of lifestyle variables within the framework of the Seemingly Unrelated Regression (SUR) model. Unobserved heterogeneity is controlled by the robust fixed-effects model extended to SUR model. It is found that educational attainment has no significant effect on the lifestyle choices of individuals.

Keywords**:** Education, Smoking, Drinking, Marijuana and Cocaine Use, Fixed-Effects Model, SUR Model

JEL Classification**:** I1, I2, I10, I12, C30

*Bichaka Fayissa, Professor, Department of Economics and Finance, Middle Tennessee State University, Murfreesboro, TN 37132, phone: 615-898-2385, fax: 615-898-5596, email: bfayissa@mtsu.edu

Impact of Education on Lifestyle Choices: What Do Longitudinal Data Show?

1. Introduction

There have been long quests by both epidemiologists and economists to explain inequalities in health. Researchers appear to reach the general consensus that while access to medical care is important, it does not fully explain inequalities in the health status of individuals (Lleras-Muney, 2005; Folland et al., 2001). Economists have widely studied and recognized that there is a positive relation between health and education (Grossman, 1972; Ross and Wu, 1995). They have also found evidence that education is strongly correlated with most health-related behaviors or lifestyle factors such as smoking, drinking, and substance abuse. These lifestyle factors are assumed to be inputs in the individual's production function of health (Park and Kang, 2008). Disputes that persist, however, relate to the issue of causation versus association. One body of literature contends that causation runs from education to health, or health-related behavior. Grossman (1972) articulates the idea that education improves the efficiency of health production, or educated people are able to use more efficient mix of health inputs (Kenkel, 1991; Rosenzweig, 1995). Similarly, Cowell (2006) argues that educated people will choose to have improved health-related behaviors or lifestyles, as the opportunity cost of being ill in the future is high.

Another body of literature, however, casts doubt on whether better education necessarily leads to better health, or lifestyle choices that improve health. One justification of this view rests on the notion that the causation may be reverse. In other words, is it better health that leads to better schooling (Currie and Hyson, 1999), or do lifestyle choices such as substance abuse affect educational attainment (Register and Grimes, 2001). Still other studies argue that there may be no causal relationship since there may be other factors such as the rate of discount, heredity, or preferences affecting both education and health or health-related behavior (Fuchs, 1982; Rosenzweig and Shultz, 1983; Farrell and Fuchs, 1982). Fuchs (1974) is also of the opinion that personal lifestyle is the most important factor causing great variations in health.

The present paper investigates the impact of education on the lifestyle variables that are assumed to be determinants of health by considering three main issues: (1) the definition of lifestyle, (2) the unobserved heterogeneity among agents or individuals that makes the empirical results biased, and (3) the joint determination of lifestyle variables treated as choice variables for individuals. Previous studies have used either only one choice variable such as smoking (deWalque, 2007; Sander, 1995), or multi-lifestyle variables such as smoking, drinking substance use, having regular breakfast, exercise, and medical checkups, or restful patterns of sleeping in finding the impact of education on lifestyle choices (Kenkel, 1991, 1997; Park and Kang, 2008). For the purpose of this essay, only smoking, marijuana use, cocaine use, and drinking are used as proxies for lifestyle variables.

In earlier studies, the unobserved heterogeneity of education among individuals or agents was not considered (Kenkel, 1991). Recent studies have, however, controlled the unobserved heterogeneity among agents and the endogeneity of the education variable by using the instrumental variable technique(Park and Kang, 2008). The use of an instrumental variable such as family background was criticized (Grossman, 2005). The trend in the choice of the instrumental variables has evolved to the wide use of institutional characteristics (Tenn et el., 2010). The difficulties associated with finding the right instruments are, however, well documented in the literature (Bound et al., 1995). This essay takes a different approach to avoid the controversy surrounding the selection of instruments by adopting the fixed-effects method on panel data drawn from the National Longitudinal Survey of Youth 1979 (NLSY79). de Walque (2010) uses the fixed-effect approach by constructing panels from retrospective histories of individuals from National Health Interview Surveys of year 1978 and year 2000 for finding the effect of education on single variable smoking. The present paper uses the longitudinal data where data are collected at the point of survey and hence contain more current information instead of relying on past memory of individuals which may add measurement errors. de Walque (2010) also assumes that college graduation takes place between ages 17-25 and education is assumed not to change after the age of 25 years. In this paper, we argue that education does, indeed, vary beyond the age of 25 years. According to the National Center for Education Statistics (NCES), the number of older students has been growing along with the number of younger students (Miller, 2001). Between 1990 and 1999, the enrollment of students under age twenty-five increased by eight percent. During the same period, enrollment of persons 25 and over rose by seven percent. In 1997, the average age of community college students was 29, with 46% of students over the age of 25. From 1999 to 2010, NCES projects an increase of nine percent in the number of students over age twenty-five (Miller, 2001).

In dealing with the multi-lifestyle variables, Kenkel (1991) estimates each equation separately without considering their joint determination in contrast to Park and Kang (2008) who do. This essay also addresses the issue of the joint determination of the lifestyle variables within the empirical framework of the Seemingly Unrelated Regression (SUR) model. It is shown that estimating each equation using the fixed-effects model is appropriate if the explanatory variables or regressors are identical in each equation. This study utilizes the fundamentals of the SUR model which is generally underutilized in educational research as argued by Beasley (2008) and Green (2008: 267) by incorporating the fixed-effects model into the SUR.

In the literature, the relation of the random-effects and fixed-effects models to SUR models are discussed by Avery (1977), Wooldridge (2002: 272-274), and Blackwell (2005), but the elimination of the fixed effects from SUR regressions is not discussed explicitly. This essay is different from previous studies in two respects: (1) it estimates the fixed-effects model equation by equation and shows that it is equivalent to the joint determination in the SUR framework and (2) it also makes use of the robust fixed-effects technique to address the issue of biasedness due to the individual unobserved heterogeneity arising from the cross-correlation and within-correlation of lifestyles.

The essay uses panel data from the NLSY79. Only two-period panels (1992 and 1994) are used for determining the impact of education on smoking, alcohol consumption, marijuana use, and cocaine use due to data the unavailability of data for the drinking variable in 1998. We, however, use the 1992, 1994, and 1998 panel data to examine the impact of education on the other three variables (smoking, marijuana use, and cocaine use). Overall, the study shows that education does not have a significant effect on lifestyle choices after controlling for individual unobserved heterogeneity.

The paper is organized as follows. The next section gives a brief review of the literature. Section 3 describes the data and methodology. The empirical results are presented in section 4. The final section draws conclusion based on the results.

2. Review of Literature

Health care is one of the factors in maintaining health. Lifestyle is also considered to be an important determinant of health (Folland et al., 2001). Fuchs (1979) compared Utah and Nevada in terms of death rates and found that such factors as abstinence from the use of tobacco and alcohol will increase individuals' longevity. In another study, Fuchs (1982) also argued that personal lifestyle is a significant determinant of health, while the Rosenzweig and Schultz (1983) study shows that maternal cigarette smoking has a significant negative effect on the newborn birth weight. Joyce et al. (1992) have also found that illicit substance abuse by pregnant women significantly harms the newborn.

Many studies have addressed the effect of education on one lifestyle variable such as smoking. Among them, Sander (1995) finds that education has a negative effect on smoking after accounting for the endogeneity of the education variable using the instrumental variable technique. Since the instrument used is related to the family-background, however, the variable is criticized as being weak instrument (Grossman, 2005). In examining the effect of mother's education on birth outcome and smoking by using the availability of college in the maternal county as an instrumental variable, Currie and Moretti (2003) find a positive effect of education on birth weight and a negative effect of education on smoking during pregnancy. After controlling for endogeneity using Vietnam veteran's status as an instrumental variable, Grimard and Parent (2007) find no significant effect of education on smoking, while de Walque (2007) finds that education discourages smoking by using the risk of being drafted to the Vietnam War as an instrumental variable for men's college education.

Using cigarette smoking, alcohol consumption, and exercise as lifestyle variables, Kenkel (1991) finds that it is the health knowledge which affects lifestyle without drawing any conclusion about schooling as a factor affecting the above lifestyle variables. On the other hand, Contoyannis and Jones (2004) use a multivariate probit model (the Maximum Simulated Likelihood variety) for determining the impact of the discrete lifestyle choices on self-assessed health. They find that lifestyle choices affect health and that education affects lifestyle choices. Park and Kang (2008) examine health-related multiple lifestyle choices such as regular checkups, exercise, smoking, and alcohol consumption. Using the instrumental variable technique, they find that education is not a significant determinant of smoking or drinking, but they find that education has a significant effect on regular exercise and checkups. As is evident from the above discussions, the effect of education on lifestyle choices is mixed. Most of these studies vary only by birth cohort and gender or birth cohort and geographical location when they employ instrument variables (Tenn et al., 2010). The unobserved heterogeneity in Tenn et al. (2010) is controlled with education differences between similarly selected same groups one year apart in age. They find that there is no evidence of education affecting smoking after controlling for fixed-effects. The study does, however, not control the individual unobserved heterogeneity. Constructing panel data from the histories of the individuals given in Health Interview Surveys of year 1978 and year 2000, de Walque (2010) finds mixed results for the impact of education on smoking after controlling for the unobserved heterogeneity by the fixed-effects model.

This study attempts to provide some evidence on the effect of education on multiple lifestyle choices using the fixed-effects model to control for the unobserved individual heterogeneity within the SUR model. We use longitudinal data from NLSY79 which give more current information about the individual and, thus, less expose to measurement errors due to the information gathered from recalling long past histories of individuals. The data and the methodology are described in the following section.

3. Data and Methodology

The NLSY79 dataset for the present study is a nationally representative sample of 12,686 of male and female who were 14-22 years old when they were first surveyed in 1979. Individuals were interviewed annually from 1979 through 1994 and biannually beginning from 1996. The collection of data for drug use began in 1988 and was repeated in 1992, 1994, and 1998. There were variations in some of the survey questions as well as the definition of new variables with respect to drug use in the different surveys. For the purpose of the present analysis, however, data regarding the four lifestyle variables were extracted only from the 1992 and 1994 surveys. The analysis of the three lifestyle variables (smoking, marijuana use, and cocaine use) is based on data for the years 1992, 1994, and 1998. The NLSY79 survey started with 12,686 respondents, but decreased to 9002 by end of 1992, 8,794 by the end of 1994, and 8,403 by the end of 1998 due to attrition. The retention rate between 1992 and 1998 is, however, is very high. The present study treats those observations as missing information to avoid systematic bias. Unanswered questions are also considered as missing data.

The dependent variables are in binary form and their definitions of are given in Table 1. Among the explanatory variables, education and wage are in continuous form while the rest of regressors are dummy variables. The descriptive statistics are presented in Table 2. The data suggest that the trend in smoking, drinking, marijuana use, and cocaine use is declining as shown in Figures 1 through 4. The years of education have increased from 12.8 years to 13.0 years, while wages declined from \$16.70 in 1992 to \$14.76 in 1998 (a decline of about 12 percent). Almost 50 percent of the respondents are males, 25 percent Blacks, and 15.7 percent Hispanics. About 58 percent of the respondents are married and more than 70 percent of the respondents lived in urban locations.

Figures 1 through 4 show that the trends in smoking, marijuana use and cocaine use have declined steeply for both genders. The drinking variable also decreased for both genders, but at a slower rate. The figures reveal that males dominate in drinking, smoking, marijuana use, and cocaine use. Whether this declining trend for these variables is due to education, or wage changes is the research objective of this study. As discussed above, this paper deals with the impact of education, wages, and other demographic variables (gender, marital status, race, and location of residence) on various lifestyle choices by heterogeneous individuals with important implications for their health outcome. Since the effect of education and other demographic factors on multiple lifestyle variables such as smoking, drinking, marijuana use, and cocaine use involve multiple equations that on the surface appear to be independent of each other, but in fact involve cross-correlations, we employ the SUR model as developed by Zellner (1962).

In the present analysis, there are two sets of equations. In the first set, there are four equations; in the other set, there are three equations. We initially set up the pooled cross-section analysis by estimating four equation for the four dependent lifestyle variables for the years 1992 and 1994 and then three equations for three dependent variables for the years 1992, 1994 and 1998. The model is in the general form in these types of studies and is specified as follows:

$$\mathbf{Y_{it,j} = \beta_j X_{it} + €_{it,j}} \qquad \text{where } \mathbf{j=1, 2,3,4.} \qquad \mathbf{(1)}$$

where $\mathbf{Y_{it,j}}$ is a vector of the dependent lifestyle variable for individual *i* at time *t*, and $_j$ indexes the equation number. $\mathbf{X_{it,j}}$ denotes the control variables such as education, gender, race, marital status, wages, and location of residence. $\mathbf{€_{it,j}}$ is the vector the disturbance terms. Model (1) can be illustrated in a vector matrix form as below:

$$\begin{bmatrix} \boldsymbol{Y_{it,1}} \\ \boldsymbol{Y_{it,2}} \\ \boldsymbol{Y_{it,3}} \\ \boldsymbol{Y_{it,4}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_{it,1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X_{it,2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{X_{it,3}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{X_{it,4}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \\ \boldsymbol{\beta_3} \\ \boldsymbol{\beta_4} \end{bmatrix} + \begin{bmatrix} \boldsymbol{€_{it,1}} \\ \boldsymbol{€_{it,2}} \\ \boldsymbol{€_{it,3}} \\ \boldsymbol{€_{it,4}} \end{bmatrix} \quad (2)$$

(i)(t)(4)×1 (i)(t)(4) ×K K×1 (i)(t)(4)×1

where $K = \sum K_j$, and K_j is the number of explanatory variables in equation j. If the regressors are same in all the equation, there is the absence of individual-specific errors, and assuming strict homoscedasticity and that disturbance terms are correlated across, but uncorrelated within equations, then the OLS equation by equation is equivalent to the joint determination in SUR model (Green, 2008, 257-258; Wooldridge, 2002, 148-150).

A test for heteroscedasticity reveals its presence, and we use the Breusch-Pagan approach to test its presence. We use the robust OLS equation by equation estimates of the coefficients to compare our results to the robust fixed-effects model.

The individual specific-effects, or fixed-effects in the above-stated SUR model in equation (2) are described as follows. The disturbance terms $\mathbf{€_{itj}}$ account for the individual-specific or omitted variables and other random errors. They are decomposed into individual specific errors $\mathbf{c_{i,j}}$ and random errors, or “noise” term $\mathbf{u_{it,j}}$. The vector matrix form becomes as follows:

$$\begin{bmatrix} \boldsymbol{Y_{it,1}} \\ \boldsymbol{Y_{it,2}} \\ \boldsymbol{Y_{it,3}} \\ \boldsymbol{Y_{it,4}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_{it,1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X_{it,2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{X_{it,3}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{X_{it,4}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \\ \boldsymbol{\beta_3} \\ \boldsymbol{\beta_4} \end{bmatrix} + \begin{bmatrix} c_{i,1} \\ c_{i,2} \\ c_{i,3} \\ c_{i,4} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\mu_{it,1}} \\ \boldsymbol{\mu_{it,2}} \\ \boldsymbol{\mu_{it,3}} \\ \boldsymbol{\mu_{it,4}} \end{bmatrix} \quad (3)$$

(i)(t)(4)×1 (i)(t)(4) ×K K×1 (i)(t)(4)×1

K = $\sum K_j$, and K_j is the number of the explanatory variable in equation j.

As in Avery (1977), we consider $c_{i,j}$ in each equation as the fixed-effects instead of random effects and extend it to the multiple equations of the SUR model. While Blackwell (2005) estimates the fixed-effects coefficients $c_{i,j}$ in the SUR model, we initially remove the$c_{i,j}$ as done in the fixed-effects model (Green, 2008, 190).

There are two ways to remove $c_{i,j}$. One way is to take the first difference, while the other way is to use the fixed-effects transformation following Wooldridge (2002). The fixed-effects transformation is done by first averaging each equation j over the period to get the cross-sectional effect for each equation j as given below and then, it is subtracted from each jth equation.

$$\bar{Y}_{i,j} = \beta \bar{X}_{i,j} + c_{i,j} + \bar{\mu}$$

If the regressors are not identical, then the FGLS method of estimation is applied to the above equations in deviation form as in SUR model which solves the problem of joint determination by taking out the effect of the within and across correlations among the equations. If the explanatory variables are identical, however, the fixed-effects model can be applied equation by equation, amounting to the SUR model which addresses the issue of joint determination. This simple assertion is not explicitly applied in literature to the best of the authors' knowledge. We estimate each equation by the fixed-effects model and the issue of the joint determination of lifestyle variables is resolved based on the above assertions.

A possibility exists that the individual time-variant effects are controlled by using the following model

$$Y_{it,jt} = \beta X_{it,t} + c_{i,j} + \delta_{it} + \mu_t$$

where $\delta_{,t}$ is the individual time-variant effects. The fixed-effects model estimation of the above equation presumes that lifestyle variables are estimated separately without considering the joint determination since the distribution of δ_{it} is unknown.

In sum, we first estimate each equation separately by the OLS method and then by fixed-effect model to address the issue of joint determination. We also estimate the equations by controlling for the time-variant factor and fixed-effects. The results from the application of the three techniques are discussed in the next section.

4. Empirical Results

The results from the system of the robust OLS regressions using the dataset for the years 1992, 1994, and 1998 are presented in Table 3. The cross-sectional results show that education discourages smoking, marijuana use, and cocaine use for both sets of data. For the drinking lifestyle, however, the educational attainment variable has a positive and statistically significant effect. The Kenkel (1991) study also shows that education discourages smoking and has a positive effect on drinking. Using South Korean data, Park and Kang (2008) also obtain similar results for the smoking and drinking variables when the unobserved individual heterogeneity is not controlled. The result of the robust OLS estimate showing that education has a negative impact on smoking is in line with other works by Sander (1995) and de Walque (2008).

The robust OLS results show that for each additional year, educational attainment reduces smoking, marijuana use, and cocaine use by 5 percent, 0.4 percent, and 0.1 percent, respectively. The analysis of the 1992 and 1994 periods indicates that education has a significantly positive effect on alcohol drinking, perhaps, confirming the recent medical information which suggests that moderate drinking of one or two servings of alcohol daily may promote the health status of individuals (Rimm et al., 1996; Davies et al., 2002; DHHS, 2005).

The robust fixed-effects model is applied to estimate equation by equation and the results are reported in Table 4. Wages are still insignificant in the robust fixed-effects model as in the OLS, while the impact of education on lifestyle choices becomes insignificant, contrary to the OLS results.

Table 5 shows the time-variant effect with the fixed-effects results, confirming again the insignificance of education in affecting the lifestyles choices. Education has a negative and significant effect on the lifestyle choice only for the year 1996.

5. Conclusion

This paper has explored the effect of educational attainment and wages on lifestyle choices by considering the correlation among lifestyle variables in pooled cross section analysis and fixed-effects model. It has been noted above that if each equation has an identical set of explanatory variables, then the OLS estimation is equivalent to the estimation by FGLS for the SUR model. The paper then introduces the time-variant individual effects in the fixed-effects model. The cross-sectional results of the OLS method shows that there is a negative and significant effect of educational attainment on smoking, marijuana use, and cocaine use and a positive effect on drinking in line with the Parks and Kang (2008) study and similar effect is found for the smoking variable in de Walque (2010) without controlling the unobserved heterogeneity.

The paper addresses the issue of heterogeneity by extending the robust fixed-effects model to multiple equations within the SUR framework. When the heterogeneity of individuals is controlled, the results are very different from the analysis based on the robust OLS method. It is found that education has no statistically significant effect on the lifestyle choices, although it has a negative overall impact. The negative and significant effect for the educational attainment variable in the robust OLS analysis may be due to the unobserved heterogeneity factor. The time trend shows that, over the period of the study, the lifestyle variables have declining trends. After controlling for heterogeneity and heteroskedacity using the robust fixed-effects model, we conclude that factors other than education may have been responsible for the relationship. Similar impacts are observed for the fixed-effects model after controlling the individual time variant factors. Wages also do not have a significant impact on lifestyle choices. This may be interpreted as suggesting that lifestyle choices are independent of the individual's income from working, perhaps, because such addiction habits supported by other sources

In the case of smoking, de Walque's (2010) results are mixed for different cohorts. Tenn *et al*. (2010) show that when unobserved heterogeneity is controlled, education has no impact on smoking. Kenkel (1991) argues that it is health knowledge that affects lifestyle variables. In case of the effect of education on smoking, de Walque (2010) suggests that the channels for discouraging smoking may be further explored. The negative trend in lifestyle choices obtained in this paper suggests that policy and its strict implementation may be the cause in explaining the negative trend of lifestyles variables. Thus, policies governing the lifestyle choices may be more effective in discouraging choices that are injurious to health.

References

Avery, R.B. (1977). Error components and seemingly unrelated regressions, *Econometrica*, 45, 199-209.

Beasley, T. M., (2008). Seemingly Unrelated Regression (SUR) Models as a Solution to Path Analytic Models with Correlated Errors, *Multiple Linear Regression Viewpoints*, 34(1):1-7.

Blackwell, J. (2005). Estimation and testing of fixed-effect panel-data systems, *The Stata Journal*, 5, 202-207.

Bound, J., Jaeger, D.A., and Baker, R.M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak, *Journal of the American Statistical Association*, 90, 443-450.

Contoyannis, P. and Jones, A. (2004). Socio-economics status, health and lifestyle, *Journal of Health Economics*, 23, 965-995.

Cowell, A. (2006). The relationship between education and health behavior: Some empirical evidence, *Health Economics*, 15, 124-146.

Currie, J., and Hyson, R. (1999). Is the impact of health stock cushioned by socioeconomic status? The case of low birthweight, *American Economic Review*, 89, 245-250.

Currie, J. and Moretti, E. (2003). Mother education and intergenerational transmission of human capital: Evidence from college opening, *Quarterly Journal of Economics*, 118, 1495-1592.

Davies, M., Baer, D., Judd, J., Brown, E., Campbell, W., and Taylor, P. (2002). Effects of moderate alcohol intake on fasting insulin and glucose concentrations and insulin sensitivity in postmenopausal women. *The Journal of the American Medical Association, 287*, 2559-2562.

Department of Health and Human Services and U.S. Department of Agriculture (DHHS). (2005). *Dietary Guidelines for Americans 2005*. Retrieved from www health.gov/ dietaryguidelines/dga2005/doucument /html/chapter9.htm.

de Walque, D. (2007). Does education affect smoking behavior? Evidence using the Vietnam draft as an instrument for college education, *Journal of Health Economics*, 26, 877-895.

de Walque, D. (2010). Education, Information, and Smoking Decisions: Evidence from Smoking Histories in the United States, 1940–2000, Journal of human resources, 45(3):682-717.

Fuchs, V. (1982). Time preferences and health: An explanatory study, In V. Fuchs, Editor, Economics Aspect of Health, second NBER conference on health in Stanford, University of Chicago Press, 83-119.

Fuchs, V. (1974). *Who Shall Live: Health, Economics, and Social Choice*, Basic Books Inc., New York.

Green, W. (2008). Econometric analysis, Pearson Education Inc., Upper Saddle River, NJ.

Grimard, F, and Parent, D. (2003). Education and smoking: Were Vietnam wardraft avoiders also more likely to avoid smoking? *Journal of Health Economics*, 26, 877-895.

Grossman, M. (1972). On the concept of health capital and the demand for health, *Journal of Political Economy*, 80, 223-225.

Grossman, M. (1973). Correlation between health and education, NBER Working Paper, No 22.

Gorssman, M. (2005). Education and nonmarket outcomes, NBER Working Paper, No 11582.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the U.S., *Review of Economic Studies*, 71, 189-221.

Joyce, T., Racine, A. and Mocan, N (1992).The consequences and costs of Maternel substance abuse in New York: A pooled time-series, cross-section analysis, *Journal of Health Economics*, 11, 297-314.

Kenkel, D. (1991). Health behavior, health knowledge, and schooling, *Journal of Political Economy*, 99, 287-305.

Kenkel, D. (1995). Should you eat breakfast? Estimates from health production functions, *Health Economics*, 4, 15-29.

Miller, R. (2001). Projection for college attendance by younger and older student, http//www. Greaterexpectations.org (Accessed Feb 4, 2011)

Parks, C. andKang, C. (2008). Does education induce healthy lifestyle? *Journal of Health Economics*, 27, 1516-1531.

Register, C., William, D., Grimes, P. (2001). Adolescent drug use and educational attainment, *Eductional Economics*, 9, 1-18.

Rimm, E., Klasksy, A., Grobbee, D., and Stampfer, M. (1996). Review of moderate alcohol consumption and reduced risk of coronary heart disease: Is the effect due to beer, wine or spirits? *British Medical Journal, 312* (7033), 731-36.

Ross, C., and Wu, C. (1995). The links between education and health, *American Sociological Review*, 60, 719-745.

Sander, W. (1995). Schooling and quitting smoking, *Review of Economics and Statistics*, 77, 191-199.

Wooldridge, J. (2002). Econometrics analysis of cross section and panel data, The MIT Press, Cambridge, MA.

Rosenzweig, M.R. (1995). Why are there returns to schooling? *American Economic Review*, 85, 153-158.

Rosenzweig, M.R., and Schultz, T. (1983). Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight, *The Journal of Political Economy*, 91, 773-746.

Tenn, S., Herman, D.A. and Wending, B. (2010). The role of education in the production of health: An empirical analysis of smoking behavior, *Journal of Health Economics*, 29, 404-417.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias, *Journal of the American Statistical Association*, 57, 348-368.

Table 1: Definition of Varibles

Variables	Variables type	Variable Definition
SMK	*Dummy*	Is smoking
MRJ	*Dummy*	Is using marijuana
COC	*Dummy*	Is using cocaine
DRK	*Dummy*	Is drinking alcohol
EDU	*Continuous*	Number of year of school attainment
WAG	*Continuous*	Hourly payment*10^{-4}
GEN	*Dummy*	Is Female
HIS	*Dummy*	Is Hispanic
BLK	*Dummy*	Is Black
MST	*Dummy*	Marital Status is “Married”
URB	*Dummy*	Lives in Urban Area
SMA	*Dummy*	Lives in Standard Metropolitan Statistical Area

Table 2: Descriptive Statistics

	1992		1994		1998	
Variables	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
SMK	0.324	0.468	0.326	0.469	0.306	0.461
MRJ	0.067	0.251	0.083	0.277	0.055	0.228
COC	0.011	0.106	0.013	0.115	0.008	0.090
DRK	0.630	0.482	0.596	0.490	-	-
EDU	12.848	2.451	12.910	2.451	13.046	2.462
WAG	16.706	39.492	12.259	9.941	14.765	15.983
GEN	0.495	0.499	0.495	0.499	0.495	0.499
HIS	0.157	0.364	0.157	0.364	0.157	0.364
BLK	0.250	0.433	0.250	0.433	0.250	0.433
MST	0.463	0.498	0.250	0.433	0.265	0.441
URB	0.816	0.387	0.805	0.395	0.700	0.457
SMA	0.778	0.415	0.812	0.390	0.800	0.399

Table 3: Robust Estimators Without Controlling Hetrogeneity

	Dependent Variables						
	SMK		*MRJ*		*COC*		*DRK*
Variables	1992-1994	1992-1998	1992-1994	1992-1998	1992-1994	1992-1998	1992-1994
EDU	-0.050**	-0.048**	-0.004**	-0.004**	-0.001**	-0.001**	0.009**
	(0.001)	(0.001)	(0.0009)	(0.0007)	(0.0003)	(0.0002)	(0.001)
WAG	-0.001**	-0.001**	-0.0004**	-0.0004**	-0.00005**	-0.0006*	0.0008**
	(0.0002)	(0.0003)	(0.0001)	(0.0001)	(0.00001)	(0.00002)	(0.0001)
GEN	-0.031**	-0.029**	-0.043**	-0.041**	-0.007**	-0.006**	-0.171**
	(0.008)	(0.006)	(0.004)	(0.003)	(0.001)	(0.001)	(0.008)
HIS	-0.0147**	-0.149**	-0.0.34**	-0.033**	0.001	0.001	-0.057**
	(0.010)	(0.008)	(0.006)	(0.004)	(0.002)	(0.002)	(0.011)
BLK	-0.072**	-0.070**	-0.037**	-0.036**	-0.002	-0.002	-0.142**
	(0.010)	(0.008)	(0.005)	(0.004)	(0.002)	(0.001)	(0.010)
MST	0.118**	0.119**	0.049**	0.050**	0.011**	0.011**	0.084**
	(0.008)	(0.007)	(0.005)	(0.004)	(0.002)	(0.001)	(0.009)
URB	-0.007	0.014	-0.001	0.007	0.005	0.002	0.064**
	(0.013)	(0.009)	(0.007)	(0.004)	(0.002)	(0.001)	(0.014)
SMA	0.028*	0.082	0.019**	0.013**	-0.0007	0.001	0.028
	(0.012)	(0.009)	(0.006)	(0.004)	(0.002)	(0.001)	(0.013)
Intercept	0.968**	0.944**	0.133**	0.127**	0.024**	0.020**	0.536**
	(0.024)	(0.0.20)	(0.013)	(0.010)	(0.005)	(0.003)	(0.026)

Note: Figures in parentheses are robust standard errors. * Significance of 5%, ** significance of 1%.

Table 4: Robust Fixed-Effect Estimators

	Dependent Variables						
	SMK		*MRJ*		*COC*		*DRK*
Variables	1992-1994	1992-1998	1992-1994	1992-1998	1992-1994	1992-1998	1992-1994
EDU	-0.0008	-0.002	-0.014	-0.005	-0.002	0.002	-0.021
	(0.0147)	(0.007)	(0.012)	(0.004)	(0.005)	(0.002)	(0.020)
WAG	-0.00008	-0.00006	-0.00006	-0.00004	-0.000008	-0.000005	0.0002
	(0.00009)	(0.00004)	(0.00009)	(0.00005)	(0.00003)	(0.00006)	(0.0002)
MST	0.0163	0.021*	0.018	0.020**	0.004	0.003	0.050*
	(0.0144)	(0.009)	(0.014)	(0.007)	(0.006)	(0.003)	(0.022)
URB	0.0257	0.016*	-0.034	0.004	0.002	-0.006	0.038
	(0.0297)	(0.008)	(0.028)	(0.006)	(0.010)	(0.002)	(0.038)
SMA	-0.0002	-0.001	0.013	0.009	-0.005	-0.001	-0.0008
	(0.019)	(0.014)	(0.014)	(0.010)	(0.006)	(0.004)	(0.030)
Intercept	0.297	0.320**	0.263	0.123	0.044	-0.020	0.865**
	(0.193)	(0.093)	(0.166)	(1.86)	(0.067)	(0.037)	(0.271)

Note: Figures in parentheses are robust standard errors. * Significance of 5%, ** significance of 1%.

Figure 1: Smoking Trend

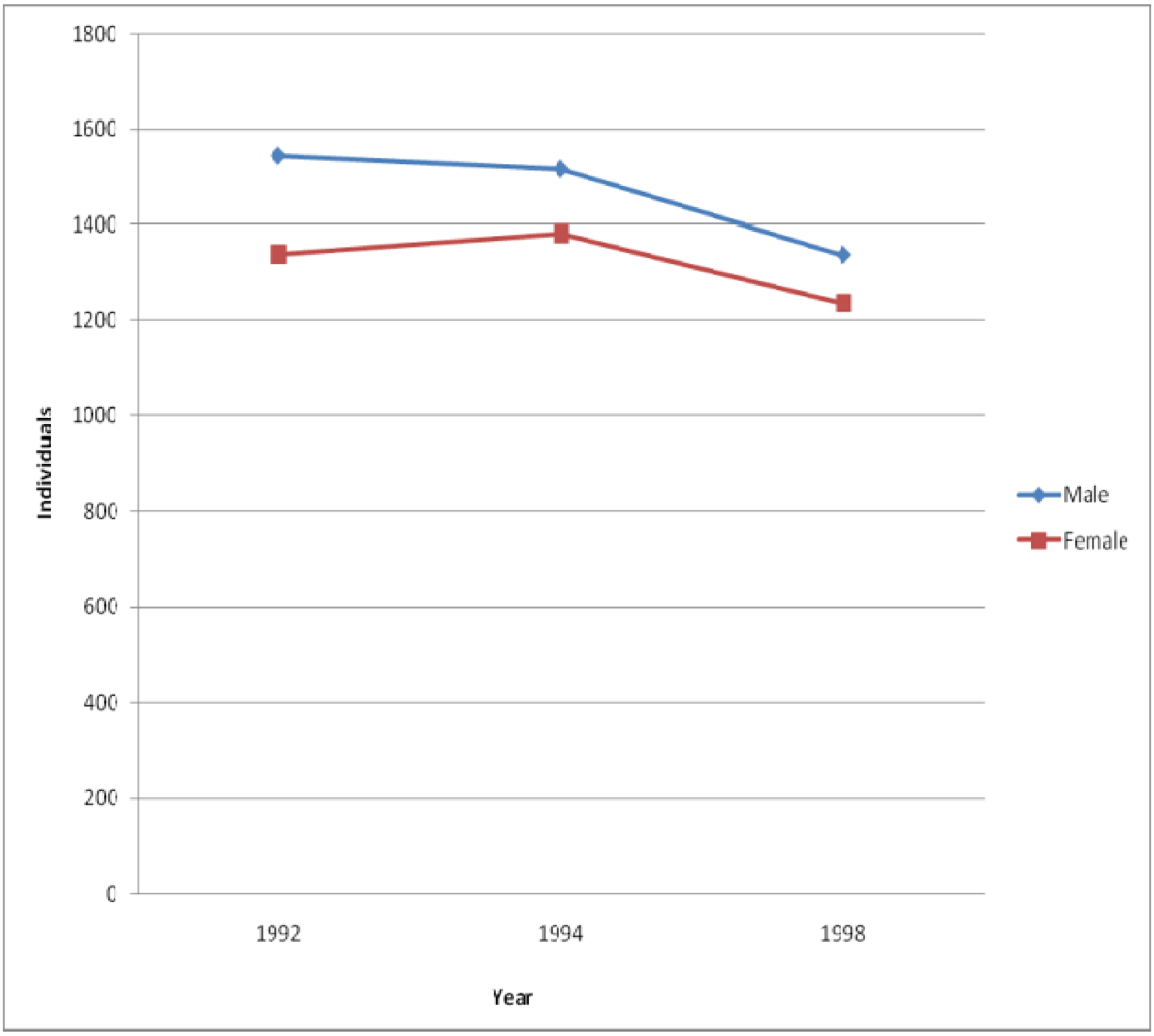

Figure 2: Marijuana Use Trend

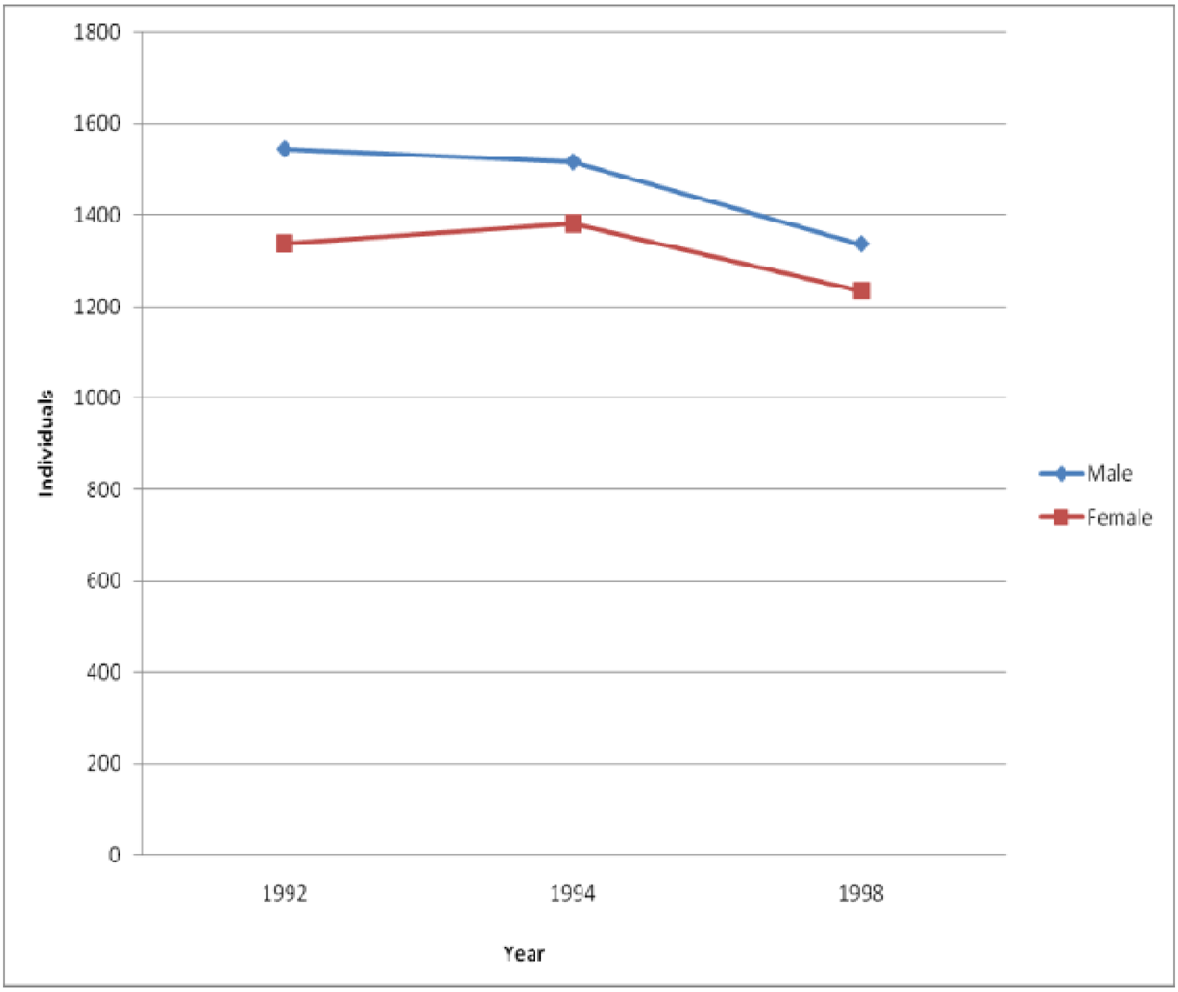

Figure 3: Cocaine Use Trend

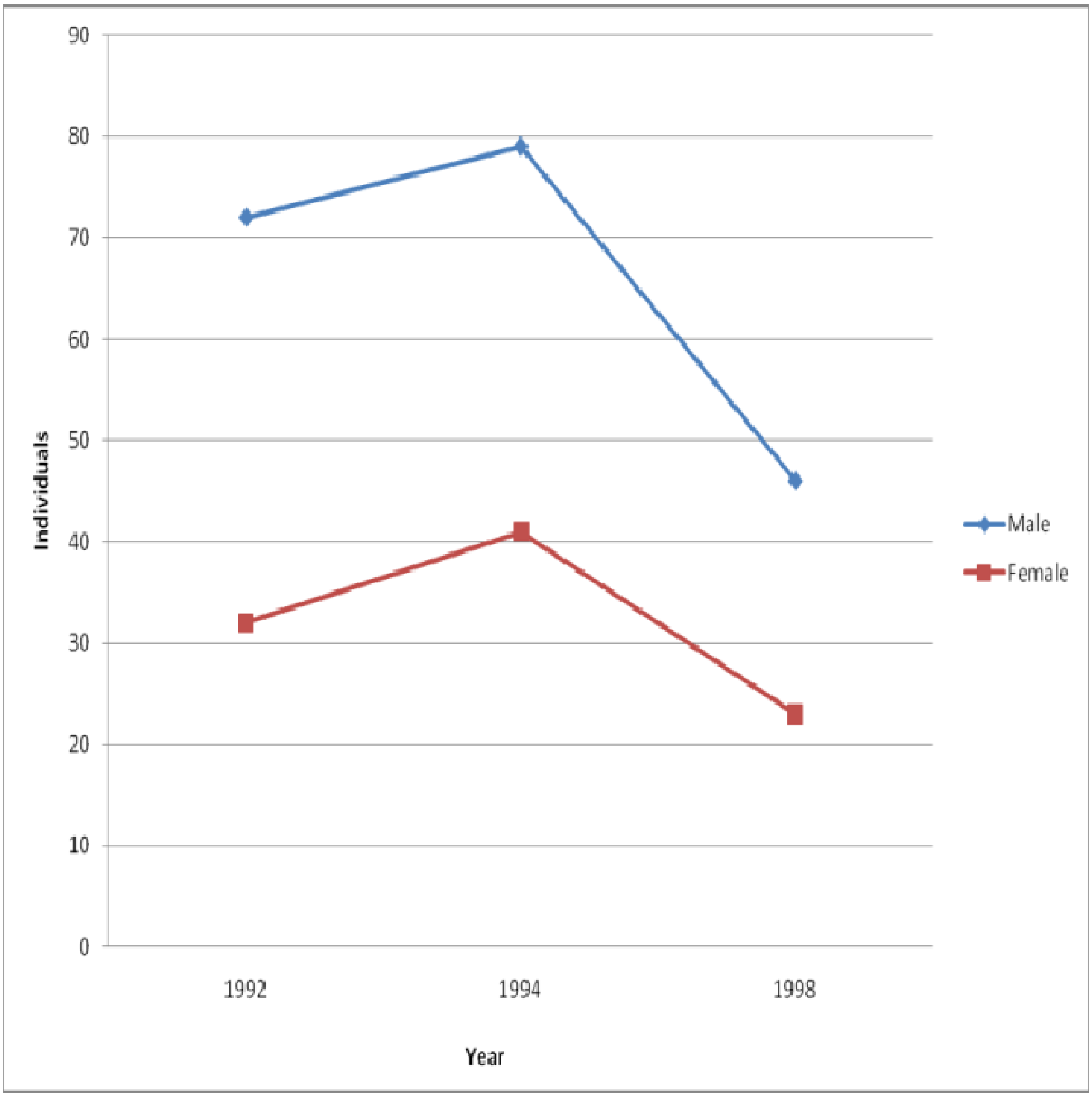
90
80
70
60
50
40
30
20
10
0
Individuals
1992
1994
1998
Year
Male
Female

Figure 4: Drinking Trend