The Effect of Discounting on Smokers' Labour Outcomes

Dan Hetherington (MSc stream)

27th January 2006

1 Abstract

Separate studies have found both that smokers exhibit higher discounting rates than non-smokers and that smokers tend to suffer a negative wage differential compared to non-smokers. However, to the author's knowledge, none has attempted to assess the impact of smokers' higher discounting on their labour outcomes. Firstly, this study aims to assess whether smoking is associated with elevated discounting of future earnings by examining the relationship between smoking and wage progression; secondly, by focusing on individuals starting and stopping smoking, some indication as to the causal relationship between smoking and discount rates is sought.

An association between smoking and a high discount rate is found, although this result is vulnerable to plausible changes in specification. There is insufficient evidence to draw any conclusions about the direction of causal relations.

Contents

2 Conceptual Framework

Several studies have shown that smokers suffer a wage penalty compared to non-smokers (see Levine (1995), Bertera (1991), Auld (1998)). Many explanations for this differential have been suggested, including health consequences for manual work, absenteeism, higher health and fire insurance premia, maintenance costs, negative effects on morale, direct productivity losses associated with the time spent smoking, and discrimination.

However, Bickel (1999) has shown that, in common with users of other addictive substances, smokers exhibit a higher discount rate than non-smokers: they have a higher preference for present consumption relative to future consumption. This finding suggests another explanation for the wage differential that has not been considered: the possibility that people with a higher discount rate self-select both as smokers, and as employees with a low rate of wage progression over their lifetimes, resulting in lower average wages. Smokers could be expected either to take a job that pays well but whose wages rise slowly in preference to one which pays badly but allows greater wage progression; or else, they might select a job which allows them to enjoy free-time and a stress-free lifestyle in preference to one which pays the same at present, but offers future payoffs in the form of wage progression.

Causality in this question is difficult to determine. Another finding of Bickel (1999) is that ex-smokers do *not* have a different discount rate to never-smokers. This suggests that a higher discount rate is a consequence of drug addiction, rather than vice versa, which is slightly counterintuitive: one would expect those who value present consumption over future consequences to self-select as drug users. However, an alternative explanation is that those who smoke for a period and then quit (becoming ex-smokers) are those with low discount rates who began smoking for other reasons (perhaps due to their parents' perception of smoking, for example), whereas those with higher discount rates never find a rational motivation to quit. A further possibility is that discount rates vary over the course of an individual's lifetime, and that people who give up smoking discount at a lower rate than they previously had whilst choosing to continue smoking. Naturally, the true explanation could be a combination of these hypotheses.

These questions greatly complicate a panel data analysis that focuses on changes in wage progression as individuals start or stop smoking, as current research tells us little about how, when, or if these individuals' discount rates change. If individuals' discount rates do not vary over time, then they would be eliminated as a fixed effect in an analysis based on traditional panel techniques.

In this study, then, the sources of the wage differential between smokers and non-smokers will be split into two main categories:

- factors such as insurance premia, maintenance costs, negative effects on morale and discrimination should create a fixed differential, constant throughout an individual's lifetime, whereas
- two reasons for the wage differential can be expected to rise with age:
  - the health differential between smokers and non-smokers is likely to increase with the length of time that the individual has smoked; in practice, more or less with his or her age, and
  - the wage differential due to career and education selection with less emphasis on future earnings can also be expected to increase with age.

3 Econometric Model

Although the data is drawn from a panel dataset, cross-sectional methods are used due to the difficulties associated with the use of panel methods in this case, as described above. The panel is therefore collapsed into a cross-section. Most variables draw on information included in different observations throughout the thirteen years; however, the data used in the following regressions only contains one observation per individual.

The basic model to be estimated takes the form

$$\overline{\Delta W} = \beta_0 + \beta_1 smoker + \beta_2 \overline{A} + \sum_k \beta_k other\ factors + u$$

Where $\overline{\Delta W}$ is an individual's average increase in monthly wage per year, *smoker* is a dummy indicating whether the individual smokes, and $\overline{A}$ is the individual's average age over the period of observation. Since wages follow a broadly parabolic variation with age, age can be expected to linearly influence the first derivative of wage. Different regressions include slightly different forms of dummy to represent the individual's status as a smoker.

4 Estimation Methods

Standard OLS methods are used throughout. In doing so, we implicitly assert that the classical assumptions are valid. The most controversial of these is the assertion that there is no correlation between each individual's status as a smoker and the unobserved factors, denoted by u, which impact wages. Several alternative specifications of the model are included to test the robustness of our findings. However, this problem is never fully resolved and all results should be treated with a degree of caution because of this uncertainty.

5 Notes on the Data

The data is drawn from the British Household Panel Survey (BHPS), waves 1–13 (from 1991 to 2003).

Variable	N	Mean	Std Dev	Minimum	Maximum
avdpaygu	11550	87.55661	383.041	−15317.32	18013.86
avsmokes	15130	.3433134	.4494397	0	1
fullsmoker	15130	.2773298	.4476955	0	1
starter	15130	.0318572	.1756257	0	1
quitter	15130	.0516854	.2213984	0	1
miscsmoker	15130	.0402512	.1965542	0	1
avhlstatus	15128	2.984602	.7107921	0	4
avdhlstatus	11548	−.0160356	.4471706	−3	3
avhllimit	2516	1.807243	.8301798	0	3
avdhllimit	878	.0150323	.8494759	−3	3
female	15130	.5099802	.4999169	0	1
avage	15130	35.83779	12.27337	16	65
avjbhrs	15022	34.51489	10.24987	1	99
avdjbhrs	11423	.1002701	4.784115	−66	74

Table 1: Summary statistics for variables used

6 Results and Analysis

6.1 Outliers

Excluding three outlying observations — those in which the absolute value of the average difference in monthly pay (`avdpaygu`) is greater than £10,000 has significant effects on results. Regressions from 3 onwards exclude these three observations. Excluding less extreme outliers has a greatly reduced impact on results.

6.2 Controlling for Health

As discussed earlier, the deterioration of a smoker's health could be expected to reduce the progression of her wage throughout her lifetime — in other words, to have the same effect as an increased discount rate.

Dependent Variable: avdpaygu			
	(1)	(2)	(3)
avsmokes	-20.88597**	—	—
	(8.257019)		
fullsmoker	—	-21.45612***	-11.72705**
		(8.19521)	(5.64259)
starter	—	-48.27315***	-11.41257
		(17.08491)	(11.77283)
quitter	—	8.645796	12.53107
		(13.60446)	(9.365298)
miscsmoker	—	-13.47426	-9.603062
		(15.21856)	(10.47644)
avhlstatus	3.727307	2.308928	8.335518**
	(5.59533)	(5.316979)	(3.661161)
avdhlstatus	10.18778	15.62972**	7.542791
	(7.927794)	(7.542414)	(5.193297)
female	-36.51473***	-13.32249*	-7.269843
	(7.101025)	(7.758972)	(5.342612)
avage	-3.475643***	-3.022858***	-2.852897***
	(.3030387)	(.292214)	(.2011723)
avjbhrs	—	2.591446***	2.803965***
		(.4139303)	(.2849984)
avdjbhrs	—	13.54845***	13.34235***
		(.7022)	(.4834018)
constant	230.0822***	114.7801***	76.68665***
	(22.06382)	(26.99626)	(18.58911)
Adjusted R^2	0.0136	0.0483	0.0940
No of observations	11548	11422	11419
Outliers removed	no	no	yes

Asterisks denote significance levels: * 90 per cent, ** 95 per cent, *** 99 per cent.

Table 2: regressions 1–3

It is therefore very important to be certain to separate the impact on wages of a high discount rate from that caused by any deterioration in health.

Two indicators of health are used in separate regressions; one based on an individual's subjective analysis of her own general health, the other based on an indvidual's perception of the amount that her state of health impacts her ability to work. In each case, both the individual's average state of health and the average change in her state of health are considered.

Regression 4 uses controls based on `hllimit`, which measure whether an individual's health limits her ability to work. Unfortunately, there were insufficient responses to this question for this measure of health to be statistically significant. Instead, controls based on a measure of general state of health are used in the other regressions (`hlstatus`) — these are less directly relevant, but data availability is much better.

6.3 Smoker Dummies

The simplest smoker dummy considered, `avsmokes`, is the proportion of years in which each individual reported being a smoker (equal to unity if a smoker throughout and zero if a non-smoker throughout). The coefficient on this variable can be interpreted to mean the expected extra increase in monthly wage each year that somebody with a discount rate typical of a smoker could be expected to receive relative to somebody with a discount rate typical of a non-smoker.

Dependent Variable: avdpaygu			
	(4)	(5)	(6)
avsmokes	8.245243	—	—
	(20.93906)		
fullsmoker	—	2.687301	-11.41773**
		(5.587092)	(5.441888)
starter	—	.5738152	—
		(11.57858)	
quitter	—	13.77288	—
		(9.199142)	
miscsmoker	—	-6.9476	—
		(10.29116)	
avhllimit	-.3524222	—	—
	(12.99198)		
avdhllimit	.7352575	—	—
	(10.74303)		
avhlstatus	—	-.7120988	8.703235**
		(3.623296)	(3.645425)
avdhlstatus	—	5.683314	—
		(5.101859)	
female	5.720297	11.07128**	—
	(18.48779)	(5.323955)	
avage	-.9903623	-3.49829***	-2.809899***
	(.8379548)	(.2001084)	(.1988297)
avjbhrs	—	.2916967	2.993719***
		(.3057566)	(.2469924)
avdjbhrs	—	13.58161***	13.33039***
		(.4749595)	(.4832492)
avpaygu	—	.0706037***	—
		(.0034561)	
constant	81.97911*	114.1645***	63.37513***
	(47.94136)	(18.35084)	(16.22065)
Adjusted R^2	-0.0037	0.1259	0.0938
No of observations	878	11419	11420

Table 3: regressions 4–6

However, by showing that ex-smokers display similar discounting rates to never-smokers, Bickel (1999) has raised an interesting question about the difference in discounting rate between those who start, continue and stop smoking, not simply between those who do and do not. For example, it might be proposed that individuals who quit smoking have the lowest discount rate of all — in choosing to fight an addiction, they are selecting a much more costly choice in favour of long-term rather than short-term payoffs than is the case for individuals who continue not smoking. Similarly, it might be expected that an individual who chose to start smoking late in life would have a particularly high discount rate, more so than an individual that continues to smoke.

A second and linked question is that of whether discount rates vary over time for each individual. It is possible that the main difference between smokers' and non-smokers' discount rates is that smokers temporarily discounted faster during a short period in which they started smoking. It might then be rational for them to continue smoking due to the higher costs of quitting relative to costs of continuing to not smoke in a non-smoker. Looking at the issue in this way would suggest that the reason that most smokers begin smoking during their teens is linked to a general tendency to discount faster at this age

than later in life, which seems to be a plausible hypothesis, though beyond the scope of this study.

In order to examine this question, regressions 2, 3 and 5 use the `fullsmoker`, `starter`, `quitter` and `miscsmoker` dummies. These separate smokers into those who smoke throughout the period observed, those who start during the period of observation, those who quit, and those who exhibit some more complicated evolution.

Regressions 2 and 3 broadly support theoretical predictions. Regression 2 suggests that smoking is associated with an elevated discount rate, and that starting to smoke is associated with an even higher discount rate. Coefficients on the other two categories are not statistically significant. The number of observations of individuals who start (481) and quit (782) are relatively small. With a larger sample size, the indication that individuals who quit smoking have a lower discount rate than non-smokers might become statistically significant. However, with available data the hypothesis that those who quit smoking have the same discount rate as non-smokers cannot be rejected, in accordance with the findings of Bickel (1999). In regression 3, the exclusion of three outlying values destroys the statistical significance previously attached to those who start smoking. It cannot reasonably be concluded that there is sufficient evidence that those who start smoking have a different discount rate to non-smokers, nor to continuous smokers.

6.4 Controls

Due to the nature of the hypothesis under test, the question of which controls to include is more than usually complicated.

The relationship of interest is between smoking and discount rate, not directly between smoking and wage progression. Wage progression may be influenced by various intermediate factors such as wage level and working hours, which are themselves affected by the individual's discount rate. *If* these intermediate factors are a consequence of the underlying discount rate, then they should *not* be controlled for.

For example, adding controls for educational level substantially reduces the coefficients associated to smoking. But that is to be expected. An individual with a high discount rate is likely to avoid high levels of education, since the payoffs of education are spread further into the future.

An hypothetical example may clarify this point. Imagine a small town in which the only sources of employment are two factories, each of which has a fixed wage structure based on a starting wage (identical in both factories) and a fixed annual increase. Factory A offers a low annual pay rise, factory B offers a high annual rise, but requires all workers to attend two extra years of school without pay before starting work. Each individual in the town chooses a career for life at one or other factory. Some individuals smoke. Those with higher discount rates are more likely to smoke than those with low discount rates, but many other factors are involved in the decision to start smoking. In contrast, individuals choose their career based *solely* on their discount rate — those with higher than average discount rate choose factory A, whilst those with lower than average choose to continue in education, then join factory B. Now, running a regression between pay rise and status as a smoker which controls for educational level will conclude that there is no relation between smoking and wage differential, because all pay rises are perfectly accounted for by education. However, this does not imply that smoking is not indicative of a high discount rate. In contrast, a regression which did not control for education would return an accurate indication of the difference in discount rate between smokers and non-smokers.

The most difficult application of this problem is in the decision of whether to include average wage level as a control. Clearly, long-term smokers with high discount rates are likely to have lower wages during most of their lives. On the other hand, there are strong arguments that pay level should influence pay increase for reasons that may not be linked to discount rate, such as ability (high ability workers could be expected to have higher wages, and higher wage increases due to a greater ability to improve their productivity over time). Even if ability could be controlled for directly, there is no justification for assuming that there is no correlation between ability and discount rate — perhaps high ability workers discount more slowly.

Regressions with (5) and without (3 and 6) controls for wage are included. No coefficients on smoking remain statistically significant with a control for average wage added, and some even change sign. Economic reasoning alone is not able to determine which is a more appropriate formulation, but this lack of robustness advises caution in drawing confident conclusions.

Finally, regression 6 removes all controls which are not statistically significant. In this, as in regression 3, the hypothesis that individuals who smoked throughout have no higher discount rates than those who did not smoke can be rejected at the 95 per cent level. Regression 6 indicates that the typical non-smoker gains an extra wage increase of £11.42 per month each year due to her lower discount rate, as compared to the typical smoker. Over a career from 16 to 65, this equates to a total difference in earnings of £165,000, without any discounting.[1]

7 Conclusions

Assuming that the exclusion of wage levels does not bias the estimated link between smoking and wage progression, evidence has been found that smoking is associated with an elevated discounting rate. The total difference in expected earnings due to this difference in discount rate (but *not* due to smoking) is very large — approximately £165,000 when not discounted.

No specific conclusions could be drawn about the discount rates of individuals starting and stopping smoking, although this might be due to the relatively small sample size available for these groups. No evidence was found that the differences in discount rates accorded these groups by theory are incorrect.

The inclusion of wage rate as a control changes these conclusions. Although the inclusion of wage rate as a control is highly likely to bias findings, there is no guarantee that its exclusion will result in unbiased estimators. The dependency of all results on the inclusion or exclusion of a control based on wage rate raises questions over the validity of assumptions stated in section 4.

[1] Because $\int_0^{65-16} 12 \times 11.42t \, \mathrm{d}t = \frac{1}{2} \times 49^2 \times 12 \times 11.42 = 164,527$.

8 Bibliography

Auld, Christopher (1998), "Wages, Alcohol Use and Smoking: Simultaneous Estimates" *Economics Working Paper Archive EconWPA* HEW 9808001

Bertera, Robert (1991), "The Effects of Behavioural Risks on Absenteeism and Health-Care Costs in the Workplace," *Journal of Occupational Medicine* Vol 33, No 11 (November 1991) pp 1119–1123

Bickel, Warren, Amy Odum and Gregory Madden (1999), "Impulsivity and Cigarette Smoking: Delay Discounting in Current, Never, and Ex-Smokers" *Pyschomarmacology* 146:447–454

Kristein, Marvin (1983), "How Much Can Business Expect to Profit from Smoking Cessation?" *Preventive Medicine* Vol 12, pp 358–381

Levine, Phillip, Tara Gustafson and Ann Velenchik (1995), "More Bad News for Smokers? The Effects of Cigarette Smoking on Labour Market Outcomes," *National Bureau of Economic Research Working* Working Paper 5270

Wooldridge, Jeffrey (2003), *Introductory Econometrics: A Modern Approach* (2nd ed.), Thompson/ Southern-Western

9 Appendix

9.1 Source Code

```
log using d:/msc/smokers-discount-rates.log, replace
clear
cd d:/msc/bhps
scalar y = 1991
foreach wave in a b c d e f g h i j k l m {
drop _all
label drop _all
matrix drop _all
cluster drop _all
eq drop _all
constraint drop _all
postutil clear
_return drop _all
discard
mata: mata clear
use `wave'indresp
renpfix `wave'
if "`wave'"=="i" {
* wave i has different questions
keep pid paygu smnow ncigs hlsf1 sex age jbsic jbsoc jbhrs
    jbstat race qfachi qfvoc
rename smnow smoker
rename hlsf1 hlstat
}
else {
keep pid paygu smoker ncigs hlstat hlltw hlendw hlltwa sex age
    jbsic jbsoc jbhrs jbstat race qfachi qfvoc
}
gen year=y
disp y
save `wave', replace
scalar y = y + 1
}
clear
use a
```

```
foreach wave in b c d e f g h i j k l m {
append using ‘wave’
}
save 13waves, replace
tsset pid year
* drop all too old, too young, not working
* 1: male, 2: female
drop if age > 65 & sex == 1
drop if age > 60 & sex == 2
drop if age < 16
drop if jbstat != 1 & jbstat != 2
gen female=1 if sex==2
replace female=0 if sex==1
sort pid year
********** INCOME VARIABLES **********
* lose anyone without pay information
drop if paygu <= 0
gen lnpaygu = ln(paygu)
by pid: gen dlnpaygu = lnpaygu - lnpaygu[_n-1]
by pid: egen avdlnpaygu = mean(dlnpaygu)
by pid: egen avlnpaygu = mean(lnpaygu)
by pid: gen dpaygu = paygu - paygu[_n-1]
by pid: egen avdpaygu = mean(dpaygu)
by pid: egen avpaygu = mean(paygu)
by pid: egen avage = mean(age)
gen sqavage = avage^2
gen propincrease=avdpaygu/avpaygu
********** WORKING HOURS VARIABLES **********
replace jbhrs=. if jbhrs<=0
gen fulltime=0
replace fulltime=1 if jbhrs>=20
by pid: egen avjbhrs = mean(jbhrs)
by pid: egen avfulltime = mean(fulltime)
by pid: gen djbhrs = jbhrs - jbhrs[_n-1]
by pid: egen avdjbhrs = mean(djbhrs)
********** HEALTH VARIABLES **********
* hlstatus rates health on a 0-4 scale, 4 meaning excellent
gen hlstatus = 5 - hlstat
replace hlstatus=. if hlstat < 1
by pid: egen avhlstatus = mean(hlstatus)
by pid: gen dhlstatus = hlstatus - hlstatus[_n-1]
by pid: egen avdhlstatus = mean(dhlstatus)
gen hlworsening=.
replace hlworsening=1 if avdhlstatus<0
replace hlworsening=0 if avdhlstatus>=0
gen hlimproving=.
replace hlimproving=1 if avdhlstatus>0
replace hlimproving=0 if avdhlstatus<=0
* hllimit varies 0-3
*  0: poor health, high impact on work
*  3: good health, no impact on work
gen hllimit=hlltwa-1
replace hllimit=. if hllimit<0
by pid: gen dhllimit = hllimit - hllimit[_n-1]
by pid: egen avdhllimit = mean(dhllimit)
by pid: egen avhllimit = mean(hllimit)
* mark first and last observation
```

```
by pid: gen lag = female - female[_n-1]
by pid: gen lead = female - female[_n+1]
gen first=0
replace first=1 if lag==.
gen last=0
replace last=1 if lead==.
********** SMOKING VARIABLES **********
* smoker dummy
gen smokes=1 if smoker==1
replace smokes=0 if smoker==2
drop if smokes!=1 & smokes!=0
replace ncigs=. if ncigs<0
* "inapplicable" has multiple meanings here
replace ncigs=0 if smoker==2
by pid: egen avsmokes = mean(smokes)
by pid: egen avncigs = mean(ncigs)
by pid: gen smokesatstart = smokes[1]
by pid: gen smokesatend = smokes[_N]
gen quitter=0
replace quitter=1 if smokesatstart==1 & smokesatend==0
gen starter=0
replace starter=1 if smokesatstart==0 & smokesatend==1
gen fullsmoker=0
replace fullsmoker=1 if avsmokes==1
gen neversmoker=0
replace neversmoker=1 if avsmokes==0
gen miscsmoker=0
replace miscsmoker=1 if quitter==0 & starter==0 & fullsmoker==0 &
  neversmoker==0
gen eversmoker=0
replace eversmoker=1 if avsmokes>0
********** INTERACTION BETWEEN SMOKING AND AGE **********
gen smavage=avage*fullsmoker
gen smsqavage=sqavage*fullsmoker
gen sqavpaygu = avpaygu^2
gen cuavpaygu = avpaygu^3
gen quavpaygu = avpaygu^4
* COLLAPSE HERE
* get rid of all but the first observation for each individual, to collapse
* this back to a cross-section
* each individual's first entry contains all the information we need now
keep if first==1
********** summarise data **********
sum avdpaygu avdlnpaygu avpaygu
sum avsmokes fullsmoker starter quitter miscsmoker avncigs avncigs
sum avhlstatus avdhlstatus avhllimit avdhllimit
sum female avage avjbhrs avdjbhrs
********** REGRESSIONS **********
* reduced controls
regress avdpaygu avsmokes female avage avhlstatus avdhlstatus
regress avdpaygu fullsmoker starter quitter miscsmoker female avage avhlstatua
  avdhlstatus avjbhrs avdjbhrs
* remove outliers
keep if avdpaygu<10000 & avdpaygu>-10000
regress avdpaygu fullsmoker starter quitter miscsmoker female avage avhlstatus
  avdhlstatus avjbhrs avdjbhrs
* alternative health specification
```

```
regress avdpaygu avsmokes female avage avhllimit avdhllimit
* added control for wage
regress avdpaygu fullsmoker starter quitter miscsmoker female avage avhlstatus
  avdhlstatus avjbhrs avdjbhrs avpaygu
* finally, without controls that aren't statistically significant
regress avdpaygu fullsmoker avage avhlstatus avjbhrs avdjbhrs
log close
```

9.2 Log Output

```
-------------------------------------------------------------------------------
       log:  d:/msc/smokers-discount-rates.log
  log type:  text
 opened on:  27 Jan 2006, 14:03:45
. clear
. cd d:/msc/bhps
d:\msc\bhps
. scalar y = 1991
. foreach wave in a b c d e f g h i j k l m {
  2.
. drop _all
  3.
. label drop _all
  4.
. matrix drop _all
  5.
. cluster drop _all
  6.
. eq drop _all
  7.
. constraint drop _all
  8.
. postutil clear
  9.
. _return drop _all
 10.
. discard
 11.
. mata: mata clear
 12.
. use `wave'indresp
 13.
. renpfix `wave'
 14.
. if "`wave'"=="i" {
 15.
. * wave i has different questions
. keep pid paygu smnow ncigs hlsf1 sex age jbsic jbsoc jbhrs jbstat race
  qfachi qfvoc
 16.
. rename smnow smoker
 17.
. rename hlsf1 hlstat
 18.
. }
 19.
. else {
```

```
 20.
. keep pid paygu smoker ncigs hlstat hlltw hlendw hlltwa sex age jbsic jbsoc
  jbhrs jbstat race qfachi qfvoc
 21.
. }
 22.
. gen year=y
 23.
. disp y
 24.
. save ‘wave’, replace
 25.
. scalar y = y + 1
 26.
. }
1991
file a.dta saved
1992
file b.dta saved
1993
file c.dta saved
1994
file d.dta saved
1995
file e.dta saved
1996
file f.dta saved
1997
file g.dta saved
1998
file h.dta saved
1999
file i.dta saved
2000
file j.dta saved
2001
file k.dta saved
2002
file l.dta saved
2003
file m.dta saved
. clear
. use a
. foreach wave in b c d e f g h i j k l m {
  2.
. append using ‘wave’
  3.
. }
age was byte now int
. save 13waves, replace
file 13waves.dta saved
. tsset pid year
       panel variable:  pid, 10002251 to 1.398e+08
        time variable:  year, 1991 to 2003, but with gaps
. * drop all too old, too young, not working
. * 1: male, 2: female
. drop if age > 65 & sex == 1
```

```
(11587 observations deleted)
. drop if age > 60 & sex == 2
(20851 observations deleted)
. drop if age < 16
(564 observations deleted)
. drop if jbstat != 1 & jbstat != 2
(39810 observations deleted)
. gen female=1 if sex==2
(47643 missing values generated)
. replace female=0 if sex==1
(47643 real changes made)
. sort pid year
. ********** INCOME VARIABLES **********
. * lose anyone without pay information
. drop if paygu <= 0
(14627 observations deleted)
. gen lnpaygu = ln(paygu)
. by pid: gen dlnpaygu = lnpaygu - lnpaygu[_n-1]
(16105 missing values generated)
. by pid: egen avdlnpaygu = mean(dlnpaygu)
(3747 missing values generated)
. by pid: egen avlnpaygu = mean(lnpaygu)
. by pid: gen dpaygu = paygu - paygu[_n-1]
(16105 missing values generated)
. by pid: egen avdpaygu = mean(dpaygu)
(3747 missing values generated)
. by pid: egen avpaygu = mean(paygu)
. by pid: egen avage = mean(age)
. gen sqavage = avage^2
. gen propincrease=avdpaygu/avpaygu
(3747 missing values generated)
. ********** WORKING HOURS VARIABLES **********
. replace jbhrs=. if jbhrs<=0
(1211 real changes made, 1211 to missing)
. gen fulltime=0
. replace fulltime=1 if jbhrs>=20
(67623 real changes made)
. by pid: egen avjbhrs = mean(jbhrs)
(157 missing values generated)
. by pid: egen avfulltime = mean(fulltime)
. by pid: gen djbhrs = jbhrs - jbhrs[_n-1]
(17647 missing values generated)
. by pid: egen avdjbhrs = mean(djbhrs)
(4173 missing values generated)
. ********** HEALTH VARIABLES **********
. * hlstatus rates health on a 0-4 scale, 4 meaning excellent
. gen hlstatus = 5 - hlstat
. replace hlstatus=. if hlstat < 1
(18 real changes made, 18 to missing)
. by pid: egen avhlstatus = mean(hlstatus)
(2 missing values generated)
. by pid: gen dhlstatus = hlstatus - hlstatus[_n-1]
(16127 missing values generated)
. by pid: egen avdhlstatus = mean(dhlstatus)
(3751 missing values generated)
. gen hlworsening=.
(75465 missing values generated)
```

```
. replace hlworsening=1 if avdhlstatus<0
(21784 real changes made)
. replace hlworsening=0 if avdhlstatus>=0
(53681 real changes made)
. gen hlimproving=.
(75465 missing values generated)
. replace hlimproving=1 if avdhlstatus>0
(19789 real changes made)
. replace hlimproving=0 if avdhlstatus<=0
(55676 real changes made)
. * hllimit varies 0-3
. *  0: poor health, high impact on work
. *  3: good health, no impact on work
. gen hllimit=hlltwa-1
(7269 missing values generated)
. replace hllimit=. if hllimit<0
(62995 real changes made, 62995 to missing)
. by pid: gen dhllimit = hllimit - hllimit[_n-1]
(73587 missing values generated)
. by pid: egen avdhllimit = mean(dhllimit)
(69237 missing values generated)
. by pid: egen avhllimit = mean(hllimit)
(59523 missing values generated)
. * mark first and last observation
. by pid: gen lag = female - female[_n-1]
(16105 missing values generated)
. by pid: gen lead = female - female[_n+1]
(16105 missing values generated)
. gen first=0
. replace first=1 if lag==.
(16105 real changes made)
. gen last=0
. replace last=1 if lead==.
(16105 real changes made)
. ********** SMOKING VARIABLES **********
. * smoker dummy
. gen smokes=1 if smoker==1
(53536 missing values generated)
. replace smokes=0 if smoker==2
(50840 real changes made)
. drop if smokes!=1 & smokes!=0
(2696 observations deleted)
. replace ncigs=. if ncigs<0
(51099 real changes made, 51099 to missing)
. * "inapplicable" has multiple meanings here
. replace ncigs=0 if smoker==2
(50840 real changes made)
. by pid: egen avsmokes = mean(smokes)
. by pid: egen avncigs = mean(ncigs)
(22 missing values generated)
. by pid: gen smokesatstart = smokes[1]
. by pid: gen smokesatend = smokes[_N]
. gen quitter=0
. replace quitter=1 if smokesatstart==1 & smokesatend==0
(5389 real changes made)
. gen starter=0
. replace starter=1 if smokesatstart==0 & smokesatend==1
```

```
(2859 real changes made)
. gen fullsmoker=0
. replace fullsmoker=1 if avsmokes==1
(14946 real changes made)
. gen neversmoker=0
. replace neversmoker=1 if avsmokes==0
(44687 real changes made)
. gen miscsmoker=0
. replace miscsmoker=1 if quitter==0 & starter==0 & fullsmoker==0 &
  neversmoker==0
(4888 real changes made)
. gen eversmoker=0
. replace eversmoker=1 if avsmokes>0
(28082 real changes made)
. ********** INTERACTION BETWEEN SMOKING AND AGE **********
. gen smavage=avage*fullsmoker
. gen smsqavage=sqavage*fullsmoker
. gen sqavpaygu = avpaygu^2
. gen cuavpaygu = avpaygu^3
. gen quavpaygu = avpaygu^4
. * COLLAPSE HERE
. * get rid of all but the first observation for each individual, to collapse
. * this back to a cross-section
. * each individual's first entry contains all the information we need now
. keep if first==1
(57639 observations deleted)
. ********** summarise data **********

. sum avdpaygu avdlnpaygu avpaygu

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    avdpaygu |     11550    87.55661     383.041  -15317.32   18013.86
  avdlnpaygu |     11550    .0865361    .2730536  -2.761173   8.363042
     avpaygu |     15130    1168.875    802.0532         26      11323

. sum avsmokes fullsmoker starter quitter miscsmoker avncigs avncigs

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    avsmokes |     15130    .3433134    .4494397          0          1
  fullsmoker |     15130    .2773298    .4476955          0          1
     starter |     15130    .0318572    .1756257          0          1
     quitter |     15130    .0516854    .2213984          0          1
  miscsmoker |     15130    .0402512    .1965542          0          1
-------------+--------------------------------------------------------
     avncigs |     15114    5.001417    7.999495          0   58.57143
     avncigs |     15114    5.001417    7.999495          0   58.57143

. sum avhlstatus avdhlstatus avhllimit avdhllimit

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
  avhlstatus |     15128    2.984602    .7107921          0          4
 avdhlstatus |     11548   -.0160356    .4471706         -3          3
   avhllimit |      2516    1.807243    .8301798          0          3
  avdhllimit |       878    .0150323    .8494759         -3          3
```

```
. sum female avage avjbhrs avdjbhrs

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      female |     15130    .5099802    .4999169          0          1
       avage |     15130    35.83779    12.27337         16         65
     avjbhrs |     15022    34.51489    10.24987          1         99
    avdjbhrs |     11423    .1002701    4.784115        -66         74

. ********** REGRESSIONS **********
. * reduced controls
. regress avdpaygu avsmokes female avage avhlstatus avdhlstatus

      Source |       SS       df       MS              Number of obs =   11548
-------------+------------------------------           F(  5, 11542) =   32.78
       Model |  23726716.8     5  4745343.37           Prob > F      =  0.0000
    Residual |  1.6707e+09 11542  144753.328           R-squared     =  0.0140
-------------+------------------------------           Adj R-squared =  0.0136
       Total |  1.6945e+09 11547  146745.443           Root MSE      =  380.46

------------------------------------------------------------------------------
    avdpaygu |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    avsmokes |  -20.88597   8.257019    -2.53   0.011    -37.07113   -4.700811
      female |  -36.51473   7.101025    -5.14   0.000    -50.43395   -22.59552
       avage |  -3.475643   .3030387   -11.47   0.000    -4.069651   -2.881636
  avhlstatus |   3.727307    5.59533     0.67   0.505    -7.240487     14.6951
 avdhlstatus |   10.18778   7.927794     1.29   0.199    -5.352036     25.7276
       _cons |   230.0822   22.06382    10.43   0.000     186.8334    273.3311
------------------------------------------------------------------------------

. regress avdpaygu fullsmoker starter quitter miscsmoker female avage
  avhlstatus avdhlstatus avjbhrs avdjbhrs

      Source |       SS       df       MS              Number of obs =   11422
-------------+------------------------------           F( 10, 11411) =   58.95
       Model |  75751922.8    10  7575192.28           Prob > F      =  0.0000
    Residual |  1.4663e+09 11411  128495.796           R-squared     =  0.0491
-------------+------------------------------           Adj R-squared =  0.0483
       Total |  1.5420e+09 11421  135015.975           Root MSE      =  358.46

------------------------------------------------------------------------------
    avdpaygu |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  fullsmoker |  -21.45612    8.19521    -2.62   0.009    -37.52014   -5.392096
     starter |  -48.27315   17.08491    -2.83   0.005    -81.76252   -14.78379
     quitter |   8.645796   13.60446     0.64   0.525    -18.02128    35.31287
  miscsmoker |  -13.47426   15.21856    -0.89   0.376    -43.30525    16.35673
      female |  -13.32249   7.758972    -1.72   0.086    -28.53141    1.886426
       avage |  -3.022858    .292214   -10.34   0.000    -3.595648   -2.450068
  avhlstatus |   2.308928   5.316979     0.43   0.664    -8.113264    12.73112
 avdhlstatus |   15.62972   7.542414     2.07   0.038     .8452911    30.41415
     avjbhrs |   2.591446   .4139303     6.26   0.000     1.780072    3.402821
    avdjbhrs |   13.54845      .7022    19.29   0.000     12.17202    14.92489
       _cons |   114.7801   26.99626     4.25   0.000     61.86279    167.6974
------------------------------------------------------------------------------
```

```
. * remove outliers
. keep if avdpaygu<10000 & avdpaygu>-10000
(3583 observations deleted)
. regress avdpaygu fullsmoker starter quitter miscsmoker female avage
  avhlstatus avdhlstatus avjbhrs avdjbhrs

      Source |       SS       df       MS              Number of obs =   11419
-------------+------------------------------           F( 10, 11408) =  119.49
       Model |  72762561.8    10  7276256.18           Prob > F      =  0.0000
    Residual |   694656270 11408  60892.0293           R-squared     =  0.0948
-------------+------------------------------           Adj R-squared =  0.0940
       Total |   767418832 11418  67211.3183           Root MSE      =  246.76

------------------------------------------------------------------------------
    avdpaygu |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  fullsmoker |  -11.72705    5.64259    -2.08   0.038    -22.78749   -.6666028
     starter |  -11.41257   11.77283    -0.97   0.332    -34.48934     11.6642
     quitter |   12.53107   9.365298     1.34   0.181    -5.826522    30.88867
  miscsmoker |  -9.603062   10.47644    -0.92   0.359    -30.13868    10.93255
      female |  -7.269843   5.342612    -1.36   0.174    -17.74228    3.202595
       avage |  -2.852897   .2011723   -14.18   0.000     -3.24723   -2.458565
  avhlstatus |   8.335518   3.661161     2.28   0.023     1.159013    15.51202
 avdhlstatus |   7.542791   5.193297     1.45   0.146    -2.636964    17.72255
     avjbhrs |   2.803965   .2849984     9.84   0.000      2.24532    3.362611
    avdjbhrs |   13.34235   .4834018    27.60   0.000      12.3948     14.2899
       _cons |   76.68665   18.58911     4.13   0.000     40.24879    113.1245
------------------------------------------------------------------------------

. * alternative health specification
. regress avdpaygu avsmokes female avage avhllimit avdhllimit

      Source |       SS       df       MS              Number of obs =     878
-------------+------------------------------           F(  5,   872) =    0.36
       Model |  131146.275     5   26229.255           Prob > F      =  0.8762
    Residual |    63620354   872  72959.1215           R-squared     =  0.0021
-------------+------------------------------           Adj R-squared = -0.0037
       Total |  63751500.3   877  72692.7027           Root MSE      =  270.11

------------------------------------------------------------------------------
    avdpaygu |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    avsmokes |   8.245243   20.93906     0.39   0.694     -32.8516    49.34209
      female |   5.720297   18.48779     0.31   0.757    -30.56547    42.00606
       avage |  -.9903623   .8379548    -1.18   0.238    -2.635006    .6542817
   avhllimit |  -.3524222   12.99198    -0.03   0.978    -25.85162    25.14677
  avdhllimit |   .7352575   10.74303     0.07   0.945    -20.34995    21.82047
       _cons |   81.97911   47.94136     1.71   0.088    -12.11483    176.0731
------------------------------------------------------------------------------

. * added control for wage
. regress avdpaygu fullsmoker starter quitter miscsmoker female avage
  avhlstatus avdhlstatus avjbhrs avdjbhrs avpaygu

      Source |       SS       df       MS              Number of obs =   11419
-------------+------------------------------           F( 11, 11407) =  150.54
```

```
       Model |  97280701.3    11  8843700.12           Prob > F      =  0.0000
    Residual |   670138131 11407  58747.9732           R-squared     =  0.1268
-------------+------------------------------           Adj R-squared =  0.1259
       Total |   767418832 11418  67211.3183           Root MSE      =  242.38

------------------------------------------------------------------------------
    avdpaygu |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  fullsmoker |   2.687301   5.587092     0.48   0.631     -8.26436    13.63896
     starter |   .5738152   11.57858     0.05   0.960     -22.1222    23.26983
     quitter |   13.77288   9.199142     1.50   0.134    -4.259015    31.80478
  miscsmoker |    -6.9476   10.29116    -0.68   0.500    -27.12005    13.22485
      female |   11.07128   5.323955     2.08   0.038     .6354104    21.50715
       avage |   -3.49829   .2001084   -17.48   0.000    -3.890537   -3.106043
  avhlstatus |  -.7120988   3.623296    -0.20   0.844    -7.814381    6.390184
 avdhlstatus |   5.683314   5.101859     1.11   0.265    -4.317208    15.68384
     avjbhrs |   .2916967   .3057566     0.95   0.340    -.3076388    .8910322
    avdjbhrs |   13.58161   .4749595    28.60   0.000     12.65061    14.51261
     avpaygu |   .0706037   .0034561    20.43   0.000     .0638292    .0773781
       _cons |   114.1645   18.35084     6.22   0.000     78.19371    150.1353
------------------------------------------------------------------------------

. * finally, without controls that aren't statistically significant
. regress avdpaygu fullsmoker avage avhlstatus avjbhrs avdjbhrs

      Source |       SS       df       MS              Number of obs =   11420
-------------+------------------------------           F(  5, 11414) =  237.29
       Model |  72258667.6     5  14451733.5           Prob > F      =  0.0000
    Residual |   695163024 11414  60904.4177           R-squared     =  0.0942
-------------+------------------------------           Adj R-squared =  0.0938
       Total |   767421691 11419  67205.6828           Root MSE      =  246.79

------------------------------------------------------------------------------
    avdpaygu |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  fullsmoker |  -11.41773   5.441888    -2.10   0.036    -22.08477   -.7506985
       avage |  -2.809899   .1988297   -14.13   0.000    -3.199639   -2.420158
  avhlstatus |   8.703235   3.645425     2.39   0.017     1.557576    15.84889
     avjbhrs |   2.993719   .2469924    12.12   0.000     2.509571    3.477866
    avdjbhrs |   13.33039   .4832492    27.58   0.000     12.38314    14.27764
       _cons |   63.37513   16.22065     3.91   0.000     31.57987    95.17038
------------------------------------------------------------------------------

. log close
      log:  d:/msc/smokers-discount-rates.log
  log type:  text
 closed on:  27 Jan 2006, 14:04:55
------------------------------------------------------------------------------
```