THE WAGE EFFECTS OF PERSONAL SMOKING

MICHELLE RIORDAN
Senior Sophister

It is well established that smoking is bad for both your lungs and your wallet, but could it also affect your payslip? Michelle Riordan conducts a rigorous econometric analysis to determine the effect of smoking on one's wage. The results show that the price of a pack of cigarettes is not the true cost of smoking, but instead one must also account for a considerable wage penalty.

Introduction

Tobacco smoking is a topic that has generated a lot of discussion in recent decades. Since the release of the 1964 Surgeon General report asserting that smoking has adverse effects on health, causing cancers, respiratory disease etc., it has now been widely established that smoking leads to a huge amount of pain and suffering being inflicted on smokers themselves, those exposed to cigarette smoke, and society at large (Levine et al., 1997). While numerous price increases, anti-smoking campaigns, and policies, such as the smoking ban, have in the past reduced the number of smokers considerably, the effectiveness of these has weakened in recent years (Murphy, 2007). This study will attempt to show the wage penalty borne by smokers, which thereby acts as a disincentive to smoke. Using a Two-Stage Least Squares estimation method, I find that current smokers are subject to an approximate 1.95% wage penalty compared to their non-smoker counterparts.

Literature Review

A vast amount of econometric literature exists on investigating the determinants of wages, concluding that factors such as education, experience, health, innate ability, gender, age, and marital status all have a significant effect on an individual's wage, most noteworthy; Angrist & Kruegar (1991), Blackburn & Neumark (1992), and Mroz (1987). However, in comparison, the effect of smoking on income has received relatively little attention in the literature. The general consensus from those who have undertaken such studies is that smoking has a statistically significant negative effect on wages. However, the magnitude of this negative effect is subject to much discussion in the literature, with estimates of the wage penalty borne by smokers ranging from 0.5%-24%. These studies test several hypotheses regarding this wage penalty.

Firstly, there is evidence that smokers are less productive than their non-smoker counterparts, which may translate into lower wages. Basic microeconomic theory suggests that the wage of an individual is related to his/her marginal productivity, with a low wage implying a lower marginal product of labour (Arrow, 1973). Authors such as Levine, Gustafson & Velenchik (1997), Van Ours (2007) and Grek (2007) examine this link between productivity and smoking. They report that a smoker's productivity may be lower than a non-smoker for two reasons:

1. "Smoking may reduce net workers' productivity by interfering with workers' ability to carry out manual tasks"
2. "Smokers' productivity would be lower if the act of smoking itself draws time away from work"

Therefore, all reach the conclusion that that there is a statistically significant wage differential between smokers and non-smokers, where the act of smoking is believed to reduce an individual's wage by way of lower productivity levels.

Secondly, some studies also explore the notion that smokers tend to earn less as smoking is adversely related to health. Grossman (1972) found that wages and health are positively related. Using this fact and since smoking has clearly negative effects on an individual's health, some authors have concluded that it may be the case that the use of tobacco has negative effects on an individuals wage (e.g. Grek, 2007 & Braakman, 2008). One argument is that current smokers have significantly greater absenteeism than those who had never smoked, with former smokers having intermediate values. They suggest that this is because smokers on average miss 6.16 days per annum due to sickness (including smoking-related acute and chronic conditions), compared to non-smokers who miss 3.86 days of work per annum (Halpern et al, 2004).

Third, Grafova & Stafford (2009) explore whether former smokers are also subject to this wage penalty. They found statistically significant wage gaps between smokers who would continue to smoke and three other groups: those who would later quit smoking, those who had already quit smoking and those who had never smoked. They found that the wage penalty was the highest for those who would continue to smoke and was almost negligible for those who had already quit smoking.

Finally, there may be unobserved factors related to both the decision to smoke and lower wages. Becker et al. (1988 & 1994) and Munasinghe & Sicherman (1999) argue that smoking may reflect a higher time preference rate, as smoking may provide utility today, with the adverse effects generally occurring later in life. A higher discount rate is a key determinant of an individual's investment in human capital and occupational choice, where individuals who have higher discount rates tend to be less future-oriented. They come to the conclusion that smokers are less future-oriented and are less likely to invest in human capital and, therefore, are more likely to select careers with lower and flatter earnings profiles compared to non-smokers. Other possible unobservables may include preferences for work/leisure time, innate ability, intrinsic motivation, desire to succeed, etc. Therefore it is not smoking per se that is causing the lower wages but the effect of the unobservables.

The vast majority of the studies listed above have used U.S. or Eastern-European data to test their various hypotheses. My major innovation in this study is to investigate whether the hypothesis that smoking can reduce an individual's wage is true for Irish/British data. Since there can be vast differences between the social and cultural backgrounds of individuals from the US, Eastern-Europe and Western-Europe, it will be interesting to find whether this hypothesis also holds for Irish/British data.

Econometric Model

In order to quantify the hypothesized relationship above, the following regression was estimated:

$$Yi=\beta 0+\beta 1X1+\beta 2X2+\beta 3X3+\beta 4X4+\beta 5X5+\beta 6X6+\beta 7X7+\beta 8X8+\beta 9X9+ui$$

Where:
Yi=Ln(wage): The log of real monthly labour income
X1=Educ: The number of years of schooling the individual has
X2=Total Exper: The total number of years of work experience an individual has
X3=Current Exper: The number of years of experience an indv. has in their current job
X4=Female: A dummy variable equal to 1 if the person is female
X5=Marr: A dummy variable equal to 1 if the person is married
X6=Age: Age of the person in years
X7=Smoke: A dummy variable equal to 1 if the person smokes
X8=Hhsize: No. of individuals currently living in the same residence as the individual
X9=Hours: Usual number of hours worked per week
ui=An error term of statistical residuals

A positive effect is expected between Ln(wage) and Educ, TotalExper, CurrentExper, Marr, Age, and Hours. Conversely, a negative effect is expected between Ln (wage), Female, Hhsize, and Smoke. The inclusion of the log of wage will enable a semi-elasticity interpretation of the coefficients, i.e. the percentage change in wage given one unit increase in the independent variable is approximately given by:

$$\%\Delta\widehat{wage} \approx 100(\Delta\widehat{\beta}_j) \tag{1}$$

Using simple algebraic properties of the exponential and logarithmic functions, we can find the exact percentage change in the predicted wage (Wooldridge, 2009). This is given by:

$$\%\Delta\widehat{wage} = 100[\exp(\widehat{\beta_j}\ \Delta x_j) - 1] \tag{2}$$

Potential Problems

It is highly likely that factors such as ability present in the error term will not be constant across observations. Due to this likely presence of heteroscedasticity in the model, robust standard errors will be used when estimating the model. This is because although heteroscedasticity does not cause bias or inconsistency in the OLS estimators of the coefficients, homoscedasticity is required to perform the standard t and F tests. In addition, the possible issue of endogeneity in wage equations has been highlighted in the literature and will therefore also be an issue in this case. It is commonly thought that education is correlated with the error term through unobserved ability, i.e. Cov(xi,ui)≠0. Failure to correct for this would result in a violation of the Gauss-Markov zero-conditional mean assumption leading to biased estimates, as OLS would incorrectly estimate the effect of education on an individual's wage. If the education variable is not exogenous, Griliches (1977) proposes the use of the instrumental variable (IV) method to tackle the problems of ability bias and endogeneity, i.e. finding a variable to instrument for education in the regression model can rectify this problem. It can be difficult to find instruments though. The use of an IV to estimate the return to education requires that the instrument satisfies the instrument relevance and instrument exogeneity conditions, i.e. an IV for education must be uncorrelated with ability (and any other unobservable factors affecting wage) and highly correlated with education (Chaung & Lai, 2010). Empirical studies have shown that more siblings are associated with lower average levels of education. Moreover, given a family's budget constraint, the greater the number of siblings there are, the smaller the educational resources that are available to each child, leading to those in larger families having lower average levels of education. I will therefore use the number of siblings as an IV for the number of years of schooling an individual has, as it will be correlated with an individual's educational achievement but have no correlation with an individual's ability.

Data

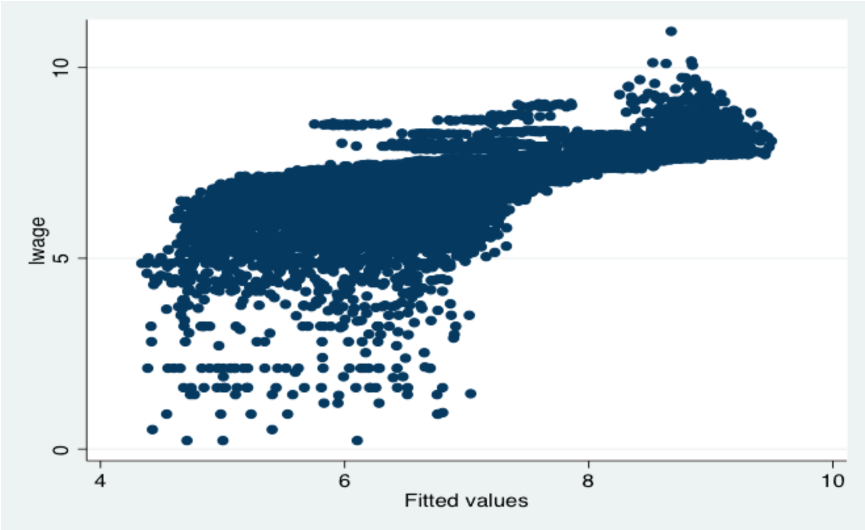
It was originally intended to use Irish data to conduct this analysis. However, after dropping cases with missing values, it was felt that the sample size was too small for this purpose. To overcome this problem, this paper will use data from the year 2009 from the British Household Panel Survey (BHPS), an annual survey carried out by the ESRC UK Longitudinal Studies Centre within the Institute for Social and Economic Research at the University of Essex (See: http://www.iser.essex.ac.uk/uisc/bhps/). For the purpose of this paper, we will focus of those of working age, i.e. those between the ages of 16-65. After dropping the cases with missing values and the cases where the individual was over 65, we arrive at a sample size of 10,344.

Variable	Observations	Mean	Std. Dev.	Min	Max
Wage	10344	1652.111	1561.319	1.25	56916.67
Ln(wage)	10344	7.014514	1.064298	0.2231435	10.94934
Educ	10344	10.49903	5.198561	0	20
TotalExper	10344	14.72322	8.552597	0	47
CurrentExper	10344	6.082367	7.197345	0	47
Female	10344	0.5446636	0.4979729	0	1
Marr	10344	0.5554911	0.4969352	0	1
Age	10344	40.90874	13.69378	16	65
Smoke	10344	0.2585073	0.4378354	0	1
Hhsize	10344	3.099961	1.39302	1	16
Hours	10344	22.6162	17.91157	0	98
Sibs	10344	3.29679	2.397121	0	9

Table 1: Summary statistics

Results

For comparison purposes, we first estimate the econometric model specified above by using the conventional OLS method with robust standard errors, noting that these estimates will be biased as the zero-conditional mean assumption will be violated. The results of this regression are shown in Table 2. We can see that all of the coefficients have the expected effect, with all coefficients being statistically significant except for current work experience. This is not a surprising result as we would expect an individual's total work experience to have a much greater effect on their wage compared to the number of years they have been in their current job. In addition, the R2 is 0.5789, implying that the model explains 57.89% of the variation in the dependent variable. From Graph 1, we can see that the model seems to explain average levels of income quite well. However, we can also see that at lower income levels, there are several significant differences between the fitted values and the actual wage of an individual. This in turn suggests that the model fails to explain 42.11% of the variation in wages, which indicates that a significant proportion of an individual's wage is determined by 'unobservables'. As already mentioned, taking the log of the dependent variable enables a semi-elasticity interpretation of the coefficients. Most noteworthy, from Table 2 we can see that if an individual smokes, on average their wage will be approximately 11.26% lower than non-smokers. Moreover, we can also see that the most common variables used in econometric wage equations, such as education, total work experience, gender, marital status, and age, all have signs consistent with previous empirical works.



Graph 1: Fitted Values of OLS Regression

Secondly, to overcome the endogeneity problem, a Two Stage Least Squares (2SLS) regression was conducted using the number of siblings as an instrument for education. For the 2SLS estimation, the variable "current work experience" was omitted, as it was statistically insignificant at any conventional statistical significance level in the first OLS estimation. Not surprisingly, the R2 fell to 0.2597, implying that 25.97% of the variation in the dependent variable is explained in the model. However, the fundamental goal of IV estimation is to correct for any endogeneity problem so that the estimates are unbiased and consistent and not solely to maximize the 'goodness of fit' (Wooldridge, 2009). From Table 2, we can see that the IV estimates are of the same direction as the OLS coefficients but several of the coefficients differ greatly in both magnitude and statistical significance. Most relevant for this paper, the coefficient on smoke is still highly statistically significant but of a much lower magnitude, with smoking reducing an individuals wage by approximately 1.95%. Furthermore there are several interesting differences between the OLS and IV estimates which include:

- The coefficient on female is now only statistically significant at the 10% level, which is quite a surprising result.
- The coefficient on current household size has become statistically insignificant.
- The IV coefficient of the return of work experience has almost tripled compared to the OLS coefficient.

Variable	OLS Coefficient (with robust std. errors	IV Coefficient (educ = sibs)
Educ	0.0745851*** (0.0016072)	0.2216247*** (0.017022)
TotalExper	0.0249403*** (0.0009600)	0.068212*** (0.0170711)
CurrentExper	0.0002228 (0.0011475)	–
Female	-0.0622505*** (0.0146029)	-0.0454236* (0.0241888)
Marr	0.0393872*** (0.016700)	0.0880125*** (0.0260882)
Age	0.0109432*** (0.0007805)	0.0070866*** (0.0009921)
Smoke	-0.1125643*** (0.0178894)	-0.0195109*** (0.00247509)
Hhsize	-0.0156371*** (0.0054907)	-0.0009336 (0.00719760
Hours	0.0201504*** (0.0005099)	0.0060065*** (0.0017096)
Constant	5.049086*** (0.0476841)	4.341796*** (0.1013211)
R^2	0.5789	0.2597

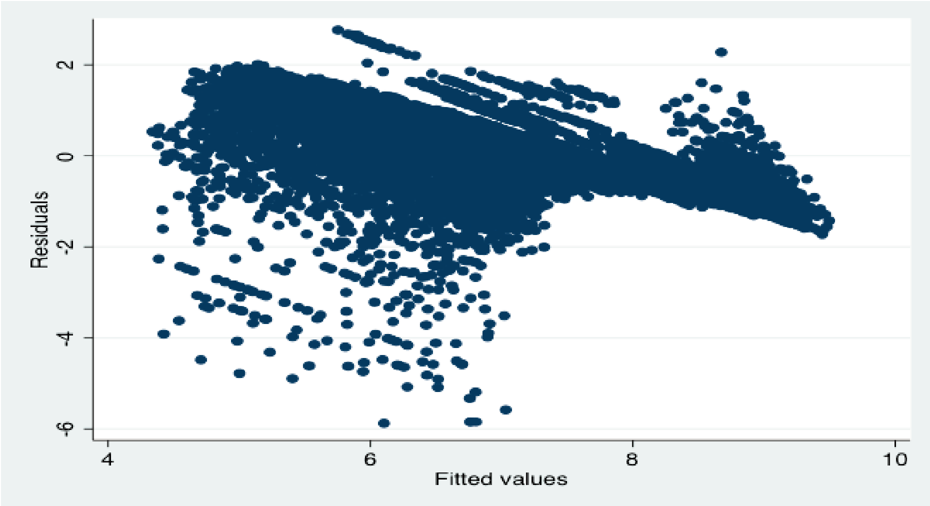
Table 2: Regression Results

Diagnostic Checks and Testing

Heteroscedasticity

If a model is well fitted, there should be no evident pattern to the residuals plotted against their respective fitted values (Wooldridge, 2009). Graph 3 plots the fitted values against the residuals from the OLS regression without robust standard errors. From this graph, we can see that the middle section and final section of the graph is not particularly scattered and, as expected, we can conclude that heteroscedasticity will be an issue in this model.

Furthermore, to confirm the presence of heteroscedasticity in the model, a Breusch-Pagan/Cook-Weisberg test for heteroscedasticity was conducted. The p-value (Prob>chi2=0.0000) confirms that heteroscedasticity will be an issue in this model. Similar tests for heteroscedasticity were conducted on the OLS model and IV model, both with robust standard errors. Both tests returned no evidence of heteroscedasticity, implying that the models and results presented in Table 2 will be valid for statistical testing.



Graph 3: Residuals vs. Fitted Values from OLS Regression without Robust SE

Omitted Non-Linear Variables

To investigate whether there are any non-linear combinations of the variable not presently in the model, which have predictive power, a heteroscedasticity-robust Ramsey RESET test was conducted. A statistically significant p-value suggests that there are relevant explanatory variables omitted from the model. This test was conducted on the OLS (with robust standard errors) and IV models. The Ramsey RESET test returned a p-value=0.000, implying that there are no non-linear combinations of the variable, not presently in the model, which have predictive power.

Endogeneity

Due to the likely presence of endogeneity in the model, a Durbin-Wu-Hausman test for endogeneity was conducted on the OLS model. As expected, the test confirmed the presence of endogeneity. Furthermore, the residuals were backed out from the OLS regression and the correlation between education and the squared residuals was calculated. The computed correlation between the two variables was -18.6. This suggests that the education variable is in fact endogenous and thus, the return to education estimated by OLS presented in Table 2 will be biased and inconsistent.

Possible Extensions

Firstly, this study focuses on individuals who currently smoke. It may be of interest to explore whether those who are currently trying to quit smoking and ex-smokers are also subject to this wage penalty. It was shown by Grafova & Stafford (2009) that the wage penalty for those who intended to quit smoking was much smaller than those who had no intention to quit and was almost negligible for ex-smokers. Due to a lack of available data, such a study was not possible in this analysis. Secondly, it also may be instructive to add cultural variables to the model. Adding such factors will enable a better understanding of the social and cultural background of individuals and may also act as a proxy for some of the 'unobservables' currently present in the error term.

Conclusion

The central goal of this analysis was to show that a wage penalty is borne by current smokers. To achieve this, I used cross-sectional data from the BHPS. As the model suffered from heteroscedasticity and endogeneity, a Two-Stage Least Squares method was used. As already mentioned, the possible issue of endogeneity in wage equations has been highlighted in the literature, with education being correlated with the error term through unobserved ability. This required the use of an IV to estimate the return to education. The number of siblings an individual has was used as an IV as the instrument satisfies the instrument relevance and instrument exogeneity conditions. After correcting for heteroscedasticity and endogeneity, I found that smokers suffer a wage penalty of approximately 1.95%. Although this is at the lower end of the estimates of the wage penalty borne by smokers, this study should act as a disincentive to smoke.

References

Arrow, K., 1973. The theory of discrimination, in Orley C. Ashenfelter and Albert Rees, eds. "Discrimination in labor markets", Princeton University Press, Princeton, pp.3-33.

Auld, C.M., 2004. Smoking, Drinking & Income, University of Calgary

Becker, G.S., 1957/1971. The economics of discrimination, 2nd edition (1971), University of Chicago Press, Chicago.

Becker, G.S. & Murphy, K.M., 1988/1994. A Theory of Rational Addiction, Journal of Political Economy, University of Chicago Press, vol. 96(4), pp.675-700, August.

Braakmann, N., 2008. The smoking wage penalty in the United Kingdom, Leuphana University Lüneburg.

Chaung, Y. & Lai, W., 2010. Heterogeneity, Comparative Advantage, and Return to Education: The Case of Taiwan, Economics of Education Review, Oct. 2010.

Graffova, I. & Stafford, F., 2009. The Wage Effects of Personal Smoking History, ILR Review.

Grek, J., 2007. The Effect of Smoking and Drinking on Wages in Sweden.

Griliches, Z., 1977. Estimating the Return to Schooling: Some Econometric Problems, Econometrica, 45(1), pp.1-22.

Halpern M.T, Shikiar R, Rentz A.M. & Khan Z.M., 2004. Impact of smoking status on workplace absenteeism and productivity, 22(3), pp.219-29.

Levine, P.B., Gustafson, T.A. & Velenchik, A.D., 1997. More Bad New for Smokers? The Effect of Cigarette Smoking on Wages, NBER Working Paper 5270.

Murphy, A., 2007. An Econometric Analysis of Smoking Behaviour in Ireland, iHEA 2007 6th World Congress: Explorations in Health Economics Paper.

Munasinghe, L. & Sicherman, N., 1999. Why Do Dancers Smoke? Smoking, Time Preference, and Wage Dynamics"

Wooldridge, J.M., 2009. "Introductory Econometrics, A Modern Approach", 4th edition, South-Western Cengage Learning, Canada.

Van Ours, Jan C., 2004. A pint a day raises a man's pay; but smoking blows that gain away, Journal of Health Economics, 23(5), pp.863-886.