Journal of Economics Bibliography
www.kspjournals.org
Volume 3 March 2016 Issue 1

Effect of Smoking and other Economic Variables on Wages in the Euro Area

By Eric A. BONDZIE [†]

Abstract. The smoking effects on wages has been examined in this work using different econometric methodologies with the use of European Community Household Panel (ECHP) datasets. We employ econometric tools like Instrumental Variable technique, Heckman correction factor, Endogenous Switching and matching estimates. The initial results from regression estimates (OLS and IV methods) revealed that the wage gap between smokers and non-smokers ranges 1% to 22.7%. Moreover, endogenous switching and matching estimator also showed a negative average treatment effect of approximately 47% and 4.3% to 6.9% respectively. Thus smokers observed less wage effects is explained in part by real effects on their health status and a measure of unobserved preferences.
Keywords. Smoking, Wages Differentials, Econometric Approaches and Euro Area.
JEL. C18, I12, J31, O51.

1. Introduction

A causal relationship between smoking and coronary heart disease was reported at Mayo Clinic in 1940 and the release of the 1964 Surgeon General's report asserting that smoking causes cancer and other serious diseases, both on smokers and on others who are exposed to cigarette smoke. Since that time several studies have been done to ascertain the various cost imposed by smokers on themselves and their surroundings.

Smoking has been shown to be the leading cause of lung cancer, chronic bronchitis, and emphysema, as well as a major cause of heart disease and stroke. It is associated with a variety of other conditions, including slowed healing from injuries and increased susceptibility to some infections (Napier, 1996; Blake et al.1988).

Individual's wage on the other hand is directly related to his or her marginal productivity, thus a low wage implies a lower marginal product of labour as compared to the marginal product for a worker with a higher wage. The marginal product of labour also relates to the level of education and how long the worker has been employed. The relationship was first introduced in Becker (1964) called the human capital model. Becker found that human capital has a positive effect on wages. Also related to the marginal product of labour are health issues such as smoking and Alcohol drinking. Grossman (1972) concluded that wages and health are positively related. And since smoking has a negative effect on an individuals' health, it may be the case that smoking has a negative effect on wages.

1.2. Motivation and Rationale

[†] Università Cattolica del Sacro Cuore-Milan, Department of Economics and Finance. 20123, Italy.
☎.
✉. bondzie2g@yahoo.com

Smoking is a health problem, the costs of which include sickness, pain, grief and misery. From a behavioural economics perspective, smoking seems to have an adverse effect on wages and imposes a significant economic burden on society.

In addition to the direct medical costs of treating smoking-induced illnesses there are other indirect costs including loss of productivity, fire damage and environmental harm from cigarette litter and destructive farming practices. The total burden caused by tobacco products more than outweighs any economic benefit from their manufacture and sale.

Smoking has been shown to decrease life expectancy and increase health care utilization and expenditures. The U.S.Centre for Diseases Control and Prevention estimates that health care expenditures attributable to smoking were over $95 billion per year in the period 2000-2004 (Adkihari et al. 2008). However, there are other costs associated with cigarette smoking besides poor health and smoking-attributable health care expenditures.

There are analyses both on the individual and the aggregate, public health level. The latter is a major policy concern as health expenditures in Western industrialized countries, with only a few exceptions, have constantly increased in the last decades (Figure 1.1).

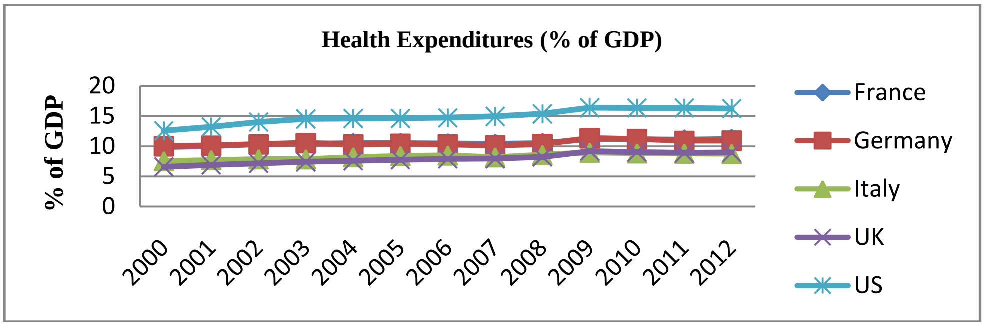

Figure 1. *Health expenditures (% GDP) in selected OECD countries*
Source: *OECD Health Data, (2014).*

The determinants for this development are of socio-political interest; this paper will concentrate on the micro-level and focus on one particular aspect of individual health behaviour, smoking, and its relationship to wages and certain economic outcomes.

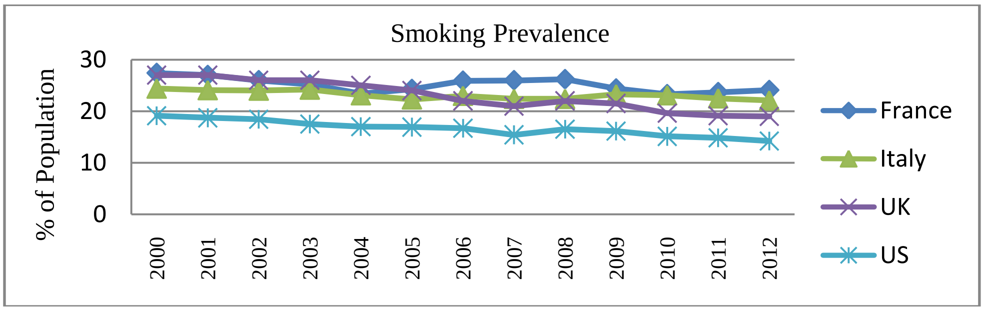

Figure 2. *Smoking prevalence (% of population) in selected OECD countries*
Source: *OECD Health Data, (2014).*

The crucial point is that people decide on whether to smoke or not, although individuals have adequate knowledge about the adverse health effects that are attributed to smoking. Despite the widespread knowledge of smoking and its negative consequences on individuals´ health, smoking is still a prevalent phenomenon in Western industrialized countries. While there is variation across countries, Figure 1.2 above shows that the rate at which individuals smoke are still on the increase.

The aim of this paper is to analyse data over a long panel period to understand better the relationship between smoking and wages, economic outcome and to evaluate possible explanations.

2. Literature Review

2.1. Theoretical Framework

In recent times both labour and health economist have examined the relationship between labour output and earnings vis-à-vis their marginal product. Whiles labour economist relates earnings to marginal product of labour (MPL), health economists on the other hand link MPL to labour's ability and health. A wide range of research for example Heijdra & Van der Ploeg (2002) and Burda & Wyplosz (2005) has found that the wage an individual receives is related to his/her marginal productivity, i.e. a low wage implies a lower marginal product compared to the marginal product for a worker with a higher wage and vice versa. As it is mostly argued in labour economics literatures, when marginal product of labour (MPL) is equal to the real wage (w), i.e. when MPL = w; firms are assumed to be maximizing their profit. The marginal product of labour is calculated as:

$$MPL{*}P = W(\text{Nominal wage}) \Rightarrow (MPL*P)/P = W/P \Leftrightarrow MPL = w(\text{Real wage})$$

Moreover, as indicated above MPL is also related to health (smoking and Alcohol drinking) and abilities of labour which means that wage is directly linked to these factors. That is for individuals that appear to be in a good health it is implied that they have a higher marginal productivity relative to individuals that don't seem to be in good health. As the aims of this paper suggest, attention will be focused on the relationship between wage and individuals' health in this case smoking. This relationship has been studied under different theories like the human capital model and the theory of efficient wages. The human capital model was first introduced in Becker (1964), where he shows that human capital (education) has a positive effect on wages, i.e. the marginal product of labour is also related to the level of education and how long the worker has been employed. Hence, the marginal product increases with skills.

The theory of efficient wages is based on the hypothesis that the net productivity of workers is a function of the wage rate they receive. The theory postulates that workers' productivity depends on the level of nutrition, high labour turnovers (increased training cost for new workers), information asymmetry in the labour market (about the characteristics of workers) and whether workers feel being treated fairly equal with their pers (Stiglitz, 1986).

2.2. Empirical Review

The explanations to wage differentials are usually point to health and productivity effects. Smoking may reduce productivity either by taking employees away from their job or by making them to absent themselves from work due to illness. The reduced productivity due to smoking-related health effects takes both readily observable forms, such as more frequent absences (Bertera, 1991; Kristein, 1983), and forms that are more difficult to quantify, such as lower physical and mental endurance. Kristein (1983), drawing together evidence from a number of studies, estimated the productivity costs of smoking to be between $80 and $160 per smoker per year, measured in 1980 dollars. Absenteeism by smokers, he argued, imposed an additional $40 to $80 in costs per smoker per year. Bertera (1991) also echoed these findings by arguing that smokers on average miss one additional work day per year due to illness, controlling for other factors such as education and age. From sociological and psychological viewpoints (Freize et al. 1991; Martel & Biller, 1987), wage and smoking relationship is based on discrimination; relative to smokers, non-smokers may tend to receive favourable treatment in the workplace. Smoking may, for example, adversely affect physical attractiveness, whether visual (notably, through skin damage, an effect of smoking found in many clinical studies) or olfactory (because of tobacco smoke's stale smelling residue). Discrimination could also be invoked to explain lower compensation for smokers, as a result of their poorer health; they incur greater benefit provision costs than non-smokers, and employers respond by imposing a negative compensating wage differential. Recent publicity about the effects of second-hand smoke makes it likely that both co-workers and customers may object to working with smokers, causing some employers to discriminate against them.

These effects might not be the same in the case of young workers since major health effects of smoking generally appear late in life, so one might be concerned that smoking would not have an impact on the current health and labour market outcomes. Several studies indeed show a negative effect of smoking on wages of young adult (Conway & Cronan, 1992; Hoad & Clay, 1992).

Several empirical works have documented the relationship between wage and smoking in different ways. For examples, Levine et al (1997), Auld (1998), Lee (1999), Grafova & Stafford (2005), Braakman (2008), and Anger & Kvasnika (2010) all examine the relationship between smoking and wages and found a consistent evidence of a negative relationship. Empirical studies reviewing the relationship between smoking and wages have found a differential in favour of non-smokers in the range of 2–10%, depending on data source, time, and country. Using the 1973Quality of Employment Survey (QES) to obtain point estimates of the earnings gap, Leigh & Berger (1989) reported a statistically insignificant differential of 1.5–3.5%. Levine et al. (1997) found that smoking reduced wages by roughly 4.2% and 6.9%, respectively, in 1984 and 1992 data samples from the National Longitudinal Survey of Youth (NLSY). International evidence is consistent with U.S. evidence. Levine et al. also suggested that the lower wages for smokers is due to such issues as employer discrimination, increased costs of employing smokers, or lower productivity by smokers.

Lee (2003) reported a 5% wage gap based on data from the Australian Twin Registry of 1980–82 and 1988–89. Heineck & Schwarze (2003) examine the effect of smoking on wages in Germany. The empirical results when using cross-sectional models show that smoking has a negative effect on wages. The empirical results when using fixed-effect estimation show that there is a positive effect on wages for males, while there is no such effect for females. van Ours (2004) reported a 10% wage gap between smokers and non-smokers using the Dutch 2001 CentER data; Van Ours estimated these relationships separately for men and women and found that the wage penalty is driven by the negative effect on men's wages as no wage penalty was found for female smokers, at least in The Netherlands (van Ours 2004). Auld (2005) found an 8% wage gap using the 1991 Canadian General Social Survey.

This work tries to decompose wage differential between smokers and non-smokers, across a range of criteria for smoking status used to gain a further understanding into the share of the wage differential that is attributed to selection into smoking. The paper will also examine the impact of the choice of the smoking status criteria, including how to capture smoking intensity (i.e., number of cigarettes consumed). Understanding the impact of smoking at different levels of intensity will aid in the interpretation of the results. As indicated above, several studies have been carried out to reveal the relationships between the subject of discussion and more of those will be discussed in the subsequent sections.

3. Data and Econometric Methodology

3.1. Data Analysis

The research uses the European Community Household Panel (ECHP) dataset; ECHP data is a harmonised cross-national longitudinal survey focusing on household income and living conditions. It also includes items on health, education, housing, migration, demographics and employment characteristics.The survey is made up of 8 waves which run from 1994 to 2001. In the first wave (1994) a sample of some 60,500 households i.e. approximately 130,000 adults aged 16 years and over were interviewed across 12 member states (Belgium, Denmark, Germany, Greece, Spain, France, Italy, Ireland, Luxembourg, The Netherlands, Portugal, the United-Kingdom). For most of the countries the surveys were carried out using the harmonised ECHP questionnaire. For some countries the institutes in charge of the production of the ECHP converted national data surveys into ECHP format to replace the ECHP from 1997 onwards. In these waves, information on life course smoking behaviour was reported, as well as for their marital status if they were married or not.

In the health module of the survey, individuals were asked whether they currently smoked or have ever smoked. Current smokers were further asked about their average daily cigarette consumption and other smoking related products (pipes and cigar) that they smoke. These questions enable us to construct smoking histories that include average daily cigarette consumption and duration of smoking.

3.1.1. Descriptive Statistics

The following section presents the characteristics of the data used and description of the variable used in the analysis. The data is limited to missing and non-applicable responses; thereby these observations were removed to avoid discrepancies in the data and the analysis. After a careful sorting-out, the sample was reduced to 41,896 observations. Table 1 gives a short presentation of the basic characteristics of the variables used in this study in the form of descriptive statistics, such as means, standard deviations, minimum and maximum values for the variables of interest. From table 1, age variable indicates a mean age of 38 years with standard deviation of 11.13966years which means that age ranges from 15-65 years of age. The wage variable on the other hand shows a mean of 55200.7with a standard deviation of 103115.1 in respective national currencies. The hourly wage also has a mean of 371.2758and a standard deviation of 708.6884per hour. The data also revealed a mean weekly working hours of 38.3 with a standard deviation of 9.393041 which means individuals work at least 2 hours a week and at most 96 hours. The mean weight for individuals in the sample is 70.4 with standard deviation of 13.52344.The rest of the variables were treated as dummies and their respective means and standard deviations are shown in table 1.The specific variables of interest included in this study is based on variables that is used frequently in previous studies and labour wage theories or is assumed to have specific effect on individual's wage and also due to the researchers own decisions and interest. For instance, age of an individual is assumed to have a positive effect on the wage, which means as individualsgrow; they earn higher wages.

Table 1. *Description of Variables*

Variables	Description	Mean	Std. Deviation	Minimum	Maximum
Age	Age of the respondent in years	38.01036	11.13966	15	65
Age Squared	Age squared	1568.877	879.8573	225	4225
Gender	1 if the respondent is male	0.4764178	0.4994495	0	1
Current Married	1 if the respondent is currently married	0.6115858	0.4873954	0	1
Education	1 if respondent works in the Education sector.	0.1372923	0.3441598	0	1
Hours worked	Hours worked per week	38.29805	9.393041	2	96
Weight	Respondents weight without clothes and shoes.	70.3654	13.52344	12	180
Admitted to Hospital	1 if the respondent has been admitted to hospital in the past 12 months	0.0567357	0.2313397	0	1
Wage	Gross monthly wage	55200.7	103115.1	90.15182	2500000
Hourly Wage	Gross hourly wage	371.2758	708.6884	0.5977511	12500
Log hourly wage	Log hourly wage	4.05171	2.146303	-0.5145808	9.433484
Smoke	1 if the respondent currently smokes daily	0.2955891	0.4563126	0	1

Individual's level of education is also assumed to have a positive effect on wages, due to the assumption that the marginal product of labour increases with skills as was discussed above. Years of education is also used as a measurement of human capital. Being married or in a marriage equivalent relationship is assumed to have a positive effect on wages. The health variables included in this study is smoking and whether the respondent has been admitted to a hospital in the past 12 months. To ascertain individual's smoking behaviour, the weight of respondents was added; since it is a well-known fact that smoking affects smoker's weight.

Based on empirical results available and the relationship between wages, marginal productivity of labour and their link to individual health, smoking is also assumed to have a negative effect on individual's wage. These evidences are shown in the baseline ordinary least square regression. The regression from the table indicates a negative correlation relation between log hourly wage and being a smoker. It is presented that individuals who smoke earn 7%less than those who do not smoke (without controlling for country and year dummies) which is in line with most of the literature on smoking and wages. Moreover, controlling for both country and yearly dummies, there was a negative and significant effect of almost 5%. However, performing Skewness/Kurtosis tests for Normality shows that the residuals are not normally distributed and it will be too early to make any conclusions without further investigations. Therefore, additional health related variableswere added as shown in equation 1 and 2.

3.2. OLS verses Instrumental Variables (IV) regression.

We estimate a standard human capital earnings model made up of personal characteristics such as age, gender, marital status, highest level of general education completed; the baseline approach is very similar to the augmented human capital model of the effects of health status and health behaviour on labour market outcomes used by previous researchers. Hence baseline model can be construed as:

$$\ln hWage = \alpha + \beta X + \gamma Smoke + \varepsilon \qquad (1)$$

where $\ln hW$ is the log hourly wage, X is individual characteristics as mentioned above and Smk is a smoking behaviour dummy (1=current daily smokers and 0= non-smoker). As a matter of fact, the Ordinary Least Square regression shows a negative relationship as shown in many literatures.

However even though smoking affect wages directly through consumer's demand for the commodity by expending their income on smoking related products, health also affect individuals wage through adverse effects of smoking but this relationship is somewhat not straightforward (Grafova & Stafford, 2005). Therefore, to account for the possibility that smoking affects wages through lower productivity due to poorer health, the baseline model is augmented by a self-reported health status indicator; how many times individual has been admitted to the hospital and the weight of the individuals without clothes and shoes[1]. This additional variable is treated with the use of Instrumental Variables (IV) which is introduced in several papers. This procedure will alleviate the unmeasured, time invariant, individual-specific variables that could be correlated with both wages and smoking behaviour which lead to biased estimates[2].

In theory, the IV technique with a valid instrument solves all these problems. A valid instrument acts as a randomization device. In a randomized trial, we can think of a fair coin toss as deciding who is selected into treatment and who is not. In a valid IV analysis, the instrument assigns subjects to either treatment or no treatment using an assignment mechanism that is independent of the outcome. For example, the presence of unmeasured risk aversion in observational data represents a clear violation of random assignment. Persons in the treatment group (smokers) would likely be less risk averse than persons in the control group (non-smokers). A valid instrument must be theoretically related to the treatment variable but not, theoretically at least, be directly related to the outcome. Invalid instruments are those for which causality arrows might run from: (1) the treatment to the instrument; (2) from the outcome to the instrument; (3) from the instrument to the outcome; or (4) from other variables to the instrument, treatment, and outcome. A weak instrument is weakly statistically correlated with the treatment variable. Unfortunately, weak and invalid instruments frequently appear in the literature. For example, Berger used mother's educational attainment as an instrument for adult child's own educational attainment in assessing the correlation between education and health (Berger & Leigh, 1989). But mother's educational attainment likely influences the child's health in early years, which certainly influences the adult's health in later years. Thus, the instrument (mother's educational attainment) affects the outcome (health) for reasons other than the effects of the instrument on the treatment.

Angrist & Krueger (1991) illustrated how IV can be used to solve the omitted variable problem by using cross-sectional regression equation to measure the rate of return to schooling. By adopting the same procedure in dealing with the omitted variable problem our model can now be estimated to be as follows:

$$\ln hWage = \alpha + \beta X + \gamma Smoke + \delta Hlt + \mu \quad (2)$$

[1] It is a well-known fact that smoking affects individual's weight.

[2] Researchers in attempt to estimate effects of treatment on outcome using observational data confront many problems. First, it could be that the outcome variable results in the treatment. Without adjustment for this reverse causality, conventional methods would underestimate the effect of smoking on wages. Second, random measurement error for the treatment variable can result in an underestimate of the effect of treatment on outcome. Third, there may be some unobserved, perhaps unmeasurable, variable or set of variables that could influence or be influenced by both the treatment and the outcome. In most cases, we would want to exclude the unobserved variables (Leigh & Schembri 2004).

In the above equation $\ln hW$ is the log hourly wage, X is an individual characteristic, Smk is a current smoking behaviour dummy and Hlt is a measure of health (number times admitted at the hospital and the weight of the individuals). The estimation of this equation can be straight forward, but in principle data on Hlt are typically unavailable, and most researchers are unsure what the right controls for health would be the in any case. Without additional information, the parameter of interest, γ, is unidentified. However, the instrument denoted by Z, which is correlated with smoking but uncorrelated with wages. Thus Z, is uncorrelated with the omitted variables and the regression error term, μ. The IV methods enable us to estimate the coefficient of interest consistently and free from asymptotic bias from omitted variables, without actually having data on the omitted variables. IV solves this issues by using only part of the variability in smoking specifically, a part that is uncorrelated with the omitted variables. Angrist & Pischke (2009) also provided a very useful overview of the challenges of causal inference in econometrics. Among other things they discussed the role of control variables in the regression to reduce the omitted variable bias problem. By their assertion, we could say in our case that individuals with good health tends to have higher wages $(\delta > 0)$ and are also likely to be non-smokers. Since most literatures uses body mass index as instrument, the novelty in our research is the weight and number of times the individual has been admitted to the hospital. The results are shown in the subsequent section.

3.3. Heckman Model

Heckman models aredesigned to deal with sample selection bias, but the same approach can be used to deal with non-random assignment to treatment as well. Selection bias can be thought of as a form of omitted variable bias. This bias results from using non-randomly selected samples to estimate behavioural relationships as an ordinary specification bias that arises because of a missing data problem. In contrast to the usual analysis of "omitted variables" or specification error in econometrics, in the analysis of sample selection bias it is sometimes possible to estimate the variables which when omitted from a regression analysis give rise to the specification error. The estimated values of the omitted variables can be used as regressors so that it is possible to estimate the behavioural functions of interest by simple methods (Heckman, 1979). Following this and the possibility of encountering similar problems in our data, Heckman correction model is later followed to estimate our wage-smoking equation to deal with such error that may arise.

The basic selection equation can be taught of as follows:

$$z_i^* = w_i\gamma + u_i \tag{3}$$

$$z_i = \begin{cases} 1 \text{ if } z_i^* > 0 \\ 0 \text{ if } z_i^* \leq 0 \end{cases}$$

The outcome equation is given as:

$$y_i = \begin{cases} x_i\beta + \varepsilon_i & \text{if } z_i^* > 0 \\ - & \text{if } z_i^* \leq 0 \end{cases} \tag{4}$$

The problem that arises from the above equation is when estimating β if u_i and ε_i are correlated and a bivariate normal distribution with zero means.The results are shown in subsequent section. This model can be estimated in two forms either by two-stage procedure or by MLE procedure. However, the latter requires making a strong assumption than those required for the two-stage procedure. Hence we use the former in our estimates which only requires that u_i and ε_i are independent of the explanatory variables with zero means and that $u_i \in N(0,1)$ (Wooldridge, 2002).This procedure is the most commonly method used for estimating the Heckman model.

4. Analysis and Discussion of Results

4.1. OLS and Instrumental Variable Estimates

This section presents the results from the baseline regression corresponding to equation 1 above; the effect of smoking on wage and other variables of interest like age, gender, marital status, working hours and those working in education sector. The results are presented in table 2 below.

Table 2. *OLS estimates of wage effects on smoking (Dep. Var = log hourly wage).*

Explanatory Variables	OLS (1)	OLS (2)	OLS (Male)	OLS (Female)
Age	0.0176952 ***	0.0641885***	0.0671177***	0.0648228***
	[0.0067651]	[0.0012928]	[0.0018199]	[0.001835]
Age squared	-0.0001723 **	-0.0006666***	-0.0006749***	-0.0007067***
	[0.0000838]	[0.000016]	[0.0000221]	[0.0000231]
Marital status	0.215251 ***	0.044521***	0.0962966***	-0.0063958
	[0.024839]	[0.0047336]	[0.0071627]	[0.0063622]
Education	0.4509665 ***	0.3061466***	0.2238809***	0.352254***
	[0.0315711]	[0.0060039]	[0.0099146]	[0.0075671]
Hours worked	0.0154326 ***	-0.0080662***	-0.0094711***	-0.0077055***
	[0.0011889]	[0.0002281]	[0.0003182]	[0.0003274]
Smoke	-0.0627336 ***	-0.0457202***	-0.0822317***	-0.0104358
	[0.0231322]	[0.0044249]	[0.005856]	[0.0066211]
Gender[3]	0.0141922 ***	0.198654***		
	[0.0221151]	[0.0042171]		
Country Dummies	No	Yes	Yes	Yes
Year Dummies	No	Yes	Yes	Yes
R-square	0.0128	0.9645	0.9683	0.9618

Note: The notation ***, **, and * denotes significance at the 1 percent, 5 percent and 10 percent respectively and Standard errors are presented in parenthesis.

Since the wage of smokers cannot be justified by individual's smoking behaviour, other variables which may influence wage have been added to the previous regression to ascertain their effects on wage. The results in table 2 above shows that individuals who smoke earn 6.2% less than those who do not smoke at a highly statistically significant value without controlling for both country and year dummies. The absolute effect of smoking on wage here is augmented as compared to the previous regression. This might be due to the fact that smoking behaviours are also influenced by other variables. The other explanatory variables also depict results with expected signs. For instance, age, marital status, gender and education all have positive and statistically significant effects on wage. However, when we control for country and year dummies, smoking effects decreases to 4.6% with the same sign but with high R-square value. These results fortify the approaches and results from previous researches as in Berger & Leigh (1989); and Levine et al (1997). Breusch-Pagan and Cook-Weisberg test for heteroscedasticity reject the null hypothesis of homoscedasticity with a very low probability andtests for

[3] OLS results for Gender under Male and Female omitted because of collinearity

Normality shows that the residuals are not normally distributed which might be due to endogeneity of our regressors.

Additionally, an individual regression for male and female is presented in appendices A1 and A2 respectively. The results show that males who smoke earn less while their female counterparts earn no wage penalty. This is in line with the results from van Ours (2004) using Dutch 2001 centER data.

Though the results presented above is in line with what most researchers say about the wage and smoking relationships, our regressors could be seen as an endogenous variables having effect on other variables which affect wage. Therefore, individual's health status was introduced to augment the previous regression as shown in equation 2. The results are shown in table 3 below.

The results presented in table 3 indicate that health related variables (admitted to hospital and weight) that were used to augment the equation all have insignificant coefficients. Additionally, Breusch-Pagan and Cook-Weisberg test for heteroscedasticity reject the null hypothesis of constant variance and test for normality on the residual was also rejected. Thereby for viability we resulted to the used of Instrumental Variables as in Angrist & Pischke (2009) where they revealed omitted variable or measurement error as the main source of such a problem. However, by estimating the model with the use of instruments these problems would be alleviated. A comparative result with the previous OLS is presented in table 3 below. The full results of both first and second-stages are presented in appendix A3.

Table 3. *OLS and IV estimates of wage effects on smoking and Health related variables.*

Explanatory Variables	OLS (2)	OLS (3)	IV Coefficient
Age	0.0641885***	0.0640339***	0.0672398***
	[0.0012928]	[0.0012978]	[0.0026623]
Age squared	-0.0006666***	-0.0006651***	-0.0007048***
	[0.000016]	[0.000016]	[0.0000332]
Gender	0.198654***	0.1948229***	0.2160202***
	[0.0042171]	[0.005094]	[0.013849]
Marital status	0.044521***	0.0442603***	0.0330674***
	[0.0047336]	[0.0047409]	[0.00993387]
Education	0.3061466***	0.3062483***	0.2938225***
	[0.0060039]	[0.0060045]	[0.0111694]
Hours worked	-0.0080662***	-0.0080765***	-0.0078553***
	[0.0002281]	[0.0002283]	[0.0002822]
Smoke	-0.0457202***	-0.0455247***	-.2269029*
	[0.0044249]	[0.0044273]	[0.137422]
Weight		0.0002483	
		[0.0001863]	
Admitted to Hospital		-0.0029464	
		[0.0085953]	
Country Dummies	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes
R-square	0.9645	0.9644	0.9630
Instrumented: Smoke			
Instruments: Gender Age Agesquared Maritalstatus Education Hoursworked WeightAdmittedHospital Countrydummies Yeardummies			

Note: The notation ***, **, and * denotes significance at the 1 percent, 5 percent and 10 percent respectively and Standard errors are presented in parenthesis.

The results from instrumental variable estimate shows that smoking affects wages negatively and also with a higher magnitude of approximately 23%at 10 percent significant level. This is probably due to the fact that the 2SLS uses the full information available to compute the fitted values of the instrumented variable (smoke) while in the previous cases we used only the information given by smoke that may suffered from endogeneity problem. Moreover, the 2SLS uses a richer set of regressors to capture information about smoking and, therefore, it allows us explain and overcome better the endogeneity problem. Both Sargan and Basmann test for over-identifying restrictions provided a very high p-value (0.814) which indicatesstrong instruments. The results from first-stage regressions also revealed statistically significant instrumental variables. The joint statistical distribution for the instrumented variables also shows that they are statistically significant.

4.2. Results from Heckman Estimation

As a results of self-selection by the individuals or data units being investigated and also non-randomly selected data, the following section presents the results obtained from Heckman correction estimates.

Table 4. *Heckman Estimation results*

Explanatory Variables	IV Coefficient	Heckman Coefficient
Age	0.0672398***	0.0741429***
	[0.0026623]	[0.0063876]
Age squared	-0.0007048***	-0.0007823***
	[0.0000332]	[0.0000803]
Gender	0.2160202***	0.2124858***
	[0.013849]	[0.0323852]
Marital status	0.0330674***	-0.0108952
	[0.00993387]	[0.0221685]
Education	0.2938225***	0.23626***
	[0.0111694]	[0.0263364]
Hours worked	-0.0078553***	-0.0067629***
	[0.0002822]	[0.0006042]
Smoke	-.2269029*	omitted[4]
	[0.137422]	
Country Dummies	Yes	Yes
Year Dummies	Yes	Yes
R-square	0.9630	-

Note: *The notation ***, **, and * denotes significance at the 1 percent, 5 percent and 10 percent respectively and Standard errors are presented in parenthesis.*

The Heckman coefficient presented above represent the Heckman first-stage estimateswhich gives similar results like the IV estimates except marital status which changes sign. Following the full results from appendix A4, the second-stage shows that all the instruments used (number of times individuals are admitted to the hospital and weight) are statistically significant which cement our results from the other estimates. Though Achen (1986) warn about including too many instruments; "Experimental data derived from nonrandomized assignments, controlling for additional variables in a regression may worsen the estimate of the treatment effect, even when the additional variables improve the specification" which in our case all seem to be good instruments.

4.3. Results from Endogenous Switching and Matching Estimates

Following the results and statistical test from our IV estimates and Heckman correction, we try to adopt endogenous switching and matching estimation procedure to generate the effect of the treatment variable (smoking) to analyse the real effect on non-smokers. The results are presented below in table 5.

Table 5. *Endogenous Switching and Matching estimator results*

	Endogenous Switching estimates	Matching estimates
ATE	-0.4738188	-
	[0.054592]	
ATT	-0.4675544	-
	[0.0543333]	
ATT (Nearest Neighbour)	-	-0.0688285
		[0.0396789]
		Bias= 0.0048384
ATT (Kernel)	-	-0.0433207
		[0.0225088]
		Bias= -.000437

[4]Smoke omitted because of collinearity (see appendix A4)

Standard errors are presented in parenthesis

The results from endogenous switching estimate *a la* Heckman correction procedure, revealed an Average Treatment Effect and Average Treatment on the Treated of 47.4% and 46.8% respectively; showing that smoking characteristics have a negative and a significant relationship on individual's wage.Matching estimator seen in few literatures also provides similar negative effects on the treatment variable with very small biases. The Average Treatment on the Treated using both Nearest Neighbour and Kernel matching presented -6.9% and -4.3% respectively which also cement the negative effect results reported in most literature on smoking and wage effects.

5. Conclusion and Policy Recommendation

So far the paper has investigated the effect of smoking on wages in some selected European countries and other economic variables of interest. The econometric methods used in this work include instrumental variable technique, Heckman correction factor, Endogenous switching and matching estimator. All these methodologies concluded with a negative relation but with different magnitudewhich is in line with previous research. It was revealed that the wage gap between smokers andnon-smokers ranges between 1% and 22.7%. Both endogenous switching and matching estimator also revealed a negative average treatment effect of 47% and 4.3% to 6.9% respectively. These results were actually expected as most literature suggests similar effects. The empirical results that hours worked is negatively related to wage was unexpected. Intuitively, we assume that as individuals work more hours, they are supposed to earn more wages and thereby have a positive relationship with wages but this data suggest the opposite.

This work has been able to identify the indirect wage effects of smoking via health status. Precisely, smoking is highly predictive of low health status, and this in turn lowers wages. Thus, while smokers observed negative wage effects appear to be explained in part by individual's health status, they are also linked to a measure of unobserved preferences and smoker's behaviour.

References

Achen, C.H., (1986). *The Statistical Analysis of Quasi-Experiments*. Berkeley, CA: University of California Press

Adhikari, B., Kahende, J., Malarcher, A., Pechacek, T., & Tong V., (2008). Smoking-attributable mortality, years of potential life lost, and productivity losses - United States, 2000–2004. *Morbidity and Mortality Weekly Report*, 57(45), 1226–1228.

Anger, S., & Kvasnicka, M., (2010). Does smoking really harm your earnings so much? Biases in current estimate of the smoking wage penalty. *Applied Economic Letters;* 17(6), 561-564. doi. 10.1080/13504850802260846

Angrist, J.D. & Kruger, A.B., (1991). Does compulsory school attendance affects schooling and earnings? *Quarterly Journal of Economics*. 106(4), 979-1014. doi. 10.2307/2937954

Angrist, J.D. & Pischke, J.S., (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton, NJ.

Auld, M.C. (1998). *Wage, Alcohol Use, And Smoking: Simultaneous estimates*. Discussion Paper, No. 98/08, Department of Economics, University of Calgary.

Becker, G., (1964). Human capital: A theoretical and empirical analysis with special reference to education, *National Bureau of Economic Research, No. 80*.

Berger, M.C., & Leigh J.P. (1989). Schooling, self-selection and health. *Journal of Human Resource*, 24(3), 433–455. doi. 10.2307/145822

Bertera, R.L. (1991). The effects of behavioral risks on absenteeism and health-care costs in the workplace. *Journal of Occupational Medicine*, 33(11), 1119-1123.

Blake, G.H., Abell T.D., & Stanley W.G.,(1988). Cigarette smoking and upper respiratory infection among recruits in basic combat training. *Annals of Internal Medicine,* 109(3), 198–202.

Braakman, N., (2008). *The smoking wage penalty in the United Kingdom: Regression and matching evidence from the British Household panel survey*. University of Luneburg Working Paper Series in Economics. No 96.

Burda, M. & Wyplosz, C. (2005). *Macroeconomics: A European Text*, Oxford University Press Inc., New York.

Conway, T.L., & Cronan, T.A. (1992). Smoking, exercise, and physical fitness. *Preventive Medicine,* 21(6), 723-734. doi. 10.1016/0091-7435(92)90079-W

Frieze, I.H., Olson, J.E., & Russell, J. (1991). Attractiveness and income for men and women in management. *Journal of Applied Social Psychology*, 21(13), 1039–1057. doi. 10.1111/j.1559-1816.1991.tb00458.x

Grossman, M. (1972). The demand for health: A theoretical and empirical investigation. Occasional Papers No. 119. New York: Columbia University Press.

Grossman, M. (1972). On the concept of human capital and the demand for health, *Journal of Political Economy*, 80(2), 223-255. doi. 10.1086/259880

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica,* 47(1), 153–161. doi. 10.2307/1912352

Heijdra, B. J., & van der Ploeg, F. (2002). *Foundations of modern macroeconomics*, Oxford University Press Inc., New York.

Heineck, G. & Schwarze, J. (2003). *Substance use and earnings: the case of smokers in Germany*, Discussion paper, IZA DP No.743.

Hoad, N. A., & Clay, D. N. (1992). Smokinig impairs the response to a physical training regime: A study of officer cadets. *Journal of the Royal Army Medical Corps*, 138(3), 115- 117.

Kristein, M.M. (1983). How much can business expect to profit from smoking cessation?, *Preventive Medicine*, 12(2), 358–81. doi. 10.1016/0091-7435(83)90245-1

Lee, Y. (1999). *Wage Effects of drinking and smoking: an analysis using Australian Twins Data*. University of Western Australia Working Paper, 99-122.

Leigha, J.P., & Schembri, M. (2004). Instrumental variables technique: cigarette price provided better estimate of effects of smoking on SF-12, *Journal of Clinical Epidemiology* 57, 284–293.

Levine, P.B., Gustafson, T.A., & Velenchik, A.D. (1997). More bad news for smokers? The effect of cigarette smoking on wages, *Industrial and Labor Relations Review*, 50(3), 493–509. doi. 10.1177/001979399705000307

Martel, L. F., & Biller, H. B. (1987). *Stature and Stigma: The Biopsychosocial Development of Short Males*. Lexington, Mass.: Lexington Books.

Napier, K. (1996). *Cigarettes: What the Warning Label Doesn't Tell You. TheFirst Comprehensive Guide to the Health Consequences of Smoking.* (with editing by William M. London, Elizabeth M. Whelan, and Andrea Golaine Case).New York: American Council on Science and Health.

Stiglitz, J.E. (1986). Theories of wage rigidity. in Butkiewicz, J. L., Koford, K. J., and Miller, J. B. editors, *Keynes' Economic Legacy: Contemporary Economic Theories*, Praeger, New York.

van Ours, J.C. (2004). A pint a day raises a man's pay; but smoking blows that gain away, *Journal of Health Economics*, 23(5), 863-886. doi. 10.1016/j.jhealeco.2003.12.005

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

Appendices

Appendix A1: OLS regression for MALE

note: gender omitted because of collinearity

Source	SS	df	MS
Model	92902.7825	16	5806.42391
Residual	3038.76389	19943	.152372456
Total	95941.5464	19959	4.80693153

Number of obs	= 19960
F(16, 19943)	=38106.78
Prob > F	= 0.0000
R-squared	= 0.9683
Adj R-squared	= 0.9683
Root MSE	= .39035

lnhwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
smoke	-.0822317	.005856	-14.04	0.000	-.09371	-.0707534
age	.0671177	.0018199	36.88	0.000	.0635505	.0706849
age2	-.0006749	.0000221	-30.58	0.000	-.0007181	-.0006316
gender	0	(omitted)				
currentmarried	.0962966	.0071627	13.44	0.000	.0822572	.1103361
educ	.2238809	.0099146	22.58	0.000	.2044475	.2433142
hworked	-.0094711	.0003182	-29.77	0.000	-.0100948	-.0088475
country						
4	1.419	.0178671	79.42	0.000	1.383979	1.454021
8	-2.304927	.0144447	-159.57	0.000	-2.33324	-2.276615
9	-2.493377	.0121859	-204.61	0.000	-2.517262	-2.469492
10	3.034297	.0131555	230.65	0.000	3.008512	3.060083
11	-2.500515	.0124139	-201.43	0.000	-2.524847	-2.476182
12	2.228073	.0125608	177.38	0.000	2.203453	2.252693
13	.3300441	.0136656	24.15	0.000	.3032583	.3568299
14	-.4726288	.0155232	-30.45	0.000	-.5030556	-.4422021
year						
2000	.0416321	.0068395	6.09	0.000	.0282262	.0550381
2001	.085304	.0067731	12.59	0.000	.0720282	.0985798
_cons	3.194988	.0378295	84.46	0.000	3.120839	3.269137

Appendix A2:OLS regression for FEMALE

note: gender omitted because of collinearity

Source	SS	df	MS
Model	93289.8911	16	5830.61819
Residual	3706.95775	21919	.169120751
Total	96996.8488	21935	4.42201271

Number of obs	= 21936
F(16, 21919)	=34476.07
Prob > F	= 0.0000
R-squared	= 0.9618
Adj R-squared	= 0.9618
Root MSE	= .41124

lnhwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
smoke	-.0104358	.0066211	-1.58	0.115	-.0234137	.0025422
age	.0648228	.001835	35.33	0.000	.0612261	.0684196
age2	-.0007067	.0000231	-30.58	0.000	-.000752	-.0006614
gender	0	(omitted)				
currentmarried	-.0063958	.0063622	-1.01	0.315	-.0188661	.0060746
educ	.352254	.0075671	46.55	0.000	.337422	.367086
hworked	-.0077055	.0003274	-23.53	0.000	-.0083473	-.0070637
country						
4	1.363977	.0164408	82.96	0.000	1.331752	1.396202
8	-2.438368	.0135981	-179.32	0.000	-2.465022	-2.411715
9	-2.535072	.0118343	-214.21	0.000	-2.558268	-2.511876
10	2.910236	.0133267	218.38	0.000	2.884114	2.936357
11	-2.618926	.0119789	-218.63	0.000	-2.642405	-2.595446
12	2.132036	.0116983	182.25	0.000	2.109107	2.154966
13	.1665448	.0128712	12.94	0.000	.1413164	.1917733
14	-.565622	.0136036	-41.58	0.000	-.592286	-.538958
year						
2000	.0438518	.0069234	6.33	0.000	.0302815	.0574222
2001	.0889932	.0068401	13.01	0.000	.0755861	.1024002
_cons	3.179632	.0369519	86.05	0.000	3.107204	3.252061

Appendix A3: Instrumental Variable Results (FULL).

First-stage regressions

Number of obs = 41896
F(18, 41877) = 102.58
Prob > F = 0.0000
R-squared = 0.0422
Adj R-squared = 0.0418
Root MSE = 0.4467

smoke	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gender	.1155002	.0055941	20.65	0.000	.1045356	.1264648
age	.0176494	.0014298	12.34	0.000	.014847	.0204519
age2	-.0002184	.0000177	-12.36	0.000	-.0002531	-.0001838
currentmarried	-.0619898	.005224	-11.87	0.000	-.0722289	-.0517506
educ	-.068428	.0066191	-10.34	0.000	-.0814015	-.0554544
hworked	.0012136	.0002519	4.82	0.000	.0007199	.0017073
country						
2	0	(empty)				
4	-.035146	.0134406	-2.61	0.009	-.0614899	-.0088022
8	-.039206	.0109958	-3.57	0.000	-.0607581	-.017654
9	-.0373828	.0094796	-3.94	0.000	-.0559631	-.0188025
10	.1030085	.0103616	9.94	0.000	.0826995	.1233174
11	.0554152	.0095986	5.77	0.000	.0366018	.0742286
12	-.0811801	.0095442	-8.51	0.000	-.0998869	-.0624733
13	-.0242244	.0104084	-2.33	0.020	-.044625	-.0038238
14	-.0568782	.0113453	-5.01	0.000	-.0791151	-.0346413
year						
1999	0	(empty)				
2000	-.0019401	.005422	-0.36	0.720	-.0125673	.0086871
2001	-.006131	.0053633	-1.14	0.253	-.0166431	.0043811
admitHosp	.0265802	.0094862	2.80	0.005	.007987	.0451734
weight	-.0012735	.0002056	-6.19	0.000	-.0016764	-.0008706
_cons	.0161241	.0308515	0.52	0.601	-.0443455	.0765937

Instrumental variables (2SLS) regression

Number of obs = 41896
Wald chi2(17) = 1.1e+06
Prob > chi2 = 0.0000
R-squared = 0.9630
Root MSE = .41261

lnhwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
smoke	-.2269029	.137422	-1.65	0.099	-.496245	.0424393
gender	.2160202	.013849	15.60	0.000	.1888766	.2431637
age	.0672398	.0026623	25.26	0.000	.0620218	.0724579
age2	-.0007048	.0000332	-21.23	0.000	-.0007698	-.0006397
currentmarried	.0330674	.0099338	3.33	0.001	.0135976	.0525373
educ	.2938225	.0111694	26.31	0.000	.2719309	.315714
hworked	-.0078553	.0002822	-27.83	0.000	-.0084085	-.0073021
country						
4	1.389247	.0131724	105.47	0.000	1.363429	1.415064
8	-2.377936	.0113662	-209.21	0.000	-2.400213	-2.355659
9	-2.516918	.0096198	-261.64	0.000	-2.535772	-2.498063
10	2.990958	.017215	173.74	0.000	2.957217	3.024699
11	-2.54737	.0121141	-210.28	0.000	-2.571113	-2.523627
12	2.163068	.0135707	159.39	0.000	2.13647	2.189666
13	.2408325	.0100728	23.91	0.000	.22109	.2605749
14	-.5365968	.0130507	-41.12	0.000	-.5621757	-.5110178
year						
2000	.042928	.0050186	8.55	0.000	.0330917	.0527644
2001	.0857389	.0050391	17.01	0.000	.0758625	.0956154
_cons	3.089287	.0275513	112.13	0.000	3.035288	3.143287

Instrumented: smoke
Instruments: gender age age2 currentmarried educ hworked 4.country 8.country 9.country 10.country 11.country 12.country 13.country 14.country 2000.year 2001.year admitHosp weight

Appendix A4: Heckman two-step Estimates.

Heckman selection model -- two-step estimates
(regression model with sample selection)

Number of obs = 41896
Censored obs = 29512
Uncensored obs = 12384

Wald chi2(16) = 408849.00
Prob > chi2 = 0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnhwage						
gender	.2124858	.0323852	6.56	0.000	.149012	.2759597
smoke	0	(omitted)				
age	.0741429	.0063876	11.61	0.000	.0616234	.0866624
age2	-.0007823	.0000803	-9.74	0.000	-.0009396	-.0006249
currentmarried	-.0108952	.0221685	-0.49	0.623	-.0543447	.0325543
educ	.23626	.0263364	8.97	0.000	.1846416	.2878784
hworked	-.0067629	.0006042	-11.19	0.000	-.0079471	-.0055787
country						
4	1.384127	.0215999	64.08	0.000	1.341792	1.426462
8	-2.416663	.0177478	-136.17	0.000	-2.451448	-2.381878
9	-2.46913	.0149646	-165.00	0.000	-2.49846	-2.4398
10	3.019026	.0152525	197.94	0.000	2.989132	3.048921
11	-2.551709	.0145395	-175.50	0.000	-2.580206	-2.523212
12	2.242474	.0157929	141.99	0.000	2.211521	2.273428
13	.2830633	.0164762	17.18	0.000	.2507706	.3153561
14	-.5198861	.0187673	-27.70	0.000	-.5566693	-.4831029
year						
2000	.0423678	.0084925	4.99	0.000	.0257227	.0590129
2001	.0857759	.0084128	10.20	0.000	.0692871	.1022647
_cons	2.395222	.2967709	8.07	0.000	1.813561	2.976882
smoke						
gender	.3457359	.0164393	21.03	0.000	.3135154	.3779563
age	.0577609	.0043012	13.43	0.000	.0493307	.066191
age2	-.0007231	.0000535	-13.51	0.000	-.000828	-.0006183
currentmarried	-.1908147	.0154125	-12.38	0.000	-.2210226	-.1606068
educ	-.2123862	.0207191	-10.25	0.000	-.2529948	-.1717776
hworked	.0042083	.0007407	5.68	0.000	.0027566	.00566
admitHosp	.0633638	.0280638	2.26	0.024	.0083598	.1183679
weight	-.0029895	.0005961	-5.01	0.000	-.0041578	-.0018211
_cons	-1.584839	.0876152	-18.09	0.000	-1.756561	-1.413116
mills						
lambda	.3633886	.143182	2.54	0.011	.082757	.6440201
rho	0.73899					
sigma	.49173903					