EC303 Applied Project
University of Reading
Department of Economics
25022053

The Effects of Smoking on Wages – A UK Panel Data Study

~ Do Smokers Face a Wage Punishment? ~

Abstract

This paper investigates the effect of smoking on wages, using panel data from the UKHLS covering periods between 2013-2019. Literature suggests a negative link between smoking and wages, for numerous reasons; Reductions in the MPL, poor health, workplace discrimination and negative personality traits. However, when moving from cross-sections to panel datasets, the negative effect dramatically falls, or in some cases becomes statistically insignificant. Further, this paper will assess the assumption that smokers might face non-favourable workplace treatment by assessing their likelihood of becoming employed. The results are unexpected using a Fixed Effects Model, concluding that using this Dataset, smokers face a wage "prize" of 1.8% more, statistically significant at the 10% level. However, Probit and Logit models find that smokers are less likely to be employed, which is consistent with existing literature.

Keywords: *UKHLS, Panel Data, Smoking, Wage Differentials, POLS, FE, Probit/Logit Regression*
Word Count. c.3070

1. Introduction & Literature Review:

Section 1.1: Introduction

The adverse effects of smoking are both very well-known and very well established, especially in our age and time. Since the first reporting of the causal relationship between smoking and heart disease at Mayo Clinic in 1940 and the 1964 General Surgeon Report, these effects have been ascertained, extensively studied and publicized by both the Public and the Private Sector for several anti-smoking campaigns. An example is the introduction of graphic anti-smoking ads on packs of cigarettes or the prohibition of indoor public smoking. According to various studies, smoking has been shown to lead to lung cancer, bronchitis and heart disease, as well as, is associated with a magnitude of conditions such as slow healing and increased susceptibility to infections and other illnesses. Therefore, smokers other than the direct private cost of buying cigarettes face numerous indirect costs from increased ill-health possibilities. In addition, they impose external costs on the society. As it was previously mentioned these have been established and are outside the scope of this paper. The purpose of this paper is to assess whether smokers face lower wages as opposed to non-smokers. Using panel data, this paper aims to understand this relationship, if it exists and to evaluate explanations based on existing literature.

The rest of this paper is concerned with the following; a literature review regarding the effects of smoking on wages, both theoretical and empirical; Discussion on the data used (benefits/limitations) and any required variable specifications; Discussion on the Econometric approach used and presentation/discussion of the results from the estimations. The software used for the estimation will be Stata/SE 16.0 and the code will be provided as do.file in the Appendix.

Section 1.2: Theoretical Work

Much literature exists on the topic of wage differentials regarding smoking and its effect on wages, theoretically. Explanations have been separated into three different directions.

The first is concerned with the effects of smoking on productivity - more specifically, the reduced productivity of smokers. This reduced productivity is a result of breaks during work and due to absence from work resulting from ill-health. These can be easily quantified and as a result the reduced productivity due to smoking can take observable forms. The same, suggest that there are forms more difficult to quantify such as physical and mental endurance (Bertera 1991; Kristein 1983). The productivity costs of smoking have been estimated to be $80 - $160 per annum, per smoker, using evidence from numerous studies and smokers on average miss at least an additional day per work year (Bertera 1991; Kristein 1983). As a result, this lower productivity takes the form of lower wages. In addition, Brune argues that "employers anticipate future lower productivity of smokers from accumulated adverse health effects ", hence may be less willing to employ smokers, ceteris paribus (Brune, 2007). Literature explains this relationship by explaining the effects of smoking in the workplace, through higher costs for the employer being reflected by lower wages, however, given no smoking policies, these do not apply in this study. However, such policies may lead to more frequent breaks and thus further lower productivity.

The second is concerned with discrimination which could explain a wage differential among smokers and non-smokers. In other words, non-smokers may receive favorable treatment as opposed to smokers in the workplace and this could be reflected in the form of lower wages. According to Amoo (2016), smoking could lead to skin damage or more generally affect visual attractiveness. Further, non-visual attractiveness could be affected due to tobacco smoke smelling residue (Amoo, 2016). Both co-workers and clients could refuse/object to working with smokers both due to annoyance leading to discrimination and due to objections as a result of health concerns due to passive smoking effects.

The third type of explanation is concerned with arguments regarding unobserved heterogeneity. Smoking can indeed be characterized with giving up long-term health for short-term (or instant) gratification. Hence, smokers relative to non-smokers, may differ in unobservable characteristics and this could be also reflected in the workplace. More specifically, in terms of the giving up of long-term benefits for instant satisfaction, according to Amoo, smokers may have higher rates of time preference (Amoo, 2016). This could lead to the fact that individuals who have lower discount rates, prefer jobs with steeper wages (Munasinghe and Sicherman, 2000). More generally, unobservable characteristics of smokers, which are reflected in the marketplace and lead to lower wages, could include; Higher discounting rate for the future, less self-control/discipline, poor judgement/less intelligence and so on.

Section 1.3: Empirical Work

On the empirical side, there exists much literature investigating the effect of smoking on wages. Consistent evidence of a negative relationship has been found by Levine at al. (1997), Auld (2005), Lee (1999), Grafova and Stafford (2005), and more recently, Anger and Kvasnika (2006). More specifically, a wage differential against smokers, in the range of 2% - 10% has been found. Levin et al. (1997) have found that smokers face reduced wages by 4.2% to 6.9% and they suggested that this is a result of both employer discrimination and lower productivity (Amoo, 2016). Levine et al. (1997), using the national Longitudinal Survey of Youth, have found that smokers earn 11% to 17% less as opposed to non-smokers in years 1984 and 1991, respectively – And after having controlled for individual/family characteristics, the wage gap falls to 4.2% and 6.9%, again for the same years (Munasinghe and Sicherman, 2000). However, a joint study on the effects of both alcohol and tobacco consumption on wages in Sweden, has found that there is no significant relationship between smoking and wages in Sweden (Grek, 2007). In terms of the time preference argument mentioned above, it has been found that "smokers have substantially flatter wage profiles" (Munasinghe and Sicherman, 2000). Using the BHPS, Brune (2007), finds "That all else equal, smokers have a 12.7% lower wage per hour than non-smokers", which is statistically significant, when using Pooled OLS. However, this result drops to 2% lower wages per hour for smokers, statistically significant, at the 10% level.

2. Data

Section 2.1 Data

The paper uses data from the UK Household Longitudinal Survey (UKHLS), drawing on 5 available waves covering the years 2013-2019. The UKHLS is an indefinite life panel survey, where sample members are interviewed every year. The survey started in 2009 (Wave 1) and approximately 40,000 households in the United Kingdom were randomly selected to participate in the study. Each year, households selected at the first stage are followed and are re-interviewed in order to collect information on household/individual circumstances changes. Respondents 16 or over complete the Adult Questionnaire and respondents under 16 complete the Youth Questionnaire. Its purpose is to provide longitudinal data on subjects such as health, work, education, income, family etc.to better understand socioeconomic changes in the UK.

Given the use of Panel Data, it is important to provide a brief discussion regarding the advantages and possible difficulties that may arise in this study. Panel/Longitudinal data refers to data that contains both a cross-sectional and time-series aspect. In other words, data that contains time-series observations for a number of observations (Hsiao, 2007, p.). Thus, as opposed to cross-sections which give us a snapshot of the population at a specific point in time, panel data allows for numerous benefits. First, it allows for controlling unobserved and time invariant factors, "often referred to as unobserved heterogeneity" (Longi and Nadi, 2015, p.3), by employing Fixed and Random Effect Methods. According to Roos et al., when fixed effect models are used for statistical analysis, variation is within units and minimize unobserved heterogeneity and omitted variable bias (Roos et al, 2019, p. 3). The same authors in analysing the limitations of fixed effect models, state that restricted time periods leads to less reliable coefficients (Roos at al., 2019, p.10). This is an important limitation to note, as this paper uses a panel with a large N but a relatively low T, specifically 5 waves of data from the UKHLS. Thus, if the responses of the cross-sectional units do not vary much, then statistical power is reduced.

This study's population of interest will be as follows; Waves with relevant variables, i.e. Wave 5 to Wave 9 will be combined into Long Format, renamed accordingly in order to form the longitudinal panel. Wave 5 does not include a "smoker" variable as waves 6 through 9 do. However, it contains an "*Are you currently smoking?"* and a "*Have you ever smoked*" variable. Thus, I make the assumption that if a cross-sectional unit has answered yes to the first question, then is a smoker. Similarly, if the answer is No, I recode that individual as a current non-smoker. This is to generate a smoker dummy taking the value of 1 if an individual is a smoker, 0 otherwise. The dependent variable is the log of monthly wage and the other controls include Marriage status, age, education and region of residence, which is line with existing literature. A table of the variables is provided in the Appendix.

3. Econometric Specification

This chapter will explain the econometric methodology for estimating the relationship between smoking and wages. A standard wage equation is given by;

$$(1) \qquad \ln(Wage) = \alpha + \beta_i X + \gamma_i Smoker$$

Where $\ln(Wage)$ denotes the wage in its natural logarithm form, X is a vector the contains the control variables and γ is a dummy variable taking the value of 1 if an individual i is currently smoking at least one cigarette per day and the value of 0 otherwise.

Pooled OLS

The first step is to estimate a pooled Ordinary Least Squares Model, in which cross sectional units are pooled across the waves; this is given by equation (2).

$$(2) \qquad \ln(Wage)_{i,t} = \alpha_{i,t} + \beta_{i,t}\boldsymbol{X_{i,t}} + \gamma_{i,t}\boldsymbol{Smoker}_{\mathrm{i,t}} + \varepsilon_i \quad t = 5, \ldots, 9$$

Where X denotes the explanatory variables including; Age, Age squared, education, marriage and region of residence. Smoker is the dummy variable taking the values 1 for smokers and 0 for non-smokers. A pooled OLS has numerous problems. First, it may be the case that Endogeneity may exist, i.e. Smoking may be correlated with factors absorbed by the error term. In this instance biased estimates would be produced. A potential solution would be the Instrumental Variable and 2SLS approach. Brune (2007) corrects for Endogeneity using Parental smoking behaviour as the Instrument, however such information was not available in the UKHLS, hence his method cannot be replicated in this paper. Another problem with Pooled OLS is that it ignores the panel nature of the data and thus does not take into account unobserved heterogeneity. Therefore, a fixed-effects model, capturing this, would be more appropriate in estimating the relationship.

Fixed Effects Model

The better specification option is to estimate a fixed-effects model. The model is presented in equation (3) and is virtually unchanged from model (2), except with the "split" of the error term in two components, where time invariant characteristics are added;

$$(3) \qquad \ln(Wage)_{i,t} = \alpha_{i,t} + \beta_{i,t}\boldsymbol{X_{i,t}} + \gamma_{i,t}\boldsymbol{Smoker}_{\mathrm{i,t}} + \varepsilon_{i,t} + a_i$$

Then, by computing deviations from the mean, unobserved heterogeneity is eliminated.

Binary Outcome Models

An important consideration that needs to be made regarding smokers and non-smokers, according to relevant literature is whether employer discrimination to smokers is an issue. As it was mentioned in the theoretical background, wage differentials between two groups can be explained by more favourable treatment (in the form of wages), towards non-smokers. Taking this one step further it is interesting to assess whether non-smokers are more likely to be employed than smokers using a Binary Outcome model. This analysis will take place as follows; using the dataset a binary outcome independent variable of the form;

$$(4) \qquad Employment = Y = \begin{cases} 0, & if\ Unemployed \\ 1, & if\ Employed\ (FT, PT) \end{cases}$$

So as to estimate the probability that $Y = 1$, as a function of the independent variables. Mathematically;

$$(4') \qquad p = p(Y = 1|\boldsymbol{X})$$

Where $\boldsymbol{X}$, is the vector of controls as in equation 1. In interpreting the results of these models, magnitude of the coefficients do not play a role. Rather, of interest are the sings of the coefficients which show the likelihood of $Y = 1$. Further, marginal effects will be reported in order to assess the magnitude of likelihood that $Y = 1$.

4. Empirical Analysis and Discussion of Results.

Section 3.1 Results from Pooled OLS Model

Results from the Pooled OLS model are presented in Table 2.1.

Linear regression

lnW	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
age_dv	0.096	0.001	95.31	0.000	0.094	0.098	***
age2	-0.001	0.000	-87.25	0.000	-0.001	-0.001	***
1b.sex	0.000	.	.	.	.	.	
2.sex	-0.179	0.004	-41.73	0.000	-0.187	-0.170	***
Current_Smoker	-0.153	0.005	-28.30	0.000	-0.163	-0.142	***
Educ_Dummy	0.367	0.012	31.44	0.000	0.344	0.390	***
MarriedDummy	0.112	0.005	23.19	0.000	0.103	0.122	***
1b.jbft_dv	0.000	.	.	.	.	.	
2.jbft_dv	-1.020	0.005	-205.04	0.000	-1.030	-1.011	***
Constant	5.260	0.022	238.49	0.000	5.217	5.303	***

Mean dependent var	7.333	SD dependent var	0.836
R-squared	0.515	Number of obs	82572.000
F-test	12511.719	Prob > F	0.000
Akaike crit. (AIC)	145097.596	Bayesian crit. (BIC)	145172.167

*** p<0.01, ** p<0.05, * p<0.1

Table 2.1 POLS Regression

As it can be seen from the table, all of the variables are highly significant. As expected both Marriage and Education, positively influence wages. Under the POLS model, smokers do face a lower wage; Ceteris paribus, smokers face 15.3% lower wage as opposed to non-smokers. However, as it was previously mentioned this method ignores heterogeneity and could also suffer from Endogeneity.

Section 3.2 Results from Fixed Effect Model

Regression results

lnW	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
age_dv	0.172	0.004	47.35	0.000	0.165	0.179	***
age2	-0.002	0.000	-37.13	0.000	-0.002	-0.001	***
1b.sex	0.000	.	.	.	.	.	
2.sex	-0.048	0.099	-0.48	0.629	-0.241	0.146	
Current_Smoker	0.018	0.007	2.45	0.014	0.004	0.033	**
Educ_Dummy	0.142	0.038	3.78	0.000	0.068	0.215	***
MarriedDummy	0.028	0.009	3.16	0.002	0.010	0.045	***
1b.jbft_dv	0.000	.	.	.	.	.	
2.jbft_dv	-0.497	0.006	-85.40	0.000	-0.509	-0.486	***
Constant	3.041	0.102	29.82	0.000	2.841	3.241	***

Mean dependent var	7.333	SD dependent var	0.836
R-squared	0.179	Number of obs	82572.000
F-test	1677.181	Prob > F	0.000
Akaike crit. (AIC)	10173.598	Bayesian crit. (BIC)	10248.169

*** p<0.01, ** p<0.05, * p<0.1

Table 2.2; FE Model

The FE model produces unexpected results, inconsistent with previous literature and the baseline POLS model. Based, on the dataset it is concluded that smokers will face higher wages than non-smokers, by 1.8%, a very small gap. However, given the inconsistency and also the fact that descriptive statistics show a higher mean wage for non-smokers, doubts and concerns regarding the model are raised and various postestimation checks

were conducted. Specifically, the Pasaran CD Test for serial correlation and the Wald Test for Heteroskedasticity, which both exist in this model.

Section 3.3 Results from Binary Outcome Models

The results of the probit and logit models are given in Tables 2.3 and 2.4.

Probit regression

Employed	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
age	-0.031	0.001	-57.69	0.000	-0.032	-0.030	***
1b.sex	0.000	.	.	.	.	.	
2.sex	-0.200	0.018	-11.02	0.000	-0.236	-0.165	***
0b.Current_Smoker	0.000	.	.	.	.	.	
1.Current_Smoker	-0.246	0.020	-12.61	0.000	-0.285	-0.208	***
mastat_dv	0.029	0.003	9.38	0.000	0.023	0.035	***
Constant	1.713	0.033	52.09	0.000	1.649	1.777	***

Mean dependent var	0.537	SD dependent var	0.499
Pseudo r-squared	0.124	Number of obs	21490.000
Chi-square	3692.921	Prob > chi2	0.000
Akaike crit. (AIC)	25992.569	Bayesian crit. (BIC)	26032.446

*** p<0.01, ** p<0.05, * p<0.1

Table 2.3; Probit Model

Logistic regression

Employed	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
age	-0.052	0.001	-55.07	0.000	-0.054	-0.050	***
1b.sex	0.000	.	.	.	.	.	
2.sex	-0.335	0.030	-11.11	0.000	-0.395	-0.276	***
0b.Current_Smoker	0.000	.	.	.	.	.	
1.Current_Smoker	-0.426	0.033	-12.97	0.000	-0.490	-0.361	***
mastat_dv	0.045	0.005	8.59	0.000	0.035	0.056	***
Constant	2.945	0.059	49.55	0.000	2.829	3.062	***

Mean dependent var	0.537	SD dependent var	0.499
Pseudo r-squared	0.127	Number of obs	21490.000
Chi-square	3773.596	Prob > chi2	0.000
Akaike crit. (AIC)	25911.893	Bayesian crit. (BIC)	25951.770

*** p<0.01, ** p<0.05, * p<0.1

Table 2.4; Logit Model

Both probit and logit models produce statistically significant results that smokers are less likely to be employed. This is consistent with several hypotheses from existing literature regarding the personality traits of smokers,

such as discounting for the future, giving up long-term benefits for short term gratification and so on, as mentioned in section 2.

4. Conclusion

Some of these results were unexpected; specifically, both based on relevant literature and on intuition, smoking is assumed to have a negative effect on earnings. Literature wise, even though a number of researchers do not find statistically significant results, they do observe a negative relationship between smoking and earnings. Intuitively, smoking was assumed to (a) influence health, thereby reducing the Marginal Product of Labour and by extension earning capacities and (b) have a negative impact on wages as smokers are assumed to sacrifice long-term benefits for short-term gratification. Further, it is worth noting that while the UKHLS has been running for 9 waves, variables on smoking behaviour were available for waves 5-9 (based on the recoding assumption for wave 5). A more appropriate dataset therefore could be the BHPS which contains extensive information regarding smoking behaviour to treat for the numerous problems explained above.

5. Appendix

Section 5.1 Stata Code do.file

```
**EC303 Applied Econometrics - Do Smokers Face a Lower Wage than Non-Smokers?
**Stata Code
//general setup - combine into long format
use pidp e_sex e_age_dv e_smever e_smnow e_ncigs e_hiqual_dv e_mastat_dv e_jbft_dv
e_paygu_dv e_employ e_jbsemp e_jbstat e_jbhrs e_gor_dv e_racel_dv using
"\\dfs.rdg.ac.uk\homes\cf022053\My Documents\Applied
Econometrics\Data\w_indresp\e_indresp.dta"
gen wave = 5
rename e_* *
save appendedfile1_ec303
use pidp f_sex f_age_dv f_smoker f_ncigs f_hiqual_dv f_mastat_dv f_jbft_dv f_paygu_dv
f_employ f_jbsemp f_gor_dv f_racel_dv f_jbhrs using "\\dfs.rdg.ac.uk\homes\cf022053\My
Documents\Applied Econometrics\Data\w_indresp\f_indresp.dta"
gen wave = 6
rename f_* *
append using appendedfile1_ec303
save appendedfile2_ec303
use pidp g_sex g_age_dv g_smoker g_ncigs g_hiqual_dv g_mastat_dv g_jbft_dv g_paygu_dv
g_employ g_jbsemp g_gor_dv g_racel_dv g_jbhrs using "\\dfs.rdg.ac.uk\homes\cf022053\My
Documents\Applied Econometrics\Data\w_indresp\g_indresp.dta"
gen wave = 7
rename g_* *
append using appendedfile2_ec303
save appendedfile3_ec303
use pidp h_sex h_age_dv h_smoker h_ncigs h_hiqual_dv h_mastat_dv h_jbft_dv h_paygu_dv
h_employ h_jbsemp h_racel_dv h_gor_dv h_jbhrs using "\\dfs.rdg.ac.uk\homes\cf022053\My
Documents\Applied Econometrics\Data\w_indresp\h_indresp.dta"
gen wave = 8
rename h_* *
append using appendedfile3_ec303
save appendedfile4_ec303
use pidp i_sex i_age_dv i_smoker i_ncigs i_hiqual_dv i_mastat_dv i_jbft_dv i_paygu_dv
i_employ i_jbsemp i_racel_dv i_gor_dv i_jbhrs using "\\dfs.rdg.ac.uk\homes\cf022053\My
Documents\Applied Econometrics\Data\w_indresp\i_indresp.dta"
gen wave = 9
rename i_* *
append using appendedfile4_ec303
save appendedfile5_ec303
save LongPanel_UKHLS_EC303
use "\\dfs.rdg.ac.uk\homes\cf022053\My Documents\Applied
Econometrics\LongPanel_UKHLS_EC303.dta"
sort pidp wave
//missing values
mvdecode _all, mv(-9/-1)
//declare panel dataset
xtset pidp wave
//variable modification and generation
replace smoker = 1 if smnow == 1 & wave == 5
```

```
replace smoker = 2 if smnow == 2 & wave == 5
gen age2 = age_dv*age_dv
gen lnW = ln(paygu_dv)
gen Current_Smoker = 0 if smoker < .
replace Current_Smoker = 1 if smoker == 1
generate MarriedDummy = 0 if mastat_dv < .
replace MarriedDummy = 1 if (mastat_dv >= 2 & mastat_dv <= 3) | mastat_dv == 10
generate Educ_Dummy = 0 if hiqual_dv < .
replace Educ_Dummy = 1 if hiqual_dv>=1 & hiqual_dv<=5
generate London = 0 if gor_dv < .
replace London = 1 if gor_dv == 7
generate Full_Time = 0 if jbft_dv == 2
replace Full_Time = 1 if jbft_dv == 1
gen Employed = 0 if jbstat < .
replace Employed = 1 if jbstat == 1 | jbstat == 2
ssc install asdoc
//analysis
reg lnW age_dv age2 ib1.sex Current_Smoker Educ_Dummy MarriedDummy
ib1.jbft_dv,vce(cluster pidp)
asdoc reg lnW age_dv age2 ib1.sex Current_Smoker Educ_Dummy MarriedDummy
ib1.jbft_dv
xtreg lnW age_dv age2 ib1.sex Current_Smoker Educ_Dummy MarriedDummy ib1.jbft_dv, fe
asdoc xtreg lnW age_dv age2 ib1.sex Current_Smoker Educ_Dummy MarriedDummy
ib1.jbft_dv, fe
estimates store fe
xtreg lnW age_dv age2 ib1.sex Current_Smoker Educ_Dummy MarriedDummy ib1.jbft_dv,
re
estimates store re
xtreg lnW age_dv age2 ib1.sex Current_Smoker Educ_Dummy MarriedDummy ib1.jbft_dv, fe
estimates store fe
hausman fe re
```

Section 5.2 Variables

- **pidp**; *cross-wave identifier*
- **sex**; sex of respondent: https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/sex
- **age_dv**; age of respondent: https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/age_dv
- **ncigs**; usual no. of cigarettes smoker per day: waves 5 – 9 https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/ncigs
- **smoker**: waves 6-9 https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/smoker
- **smnow;** Smoke nowdays: waves 2,5 https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/smnow

- **smever;** ever smoked cigarettes; waves 2,5 https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/smever
- **marstat;** legal marital status: wave 1-9 https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/marstat
- **hiqual_dv:** Highest Qualification: waves 1 – 9 https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/hiqual_dv
- **jbft_dv:** Full or Part Time Employee: waves 1 – 9 https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/jbft_dv
- **paygu_dv**; usual gross pay per month: current job; waves 1-9 https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/paygu_dv

6. Reference List & Bibliography

A Eric (2016). Effects of smoking and other economic variables on wages in Sweden. Available at: https://mpra.ub.uni-muenchen.de/69230/1/MPRA_paper_69230.pdf

A M. Christopher (2005). "Smoking, Drinking, and Income". Journal of Human Resources, 40(5):505–518.

B. L. Robert (1991). The Effects of Behavioural Risks on Absenteeism and Health-Care Costs in the Workplace. *Journal of Occupational Medicine.* 33(11) pp.1119-1124

B. Lasse (2007). The Smoker's Wage Penalty: Evidence from Britain. Available at: https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2007-31.pdf

G. Jenny. (2007). The effect of smoking and drinking on wages in Sweden. Available at: http://www.diva-portal.org/smash/get/diva2:3308/FULLTEXT01.pdf

G.B. Irina and P. S. Frank (2013). The Wage Effects of Personal Smoking History. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3717362/pdf/nihms472250.pdf

H. Cheng (2005). Why Panel Data? IEPR Working Paper No. 05.33. Available at SSRN: https://ssrn.com/abstract=820204 or http://dx.doi.org/10.2139/ssrn.820204

H. Cheng (2007). Panel data analysis—advantages and challenges. *TEST* 16, pp.1–22 https://doi.org/10.1007/s11749-007-0046-

H. D. Terence et al. (2019). Limitations of Fixed-Effect Models for Panel Data. https://doi.org/10.1177/0731121419863785

Hout, M. (2016). Models for Three-Wave Panel Data. *Sociological Methods & Research*, 46(1), pp.41-43. doi: https://doi.org/10.1177/0759106317710866

L. J. Scott (2009). The Workflow of Data Analysis Using Data. Texas: Stata Press

L. Simonetta and N. Alita (2015). A Practical Guide to Using Panel Data. London: Sage Publications Ltd

Lee, Yew Liang (1999). "Wage Effects of Drinking and Smoking: An Analysis Using Australian Twins Data". Discussion Paper 22, University of Western Australia.

M. Kristein (1983). How Much Can Business Expect to Profit from Smoking Cessation. Preventive medicine. 12. 358-81. 10.1016/0091-7435(83)90245-1. Available at: https://www.researchgate.net/publication/16786671_How_Much_Can_Business_Expect_to_Profit_from_Smoking_Cessation

M. Lalith and S. Nachum (2000). Why Do Dancers Smoke? Time Preference, Occupational Choice and Wage Growth. Available at: https://core.ac.uk/download/pdf/6586741.pdf

R. T. O (2007). Panel Data Analysis Fixed and Random Effects using Stata. Available at: https://www.princeton.edu/~otorres/Panel101.pdf

University of Essex, Institute for Social and Economic Research, NatCen Social Research, Kantar Public. (2019). Understanding Society: Waves 1-9, 2009-2018 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 12th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-13.

Y. Ayda and P. M. Anne (2011) World Health Organization Economics of Tobacco Toolkit: Assessment of the Economic Costs of Smoking. Available at: https://apps.who.int/iris/bitstream/handle/10665/44596/9789241501576_eng.pdf;jsessionid=DC464AE0A781ED45FF3054BB0A7E9439?sequence=1

Anger, Silke and Michael Kvasnicka (2006). "Biases in Estimates of the Smoking Wage Penalty". Discussion Papers of DIW Berlin 654, DIW Berlin, German Institute for Economic Research.

Levine, Phillip B., Tara A. Gustafson, and Ann D. Velenchik (1997). "More Bad News for Smokers? The Effects of Cigarette Smoking on Wages". Industrial and Labor Relations Review, 50(3):493–509.