ABSTRACT

BURNING A HOLE IN YOUR POCKET: THE EFFECT OF SMOKING CIGARETTES ON WAGES

by Zach Wesley Sanderson

This study measures the impact of smoking on wages for young adults, aged 18 to 30. Economic theory would suggest that smoking can potentially carry a negative wage effect. Smoking carries a number of health effects that have the ability to decrease a person's productivity, reducing their marginal product of labor. Economic theory states that employers set a worker's wage at the marginal product of labor. Therefore, if an individual experiences decreased productivity due to smoking, they theoretically could have a low wage. By applying OLS and first differences methods to individual and sibling pair cross-section data from the 1997 National Longitudinal Survey of Youth and following the research method outlined in Levine et al. (1997), I find that smoking cigarettes does not have a statistically significant impact on the wages of young adults. The point estimates from the OLS and first differences models lie between 6% and 11%, which match the results of previous studies that have found between a 4% and 11% negative wage effect associated with smoking. These results are confirmed by a series of robustness tests. In addition, the results of the OLS and first difference models are extremely similar to the results obtained by Levine et al., who find a statistically significant negative wage effect associated with smoking. The fact that my results line up with previous literature may suggest that smoking does carry a negative wage effect. This paper adds to the current literature by providing more research on the effects of smoking on a younger population, as well as providing more research to help validate the results of Levine et al. (1997).

BURNING A HOLE IN YOUR POCKET: THE EFFECT OF SMOKING CIGARETTES ON WAGES

Thesis

Submitted to the

Faculty of Miami University

in partial fulfillment of

the requirements for the degree of

Master's in Economics

by

Zach Wesley Sanderson

Miami University

Oxford, Ohio

2018

Advisor: Professor Gerald Granderson

Reader: Professor Austin Smith

Reader: Professor Kimberly Berg

This Thesis titled

BURNING A HOLE IN YOUR POCKET: THE EFFECT OF SMOKING CIGARETTES ON WAGES

by

Zach Wesley Sanderson

has been approved for publication by

the Farmer School of Business

and

Department of Economics

____

Professor Gerald Granderson, Advisor

____

Professor Austin Smith, Reader

____

Professor Kimberly Berg, Reader

Table of Contents

List of Tables

Acknowledgments

I would like to thank my advisor, Professor Gerald Granderson of the Farmer School of Business at Miami University for his patience and support through this process. I would also like to thank Professor Austin Smith and Professor Kimberly Berg for being the readers on my committee, and for their helpful comments with the goal of improving this paper. This paper would not be possible without the effort and dedication that they have shown in this process

1. Introduction

Despite efforts to reduce the rate of cigarette smoking in the United States, this phenomenon still occurs with a widespread impact among those who participate. The negative health effects of smoking cigarettes have been well documented throughout the years. Even though the health consequences are apparent in cigarette smoking, the rates of smoking have only slowly declined over the past decade. Jamal et al. (2018) examine data from the 2016 National Health Interview Survey and find that in 2005 roughly 20.9% of Americans were current smokers, which has declined slightly to a rate of 15.5% of Americans smoking as of 2016. Overall cigarette smoking remains a part of our society, so it is important to monitor just how impactful smoking can be.

The purpose of this paper is to estimate the impact of smoking cigarettes on wages for young adults (defined here as ages 18 to 30). Economic theory would suggest that smoking can potentially carry a negative wage effect. Smoking carries a number of health effects that have the ability to decrease a person's productivity, reducing their marginal product of labor. Economic theory states that employers set a worker's wage at the marginal product of labor. Therefore, if an individual experiences decreased productivity due to smoking, they theoretically could have a low wage. All of the papers, with the exception of Levine et al. (1997), that I discuss in the next section deal with the effects of smoking on older populations, using samples where the average age is 35 and above. With a legal smoking age of 18 years old, individuals can potentially enter their 30's with at least a decade of smoking under their belts. Therefore, it's important to examine how early and immediate the effects of smoking on wages can be. This paper will add to the current smoking literature by contributing to current research regarding the effects of smoking on a younger population, rather than an older one.

The U.S. Department of Health and Human Services (HHS) (2010) finds that smokers who quit by the age of 30 have a higher chance of recovering their health to match that of a non-smoker, so this study can provide more motivation for smokers to quit before an age where there is high chance of irreparable damage to the body due to smoking. To further this cause, I will be using panel research methods to analyze individuals and pairs of siblings from two cross-sections in the 1997 National Longitudinal Survey of Youth (NLSY97). I will be following the methods outlined by Levine et al. (1997), who examine the effects of smoking on wages using the 1979

NLSY. By applying their research methods to a modern context, I not only can extend the current literature regarding smoking wage effects on younger adults but can also potentially strengthen the results found by Levine et al (1997). Additionally, while alternatives to cigarettes such as vapes and e-cigarettes have become popular, especially among younger adults with the introduction of the Juul, I choose to focus on cigarette smoking in order to narrow the focus of this paper and prevent potential complications due to the effects of other substances.

While the rate of smoking has slowly declined, the rate of mortality due to smoking cigarettes has not changed in the past decade. HHS (2014) conducts a comprehensive study on the past fifty years of smoking and finds that smoking remains the leading cause of preventable death in the United States, with the death toll of smoking staying above 400,000 deaths per year since 2005. Those that are lucky (or unlucky) enough to not die prematurely must live with various diseases that are directly caused by smoking. The last comprehensive national study of morbidity due to cigarette smoking was conducted in 2000 by the Center for Disease Control (CDC). Using data from the Behavioral Risk Factor Surveillance System, the National Health and Nutrition Examination Survey III, and the U.S. Census, the CDC estimates that 8.6 million Americans are living with morbidity due to smoking related diseases with a total of 12.7 million smoking related diseases present. The leading disease is chronic bronchitis, which makes up about 49% of total smoking related diseases, followed by emphysema, which makes up about 24% of the total. These two diseases are characterized by difficulty breathing and performing tasks, which are some of the milder symptoms of smoking related diseases. Following the two most prevalent diseases are heart attacks (13% of total diseases), all cancers except lung cancer (7% of total diseases), strokes (7% of total diseases), and lung cancer (1% of total diseases).

With so many Americans living with such debilitating diseases, this begs the question as to how smoking cigarettes, with so many direct and indirect effects (which I discuss in further detail in the next section) can affect areas of life other than health. For this paper I decide to focus on the effects of smoking on wages to determine what kind of financial impact is caused by smoking. By applying OLS and first differences methods to individual and sibling pair cross-section data from the 1997 National Longitudinal Survey of Youth and following the research method outlined in Levine et al. (1997), I find that smoking cigarettes does not have a statistically significant impact on the wages of young adults. The point estimates from the OLS and first differences models lie between 6% and 11%, which match the results of previous studies that have found between a 4% and 11% negative wage effect associated with smoking. In addition, the results of the OLS and first difference models are extremely similar to the results obtained by Levine et al., who found a statistically significant negative wage effect associated with smoking. The fact that my results line up with previous literature may suggest that smoking does carry a negative wage effect.

2. Background and Literature Review

Smoking carries various detrimental economic effects that can affect employees and employers alike. Various studies have examined the direct and indirect economic costs caused by smoking. The economic costs of smoking can be summarized into three categories: direct medical and healthcare costs, lost wages as a result of premature mortality, and indirect costs as a result of lost productivity at work. These economic costs can potentially translate into negative wage consequences for a smoker.

With a slew of maladies associated with smoking, it's easy to see how smoking can carry direct medical costs. Smoking can affect the entire body because the toxins in cigarettes circulate through the bloodstream, spreading their negative effects throughout all parts of the body. A person that has been smoking for a long time suffers from permanent respiratory problems, as well as increased chances of heart attacks and strokes due to blood clotting. These serious problems directly translate into high healthcare costs for individuals and for employers who offer healthcare to their employees. Xu et al. (2015) estimates the healthcare costs related to smoking using data from the 2006-2010 Medical Expenditure Panel Survey and the 2004-2009 National Health Interview Survey and find that as of 2010, approximately 8.7% of total annual healthcare spending was attributed to smoking. This 8.7% represents a total of 170 billion dollars in direct healthcare expenses due to smoking. Suárez-Bonel et al. (2015) likewise estimate healthcare costs related to smoking. Using a cross-section of patients in an urban healthcare district in Europe, they find that smokers were more than twice as likely to generate high healthcare costs for their employers compared to nonsmokers. On average, direct healthcare costs in 2011 for the cross-section was 474.71 euros for non-smokers compared to 848.64 euros for smokers.

Smokers tend to have more health complications compared to non-smokers. This potentially means that a smoker will have to visit the doctor more often than a non-smoker as there is a chance that they will need more medical attention for their increased health complications. Insurance companies are aware of this fact and adjust coverage rates accordingly. Tobacco use is one of the factors used to calculate insurance premiums for group coverages, which means that companies with employees that smoke have a higher chance of receiving increased health insurance costs compared to companies without smokers. Under the Affordable Care Act, smokers can be charged up to 50% more for insurance than non-smokers through a tobacco surcharge (also referred to as a premium incentive).[1] In order to charge smoking employees more, employers must follow certain guidelines, the foremost having some sort of wellness program in place to help employees quit smoking. By following these guidelines, employers can reward non-smoking employees with up to a 50% discount (depending on the state) on their monthly health insurance costs. Therefore, a smoker can potentially pay twice as much for their health insurance compared to a non-smoker, which could negatively impact their wages.

Beyond higher health insurance costs, smokers can bring other costs to the employer as well. Examples of this can be increased facility maintenance costs in order to provide more ventilation or cleaning costs associated with removing the after-effects of smoking, such as discarded cigarette butts. Workplaces in which smoking is allowed on average spend $728 more per 1000 square feet annually than non-smoking facilities in terms of cleaning and maintenance costs. This translates into spending between $8,736 and $13,832 annually given the average size of commercial buildings in the U.S.[2] Therefore employers that have smoking facilities may not be able to offer as high of wages as a comparable employer with a non-smoking facility due to these increased costs. A smoker could therefore potentially make the tradeoff of lower wages for a smoking facility, which is yet another way in which smoking could negatively affect wages.

Smoking produces a number of negative health consequences, defined here as factors that negatively impact health and the body's performance. These negative health consequences have the capability to impact productivity within the workplace. One major health consequence is reduced oxygen levels in the blood. The carbon monoxide within cigarette smoke binds to the body's red blood cells, displacing the oxygen within the blood and preventing said oxygen from circulating to the muscles and body tissue. The lack of oxygen circulation causes lactic acid buildup, which is the substance that causes fatigue. Bodies of smokers therefore experience

[1] https://www.lung.org/assets/documents/tobacco/factsheet-tobacco-surcharges-v2.pdf
[2] https://www.cdcfoundation.org/businesspulse/tobacco-use-infographic#workplacecosts2

greater levels of fatigue and exhaustion more quickly than non-smokers. This has the capacity to limit physical activities within the workplace, causing activities even as basic as walking up stairs to be harder for smokers. The lack of oxygen also increases the development of respiratory issues, such as bronchitis and emphysema, as discussed previously. These respiratory complications can exacerbate even minor illnesses like the common cold or influenza, as well as increasing the likelihood that an individual develops these illnesses.[3] These illnesses carry side effects such as migraines and fatigue that could potentially hinder an employee's productivity at work, as well as making it more likely that the employee misses work altogether, both of which can potentially lead to a negative wage effect culminating in lost wages from lower production at work or from missing paid work days entirely. Economic theory provides some support for this argument in the fact that employers set their wage levels at the marginal product of labor of their employee. Therefore, an employee with a low marginal product of labor would expect to receive a lower wage.

Halpern et al. (2001) study absence rates and productivity levels of employees in a major U.S. airline and find that current smokers on average missed two more work days per year than non-smokers due to illness. Subjective productivity measures, which consisted of evaluations by coworkers and supervisors regarding factors such as quality of work and amount of work, were highest for non-smokers and lowest for smokers in the study. Bunn III et al. (2006) study panel data from the Wellness Inventory survey between 2001 and 2005 and find similar results in that there is a greater rate of work absence in smokers compared to non-smokers. In their study, it was found that smokers missed on average 6.7 work days per year due to illness compared to only 4.4 days per year for non-smokers.

Bunn III et al. (2006) also find that more wages were lost as a result of employees being hindered at work (presenteeism) rather than employees missing work entirely (absenteeism). Presenteeism resulted in over fifty percent of total lost wages for smokers. Absenteeism losses were calculated by multiplying an imputed average hourly wage (which accounts for salary and benefits) by the average number of days missed due to illness. Presenteeism losses were calculated by multiplying the same average hourly wage by the average number of days spent with illness subtracted by the average number of days of work missed due to illness (this is to estimate the number of days spent at work with the illness). On average, lost wages due to

[3] https://my.clevelandclinic.org/health/articles/10643-smoking-and-physical-activity

absenteeism were $1,156 per year for non-smokers and $1,811 per year for smokers, while lost wages due to presenteeism were $1,466 per year for non-smokers versus $2,619 per year for smokers. From these studies it's clear that smoking can potentially lower wages through missed work and decreased productivity at work.

Lastly, smoking-related mortality and morbidity result in increased costs to the employer in terms of early retirement due to disability and increased costs to the individual in terms of lost potential wages due to an early death. Jha et al. (2013) study data from the U.S. National Health Interview Survey between 1997 and 2004 and find that the life expectancy of smokers was about a decade shorter than non-smokers due to smoking related diseases. Quitting smoking before 40 was found to reduce the risk of smoking-related death by 90%. If an individual did not stop smoking, their probability of survival between the ages of 25 and 79 was roughly half that of a non-smoker. The decreased life expectancy of smokers represents a number of years of lost wages that an individual will not be able to claim.

HHS (2014) conducted a study on the productivity effects of smoking, in which productivity was represented by the present value of total lifetime earnings for an individual. They estimated the number of potential years of life left for individuals between the ages of 35 and 79 in the United States who died between 2005 and 2009 from smoking related diseases. The total years of potential life were then multiplied by an estimate of the present value of future earnings that are lost as a result of early mortality in order to obtain the final estimates for lost earnings as a result of smoking-related mortality. The study found that approximately $105.7 billion was lost in earnings each year due to premature deaths caused by smoking. Additionally, about $5.7 billion was lost in earnings each year due to premature deaths caused by secondhand smoke.

It's possible to see that there are potentially a number of ways in which smoking may affect wages. In order to measure the impact of smoking on wages, I examine the impact of smoking on average hourly wages. This is a popular way of examining the wage effects of smoking and various studies have examined wage loss in this manner. Bondzie (2016) studied hourly wages in Europe and found through Instrument Variable (IV) regression (using health measured by number of hospital visits and weight of individual as IVs) and Matching that smokers on average experienced a 4% to 7% statistically significant decrease in hourly wages. Likewise, Grafova and Stafford (2009) analyzed panel data from the U.S. Panel Study of Income

Dynamics and found on average a 4% to 11% statistically significant decrease in hourly wages for smokers. Finally, Levine et al. (1997) used first difference techniques on panel data from the 1979 U.S. National Longitudinal Survey of Youth and found on average a 4% to 8% decrease in hourly wages for smokers. It's important to note that these studies (in addition to the ones mentioned throughout this section), with the exception of Levine et al. (1997), examined data where the average age was beyond 35. As noted earlier, I plan to study the impacts of smoking on the wages of a younger sample, as is the case with Levine et al. (1997). By examining the impact on a younger population, I can add to the existing literature as well as check the results of Levine et al. (1997) in a modern context.

3. Data

The data for this paper comes from the 1997 National Longitudinal Survey of Youth (NLSY97). This is a U.S.-based annual survey that follows a total of 8,984 men and women, starting in 1997 when the participants were between the ages of 12 and 18. As of the latest round of interviews, conducted in 2016, these participants are between the ages of 30 and 36. The NLSY97 is broken up into two sub samples: a sub sample of 6,748 men and women designed to be representative of the U.S. population, and another sample consisting of 2,236 men and women designed to oversample Latino and black populations in the U.S. For the results of this paper to be representative of the U.S. population, the oversampled sub sample has been removed from the final sample used in this paper.

The NLSY97 contains a wide range of information on each participant, with questions in the survey covering various topics such as employment and wage history, education and ability levels, family characteristics, individual demographic characteristics, health measures, etc. The NLSY97 obtains this information through an annual interview conducted in person or over the phone if necessary. The interviewer is assisted by computer software that automatically selects the next question based on the answers to the current question, prevents interviewers from entering incorrect values, and will warn the interviewers if implausible answers are present. These checks serve the purpose of increasing consistency in the data over time, which serves to reduce some level of measurement error, which is helpful in studies such as this one that use self-reported panel data.

Panel data is subject to attenuation bias from measurement error, especially in situations like this where individuals are self-reporting their income. An individual might forget their income or exaggerate and report an income higher than what they actually make, which can ultimately bias the results when switching over to the fixed effects models. If individuals tend to report their income as higher than it is, the final estimates might be attenuated toward zero, which would understate the true wage effect of smoking. Switching over to the “within” variation in the data has the possibility to take away the variation that was protecting the data from this attenuation bias caused by the measurement error. If coefficient estimates move toward zero when switching from OLS to fixed effects models, then this could indicate that measurement error is involved.

The retention rate of the NLSY97 is fairly high, with approximately 80% of the original sample remaining in the most recent round. The NLSY97 seeks to limit their attrition rates by implementing various processes such as locating efforts, which involve participants leaving contact information for themselves and close relatives. The reasons for attrition are not likely to be connected to the mechanism of smoking or the wage of the participant, and so this attrition should not affect the results of this paper.

For the purposes of this paper, the most important questions cover the participant’s smoking behaviors as well as their employment history, which specifically involves information regarding their yearly income and the number of weeks and hours that the participant has worked within the survey year. Since this paper will be focusing on the effects of smoking on a person’s hourly wage, I have calculated hourly wages for all participants by dividing the participant’s yearly work income in a given year by their total number of work hours in that same year. In order to avoid potential complications on my wage analysis by involving factors like part-time workers or participants that enter and leave the labor force, I am restricting the sample to full time and full year workers. My definition of a full time and full year worker is any worker that has worked at least 50 weeks and 1500 hours during the calendar year. This definition is based on the definition of full time and full year used in the Levine et al. (1997) paper, in which they define full time/full year as at least 50 work weeks and 1750 work hours. Levine et al. (1997) note that the results of their study are robust using a range of 1500 to 2000 work hours as the cutoff point, and to increase the sample size as much as possible, I incorporate the lower boundary into my own definition.

Furthermore, I focus on workers in the years 2006 and 2011. From 2006 to 2011 the participants range from their early-mid 20s to their late 20s and so the majority are in the labor force. This paper's goal is to estimate the effects of smoking on wages for young adults, so the two years in question contain valuable information to this cause. In addition, a five-year gap between the two years should be enough to examine meaningful wage changes in response to smoking due to the fact that wage changes should be more pronounced with longer term smoking. The year, full time, and full year sample restrictions cut the sample down to 4,276 people.

3.1 Individual Descriptive Statistics

In order to create an indicator for whether or not a person is a smoker, I utilize a question in the NLSY97 that asks the participant how many cigarettes they usually smoke each day. A person is considered to be a smoker if the answer to this question is equal to or above 1. Using this data, I create an indicator that is 1 if the participant is a smoker, and 0 if the participant is not. Summary statistics for smokers and non-smokers are shown in Table 1. From Table 1, it's possible to see that 37% of the sample were smokers in 2006, while 29% of the sample were smokers in 2011. The NLSY97 smoking rates are slightly higher than actual smoking rates at the time, but the average smoking rate for these age groups has been between 20% to 30% for the past decade.[4]. Since my sample has a smoking rate of around 30%, I feel that this sample is still representative of their age groups.

Comparing smokers to non-smokers, we can see that there is a statistically significant mean wage difference between the two groups. Non-smokers on average receive a higher hourly wage than smokers, with a 12.2% difference in mean wages in 2006 and an 18.1% difference in 2011. Looking at the other observable characteristics, it's possible to see some potential explanations for this wage gap between the two groups. Compared to non-smokers, smokers are less educated, less likely to be married, less likely to have health insurance, less likely to be female, and more likely to be white. All of these characteristics can affect wages in some way in a manner that is uncorrelated with smoking.

---

[4] The actual 2011 U.S. smoking rate for 25-44 year old adults per the Center for Disease Control's National Health Interview Survey (which lines up with the ages of the sample in 2011) was 22.1%, while the actual 2006 U.S. smoking rate for 18-24 year old adults (which lines up with the ages of the sample in 2006) was 23.7%. According to the National Survey on Drug Use and Health (NSDUH) however, the smoking rate for 21 to 30 year olds has fallen around 30% from 2006 to 2011.

More education can lead to “better” jobs, which would lead to higher wages, which could explain the wage gap in my sample. The concept of a “marriage premium” has been researched in the realm of Economics, which states that married men statistically earn higher wages than single men, which could also account for the wage gap. There are two potential explanations for why marriage could cause a man to earn more money. The first is that men who get married may have qualities that are appealing to both employers and potential spouses. Therefore, a married man may potentially be perceived as a better worker than an unmarried man, which can cause a married man to be more likely to obtain a higher paying job versus an unmarried man. The other explanation is that marriage may cause higher wages by affecting the married man directly. For example, maybe married men obtain a goal of supporting their new partnership, which causes them to work more hours than an unmarried man, ultimately increasing their wages (or potentially the married man misses his freedom and works more hours to delay returning home). The fact that non-smokers are more likely to be married in my sample might explain the wage gap between the two groups.

Health insurance can either positively or negatively affect a person’s wages depending on which story is more likely. A “better” job might not only offer higher wages, but also more benefits. Therefore, a job that offers health insurance might have better pay than a job that does not. For example, under the Affordable Care Act, companies with less than fifty employees are not required to offer health insurance. Therefore, family-owned restaurants with less than fifty employees most likely do not offer as high of wages as larger offices in which health insurance is offered. As a result, individuals with health insurance may be more likely to make more money than those that do not have health insurance. This could explain the wage gap between smokers and non-smokers in my data. On the other hand, health insurance is an additional cost to an employer, and therefore if an employer offers health insurance, it may be likely that they have less money overall, which could indicate that they have less money to pay their employees, resulting in lower wages than companies that do not offer health insurance. Smokers may have a preference of health insurance over wages, since smokers are more likely to develop health complications that require medical assistance. As a result, a smoker may select a lower paying job with health insurance over a higher paying job without health insurance, which would cause smokers on average to have a lower wage than non-smokers. This particular story would not explain the wage gap present in my sample, as health insurance in my sample is associated with a higher average hourly wage. Depending on which story you believe, health insurance can bias my results in either direction, but more importantly this shows that health insurance is an important factor that needs to be accounted for.

White workers statistically earn higher wages than their non-white counterparts, especially when compared to black workers. As of 2015, a white worker with the same education, experience, and living in the same region as a black worker makes more money on average. In 2015, black men on average made 22% less than white men with the same qualifications and characteristics. Similarly, white female workers earn more than their black female counterparts. In 2015, black women on average made 11.7% less than white women with the same qualifications and characteristics (Wilson and Rodgers III 2016). In my sample smokers are more likely to be white, so this notion doesn't necessarily explain the wage gap in my sample but shows that this is a factor that can affect wages in a manner that is not linked to smoking.

Finally, it is statistically proven that female workers earn less than equivalent male counterparts. Since non-smokers have higher mean wages, but also a higher mean number of women, this last observable characteristic would explain a lower mean wage and not a higher one. Therefore, it doesn't explain the wage gap in my sample, but shows that this is another factor that needs to be considered due to its influence on wages outside of smoking.

When considering the observable factors, how they differ among the two groups, and how they can potentially account for wage differentials, it's important that I condition on these observable factors or my estimates regarding the effects of smoking will be biased. For example, if marriage were to be omitted from the regressions, assuming that smoking has a negative effect on wages, our estimates would be biased downward, away from 0 because of the omitted variable bias. This negative omitted variable bias comes from the fact that these summary statistics indicate that a married individual is less likely to smoke, but as discussed earlier, there may be a marriage premium that makes a married individual more likely to earn higher wages. To prevent my results from being biased I need to account for these observable characteristics. By conditioning on these observable characteristics, I can also ensure that the conditional independence assumption holds and potentially make a casual interpretation of the data.

3.2 Sibling Pair Descriptive Statistics

Summary statistics for the differences in characteristics between sibling pairs are shown in Table 2. Using the sibling pairs that have the same smoking status as a reference, Table 2 compares the differences between sibling pairs in which only the younger sibling or only the older sibling smokes to this reference group. From Table 2 we can see that in both years, for siblings with the same smoking status, on average the older sibling earned a higher wage than the younger sibling. This is what we would expect to see, because the older sibling likely has more job experience that translates into higher wages. Wage differences between sibling pairs where the younger sibling smokes and sibling pairs with no difference in smoking status is not statistically significant from 0. However, in 2011 if the older sibling smoked and the younger sibling did not, then the older sibling earned a much lower wage on average compared to the younger sibling. As just discussed, older siblings should typically earn more on average than their younger siblings due to having more job experience. In this case, the smoking older sibling on average earned less than their non-smoking younger sibling. This is the opposite of what we would expect when considering wage differentials between older and younger siblings. On the surface, this could be evidence toward the fact that smoking may be a factor that causes the wages of the smoking older sibling to be lower. Deeper analysis can provide further evidence toward or against this notion, but the fact remains that with smoking present, there is a negative smoking differential between the siblings where we would expect a positive differential without smoking.

Looking at the other observable differences, we can see that these two groups of sibling pairs in 2011 were not functionally different from one another except in terms of marriage. Sibling pairs where the older sibling smoked were likely to have an older sibling that wasn't married while the younger sibling was. The marriage premium could potentially cause the younger sibling to have a higher wage even without smoking being a factor, and so this provides evidence that we need to condition on observable characteristics in order to control for these characteristics and prevent them from biasing the estimates.

4. Research Design

As mentioned earlier, the goal of this paper is twofold: to provide an unbiased estimate of the effect of smoking on wages, and to generalize the results of the Levine et al. (1997) paper by applying their research methods to a more modern context. To further this joint goal, I follow the research method outlined by Levine et al. (1997). In order to obtain an unbiased estimate of the effect of smoking, I need to partial out the observed and unobserved effects that have the potential to be captured in our smoking status measure. This is particularly challenging for this paper, since there are many unobservable factors that are correlated with labor market behavior and smoking. In the previous section, I discuss an example of failing to include observable characteristics that are correlated with both wages and smoking behavior. An example of an unobservable factor that can affect a person's wages and smoking status would be anxiety. The side effects of anxiety are certainly observable and can lead to anxiety being diagnosed, but the condition itself is an unobservable cognitive process. A person with a lot of anxiety might pass up a promotion or might be too afraid to put themselves out there to obtain a higher-level position, which can potentially cause them to have lower wages. A common misconception of smoking is that it helps with anxiety. This is due to the fact that nicotine provides a short-term sense of relaxation, which can be improperly linked with anxiety relief. Therefore a person with a lot of anxiety who is ill-informed may be more likely to start smoking in an attempt to reduce their anxiety. As a result, not controlling for anxiety would bias our estimate downward and overstate the negative effect of smoking.

One method to attempt to prevent omitted variable bias would be to run an OLS regression that controls for various observable characteristics using a standard human capital equation in the form:

$$(1) : \ln Wage_i = \beta_0 + \beta_1 Smoker_i + \beta_2 X_i + \beta_3 F_i + e_i$$

where $Wage_i$ is an individual's hourly wage, $Smoker_i$ is an indicator that equals 1 if the individual is a smoker and 0 if they are not, $X_i$is a vector of individual characteristics, and F_i is a vector of family background characteristics. One shortcoming of the NLSY97 is the fact that it does not contain an abundance of family background information. Within the small pool of variables present, many of them were highly subjective and did not seem to provide much value (i.e. "On a scale of 1 to 24, how supportive is your mother of your father's decisions?"). Unfortunately, more helpful variables such as "Did the parents smoke?" are not present, and so it's important to note that the family background vector is not as well developed as it could be, containing only information related to the education of both parents. The fact that the family background vector is not as fleshed out as it could be hurts the conditional independence assumption, preventing the results from firmly being interpreted as the causal effect of smoking. A more robust family background vector would allow me to confidently interpret the results as the causal effect, but since the conditional independence assumption is not as strong as I would like it to be, I abstain from causal interpretations in the remainder of the paper.

It is likely that a person's individual characteristics and their family upbringing can play a factor in labor market behavior and smoking behavior (i.e. Your parents are more educated, which means they place emphasis on your own education, which allows you to get a higher paying job. At the same time, their higher education might have provided them with full knowledge of the dangers of smoking, which they could impart on you and make it less likely that you smoke.), so both need to be controlled for so that I can obtain an unbiased estimate. Equation (1), while accounting for observable characteristics, fails to account for the unobservable characteristics, and if these unobservable factors are not controlled for, then these standard OLS estimates will be biased.

Equation (1) can be modified in two different ways to account for these unobserved effects. The first method utilizes an individual first differences model. The nature of panel data allows us to estimate the relationship between changes in wages and changes in smoking behavior over time. Adding a time element to Equation (1) produces:

$$(2) : \ln Wage_{it} = \beta_0 + \beta_1 Smoker_{it} + \beta_2 X_{it} + \beta_3 F_i + \gamma_i + e_{it}$$

where all variables (except the vector of family characteristics since my family background variables do not change over time) now measure characteristics of an individual at a certain time *t*. Now the equation contains a fixed effect component, γ_i. This is specific to each individual and contains the variation in wages that would be explained by unobserved individual factors. Due to my panel data having multiple time periods, I can take the difference between two time periods to eliminate this individual fixed effect and therefore control for the heterogeneity caused by unobservable differences between individuals. Taking the difference between time periods *t* and *t-j* produces:

$$(3) : \ln Wage_{it} - \ln Wage_{it-j} = \beta_0 + \beta_1 Smoker_{it} + \beta_2 X_{it} + \beta_3 F_i + \gamma_i + e_{it}$$

$$-\beta_4 - \beta_1 S_{mokerit-j} - \beta_2 X_{it-j} - \beta_3 F_i - \gamma_i - e_{it-j}$$

and combining like terms provides us with a first differences equation that we can run through a standard OLS regression like so:

$$(4) : \Delta \ln Wage_i = (\beta_0 - \beta_4) + \beta_1 \Delta Smoker_i + \beta_2 \Delta X_i + \Delta e_i$$

It’s possible to see that the unobservable time invariant individual effects that influence a person’s wages have been eliminated by taking this difference, allowing β_1 to provide us with an estimate that isn’t biased by individual fixed effects.

With first differences and fixed effects models, a common concern is that switching over to the within variation can kill the variation that produces meaningful results. By using individuals as the controls to partial out fixed unobservable effects, there can be a concern that there isn’t enough variation within the individuals to measure an effect of our variable on interest. Around 15% of the individuals in the sample changed smoking status between 2006 and 2011, with almost twice as many individuals quitting smoking as starting smoker. This indicates that there is variation within the individuals to produce a meaningful result. In addition, the elimination of variation would produce high standard errors that would take credibility away from the results. For this paper, as will be evident in the results section, the standard errors for the individual first differences model are not incredibly high, further indicating that there is still meaningful variation within these individuals, which should alleviate concerns in using a first differences model. As mentioned previously, switching over to within variation has the capability to exacerbate measurement error, so when examining the results, it is important to see how much the coefficient estimates change when moving from OLS to first differences to determine the magnitude of measurement error.

As an alternative method to the individual first differences method just outlined, I can use the fact that the NLSY97 contains groups of siblings that grew up in the same household. Assuming that siblings that were raised in the same household have the same family background characteristics, I can modify equation (2) to include the sibling information, and then use a modified equation (3) to take the difference between two siblings from the same family, which would eliminate the unobservable effects that do not change over time. Modifying equation (2) produces:

$$(5) : \ln Wage_{st} = \beta_0 + \beta_1 Smoker_{st} + \beta_2 X_{st} + \beta_3 F_s + \varphi_s + e_{st}$$

All variables are now indexed by *s* which represents whether or not the individual is the older sibling (*o*), or the younger sibling (*y*). φ_s represents unobservable family background characteristics that do not vary over time. We can take the difference between an older sibling and a younger sibling at time *t* to produce:

$$(6) : \ln Wage_{ot} - \ln Wage_{yt} = \beta_0 + \beta_1 Smoker_{ot} + \beta_2 X_{ot} + \beta_3 F_o + \varphi_o + e_{ot}$$
$$-\beta_4 - \beta_1 Smoker_{yt} - \beta_2 X_{yt} - \beta_3 F_y - \varphi_y - e_{yt}$$

With our assumption that observable and unobservable family background characteristics are the same for both siblings ($\beta_3 F_o = \beta_3 F_y$; $\varphi_o = \varphi_y$) we can combine terms to produce:

$$(7) : \ln Wage_{ot} - \ln Wage_{yt} = (\beta_0 - \beta_4) + \beta_1 (Smoker_{ot} - Smoker_{yt})$$
$$+\beta_2 (X_{ot} - X_{yt}) + (e_{ot} - e_{yt})$$

The assumption that family background is the same ($\beta_3 F_o = \beta_3 F_y$ and $\varphi_o = \varphi_y$) for both siblings can be a fairly strong one, depending on how willing you are to believe that family effects are the same for both siblings. Parents tend to tell their children that they love them all equally and do not play favorites, but this might not always be the case. Some parents might treat one child differently from another. For example, if two siblings have strict parents, the assumption here would state that both parents treat each of their children exactly the same with an equal level of strictness. For the purposes of this paper, I am assuming that parents do not play favorites and therefore have the same attitude toward each sibling and give each one equal treatment.

With the removal of the unobservable family background characteristics, our estimate of β_1 will not be biased by family fixed effects. It is important to note that these methods do not entirely eliminate all sources of bias that can potentially affect the results. Regarding the individual first differences method, it is possible that changes in an individual's smoking status over time can be influenced by certain changes in that individual's life. The individual first differences method can't account for these factors, which means that a source of bias can still potentially exist. Regarding the sibling differences method, individual differences between two siblings can influence their smoking behavior as well as their wages. This method cannot account for the individual differences, which means that a source of bias can still potentially exist in this model as well.

A combination of the two methods can account for both individual and family fixed effects in addition to effects that can vary over time. By controlling for time invariant and time variant effects, I can move closer to obtaining an accurate estimate of the effect of smoking on wages. I can utilize equations (4) and (7) in order to first take the individual difference between each sibling between time *t* and time *t-j*, and then take a difference of the differences of both siblings. In this method, differences in wage growth between siblings is modeled as a function of differences in smoking status between siblings. To formalize this, consider equation (2) with an added unobservable time-varying element:

$$(8) : \ln Wage_{st} = \beta_0 + \beta_1 Smoker_{st} + \beta_2 X_{st} + \beta_3 F_i + \rho_{st} + e_{st}$$

Where ρ_{it} represents the unobservable characteristics of an individual that change over time. Taking the difference for each individual sibling across both time periods, and then taking a difference between siblings produces:

$$(9) : \Delta \ln Wage_o - \Delta \ln Wage_y = (\beta_0 - \beta_4) + \beta_1(\Delta Smoker_o - \Delta Smoker_y)$$
$$+ \beta_2(X_o - X_y) + (\Delta e_o - \Delta e_y)$$

The key assumption for this method is that the change in the unobservable time-varying characteristics is equal for both siblings ($\Delta\rho_o = \Delta\rho_y$). This can be a particularly strong assumption because it assumes that each sibling will experience the same type of life changes that can affect either their smoking status or wages. For example, under this assumption we would assume that all children in the same family experience the same change in maturity over time. In reality it could be the case that the younger child is spoiled a lot and remains immature, while the older child is not spoiled and becomes more mature. For this example, I assume that each child is treated equally and therefore experience the same changes in maturity. For the purposes of this paper I am assuming that equal family background observable and unobservable characteristics will cause each sibling to have the same kind of time-varying unobservable characteristic changes.

From equation (9) it's possible to see that I have removed both fixed effects as well as individual effects that vary over time. By removing these unobservable effects, I can obtain an unbiased estimate of the effects of smoking on wages. As mentioned earlier, there is a concern when employing these differences methods that we are eliminating variation that produces meaningful results. As will be evident in the results section, the large standard errors from estimating equation (7) and equation (9) are due to the small sample size that comes from the requirement that both siblings have to be full-time and full-year workers in both years. This small sample size does not allow for significant variation in smoking status among the small group of siblings, and so the standard errors have increased a great amount compared to the standard error growth between the OLS and individual first differences methods.

Another method I can employ to estimate the effects of smoking on hourly wages would be to use an instrument variable that is correlated with smoking, but only affects wages through the pathway of smoking. A popular instrument variable in the literature is the excise tax level for cigarettes. The excise tax for cigarettes should only affect wages through the pathway of smoking, since only people that buy cigarettes to smoke will have their income affected by the purchase of cigarettes. Excise taxes also should be correlated with smoking, as raising the sales tax on cigarettes should make a person less likely to buy cigarettes assuming that cigarettes are a normal good. To illustrate this, I can formalize the IV estimation process into the structural, first stage, and reduced form equations:

$$(10) : \ln Wage_{it} = \beta_0 + \beta_1 Smoker_{it} + \beta_2 X_{it} + \beta_3 F_i + e_{it}$$

$$(11) : Smoker_{it} = \beta_4 + \beta_5 Cigtax_{it} + \beta_6 X_{it} + \beta_7 F_i + e_{it}$$

$$(12) : \ln Wage_{it} = \beta_8 + \beta_9 Cigtax_{it} + \beta_{10} X_{it} + \beta_{11} F_i + e_{it}$$

As you can see, by using the tax level of cigarettes as a proxy for smoking, I could obtain unbiased estimates of the effect of smoking on wages. However, the sales tax on cigarettes needs to pass both the relevancy and exclusion restriction for this to be an effective IV. Upon running the first stage regression on the NLSY97 data, I find that the sales tax did not have a statistically significant effect on smoking behavior. Levine et. al (1997) find similar results with the NLSY79 dataset, indicating that this particular survey may not be effective for IV estimation using excise tax as the IV. Yuda (2011) also finds that tax levels don't impact the smoking decision of a current smoker who is attempting to quit, so there are situations in which excise tax is a weak IV for smoking. Since the tax level seems to be a weak instrument for the NLSY data, I stick to the OLS and first differences approaches outlined above.

5. Results

5.1 Central Analysis

The results of the OLS regressions (equation 2) are reported in Table 3. In columns 1 and 5 we can see the wage differences between the non-smokers and smokers that were reported earlier in the summary statistics. In columns 2 and 6 we can see the effects of just including education into the regression. By simply controlling for education I have reduced the estimate for smoking by nearly 50% in both years, which highlights the importance of controlling for observable characteristics. Controlling for the observable factors that are associated with both smoking and wages push these estimates upwards closer to 0, which indicates that my OVB was negative, which is what I expected from the discussion earlier in the paper. Adding the full range of controls slightly increase the estimate further and the final OLS results show that smoking is associated with around a 6% to 8% statistically significant decrease in wages on average, which is reflected in columns 3 and 7. The 2006 estimate of a 6.2% decrease is statistically significant at the 10% level, while the 2011 estimate of a 7.75% decrease is statistically significant at the 1% level. These estimates coincide with the results of prior research. Levine et al. (1997) find between a 4% and 7% statistically significant decrease in wages on average associated with smoking, so the fact that I obtain very similar results here can potentially lend validation to their results. Columns 4 and 8 add in a quadratic term for experience, based on the fact that work experience has been shown to eventually have a negative return to wages after a certain point. This specification doesn’t change our coefficient estimates for smoking status, but it does provide a different way of thinking about the impact of work experience in the long term (in this sample, work experience eventually beings to be associated with a negative effect on average for wages).

In order to take the OLS analysis one step further, I added three interaction terms to study the impact of smoking on different age groups, different education levels, and different experience levels. The idea is that if the coefficient on the interaction terms is significant, then smoking can have a different effect at certain age levels, education levels, and years of work experience. Hauber (2003) studies data from the Current Population Survey and finds that the wage penalty associated with smoking increases with years of education. Smokers who did not finish high school earned on average 7% higher wages associated with smoking than non-smokers with the same education level, while smokers who finished college earned on average 7% lower wages associated with smoking than non-smokers with the same education level. From this they conclude that smoking can have a different impact on different education levels. I explore this possibility in addition to exploring whether or not smoking can have a different impact on different age groups and different work experience levels.

The results of this regression are shown in Table 4. These regressions are the same as the ones run in columns 3 and 7 of Table 3 with the addition of the three interaction terms. From the table, it’s possible to see that in the 2011 cross-section, both the effect of smoking and the combined effect of smoking and age is statistically significant and negative. This indicates that the negative wage effect associated with smoking gets worse as an individual gets older. For this cross-section, an individual on average experiences a negative wage effect associated with smoking after turning 30 years old, and the wage effect associated with smoking continues to decrease with each year. This could indicate that younger adults don’t experience as negative of a wage change associated with smoking as an older population.

This lines up with the notion that the negative health effects due to smoking only get worse with age and quitting before 30 makes it more likely that your health fully recovers. Theoretically it could be the case that older individuals have lowered productivity from the negative health effects and therefore earn lower wages associated with smoking compared to younger individuals who most likely haven’t developed these negative health effects that would decrease their productivity at work. Additional analysis could be done to examine health status in addition to wages and smoking to take this further, but the NLYS97 does not have this kind of information readily available. The only indication of health in the study involves a question where respondents were asked at age 29 if there were any health limitations on their productivity at work in the past month, and there were a vastly small amount of "yes" answers. This isn't an efficient measure of health, and so more health-centric surveys like the National Wellness Inventory would be better suited to pursue this kind of study.

As mentioned previously, OLS regressions suffer from potential bias as a result of failure to control for unobservable characteristics. By employing the individual first differences model (equation 4) I can control for the unobservable characteristics that are caused by individual differences among the participants of the NLSY97. This takes care of the potential unobservable effects on wages that are the result of time invariant differences between individuals. The results of this process are shown in Table 5. Column 2 adds a quadratic term for the experience variable, which slightly increases the point estimate for the coefficient of the smoker variable. As you can see from the table, changes in smoking status are associated with around an 8.5% decrease in hourly wages on average. Moving from OLS to first differences, it's possible to see that the coefficient estimates decreased slightly, potentially indicating that the measurement error is not solely affecting the coefficient estimates. This does not necessarily mean that measurement error does not exist. It may be that the elimination of OVB is producing a larger effect on the coefficient estimate than the attenuation bias resulting from measurement error. The coefficient estimates are still in line with prior research, which indicates that measurement error is not significantly altering the results.

This estimate is not statistically significant, which indicates that changes in smoking status might not have an impact on an individual's wages. The OLS results deal with variation among individuals, but this first difference regression deals with variation *within* individuals. While the significant OLS results indicate that smoking is associated on average with a drop in wages among the sample, the fact that I obtain a statistically insignificant result here may tell us that smoking might not have an impact *within* individuals who begin smoking, which overall can potentially provide evidence toward the fact that smoking does not have an impact on an individual's wages. However, it's important to note that with a standard error of around 5.5%, a 95% confidence interval would span from 2.28% all the way down to a -19.28% effect on wages. Even though my model has not found statistical significance, it cannot rule out economical significance due to the precision of my results. The coefficient point estimates here line up with prior studies that have shown a statistically significant negative wage effect associated with smoking. Levine et al. (1997) in their individual first difference regression finds around a 6% negative effect. My results are similar to Levine et al. (1997), but slightly noisier (which prevents statistical significance), and so this similarity can potentially lend more validation to their study.

An alternative method to individual first differences would be to use sibling differences. Using a regression with sibling differenced data (equation 7) can account for the bias that stems from unobservable differences in family characteristics (family fixed effect). These results are shown in column 1 of Table 6. The estimate is made using the pooled data from both cross-section years. Using pooled data produces more precise results than individual cross-section data, and due to the small sample sizes associated with the sibling data, as much precision is appreciated as possible. From the table, it's possible to see that a change in smoking status is associated on average with a 11.77% decrease in wages. This result is not statistically significant. This may provide more evidence to the fact that a changing smoking status does not have an impact on the wages for young adults, but just like before, the standard errors need to be taken into account. Due to a small sample size, the standard error is fairly large, which affects the precision of the results, and does not allow for a firm conclusion that there is no effect because an economically significant effect can still be present. This estimate is within the range of 4% to 11% that has been established in previous literature and so smoking may in fact be associated with a negative wage effect. Levine et al. (1997) finds around an 8% average negative wage effect using this method with standard errors around 4%. My standard errors are around 10%, indicating that Levine et al. (1997) had more variation within their sibling pairs, but my point estimate is only slightly higher than Levine et al. (1997). My results are similar, but with slightly more noise, which could lend some validation to their results.

As discussed previously, like the individual first difference regressions, the sibling difference regressions are subject to bias because they only account for unobservable effects that come from family background. In order to account for sources of bias from unobservable time varying and time invariant factors I can run a regression using equation 10 to simultaneously account for individual fixed effects, family fixed effects, and time varying effects. The results of this regression are in column 2 of Table 6. As you can see, an older sibling who starts smoking is associated with a 22.96% smaller increase in wages compared to a younger sibling whose smoking behavior did not change. Similar to the results from regressing the sibling differenced data, this estimate is also not statistically significant. It's important to note that the sibling data, due to the fact that both siblings have to be full time and full year workers in both years, is very small compared to our original sample. Levine et al. (1997) finds a 25.1% negative wage effect, which further expands on the similarities between our papers and indicates that this particular method is not completely effective due to the small sample sizes associated with the sibling pairs. In theory the method should work to eliminate many sources of bias, and so it would be interesting to see how the method holds up in the presence of many sibling pairs.

5.2 Robustness Checks

In order to determine the robustness of my results, I have broken the original sample into four subgroups: male only, female only, less educated only (years of education less than or equal to 12), and more educated only (years of education greater than 12). Since smoking has the potential to affect these groups differently, it's important to examine this effect within each individual subgroup in order to determine how the results of my original analysis hold up within each subgroup[5]. The same analysis done in tables 3, 5, and 6 are done on each of these subgroups. In order for our previous results to be robust, each subgroup should display similar results to the ones that were found in the main analysis.

From Table 7, it's possible to see that my results are fairly robust. The point estimates in each method are mostly similar to the results from the main analysis. For the OLS section, five of the eight results are within the range of 6% to 9% established by the main OLS results. For the three that don't fall within my range, two still fall within the 4% to 11% range established by prior literature. The standard errors prevent a conclusion that there is a statistically significant negative wage effect present for half of the subgroups, but due to the point estimates being so close to prior literature, the fact that there may be an economically significant effect can't be ruled out. This leads to the conclusion that the OLS results are robust. Likewise, three of the individual first differences results are only slightly lower than the main result of around an 8% negative effect and fall within the range established by prior literature. This leads to the conclusion that the individual first differences results are fairly robust. For the robustness checks that involve sibling data, the small sample sizes prevent strong results in some of the subgroups

[5] For example, Bauer (2006) finds that gender differences in smoking are attributed to differences in smoking behavior rather than differences in characteristics between genders. Since men smoke more than women, on average men experience greater effects.

but half of the point estimates for the sibling differences method are near the results in the main analysis as well as the range established by prior literature, so sibling differences are fairly robust. For the combined method involving individual and sibling differences, the sample sizes are extremely small and don't allow for meaningful results, so I cannot conclude that the final method is robust, but overall the rest of the main analysis is fairly robust.

6. Conclusions

Smoking carries a number of direct and indirect effects that all have the ability to affect an individual's wages at work. Some of these effects, such as health complications, only get worse with age, so it's important to examine how early these effects can take hold in order to get a better picture of the impact of smoking on wages as a whole. By running an analysis that uses individual and sibling first differences methods using cross-sections from a panel of young adults, I have found that there is no statistically significant impact on wages as a result of smoking for adults below the age of 30. This doesn't necessarily mean that smoking doesn't have an impact on wages though, as lack of precision in the results prevents me from ruling out economically significant effects. The fact that other studies have shown negative wage effects due to smoking in older populations lends evidence toward the notion that smoking does have a wage effect on younger populations. The coefficients for my OLS and differences regressions do line up with previous research, which could indicate that smoking does in fact have a negative wage effect on this sample, but precision presents any definite conclusions from being made.

Preliminary IV tests showed that in this sample, the excise tax on cigarettes did not affect an individual's smoking behavior in a statistically significant way. Some other researchers have also found excise tax to be a weak IV, which provides evidence toward the fact that taxation isn't a completely effective way to deter smoking. Instead, efforts to target younger adults, such as the Truth anti-smoking ad campaign which targets teenagers, should also be focused on, especially in foreign countries where there are not effective anti-smoking controls.

This study, while not producing firm results in which to draw solid conclusions, does bring up some important points regarding the research process as a whole. Similar to this study, Levine et al. (1997) experienced problems with precision when using the sibling-differenced first difference equations. The method is an appealing way to attempt to account for bias from family fixed effects due to the fact that it is easy to understand and implement. The downside of this method is that the small sample size makes it hard to draw valid conclusions. Since precision can be a problem in first differences and fixed effects models, other research methods could be implemented in order to obtain firmer conclusions and really determine whether there are economically significant effects of smoking on wages for young adults. This way, we can obtain a better idea of the impacts of smoking at younger ages and from there potentially determine whether or not smoking-related health consequences are actually the driving factor in declines in wages as a result of smoking. The OLS interaction regressions provided some notion that the effect of smoking can worsen with age, and this worsening effect may be linked to the worsening health effects of smoking as age increases. Further analysis on data involving wage information, smoking information, and health information would be a great step to continue the analysis of the effects of smoking cigarettes on wages.

References

Bauer, Thomas, Silja Göhlmann, and Mathias Sinning. "Gender Differences in Smoking Behaviour." CEPR Discussion Paper No. 5848, 2006.

Bondzie, Eric A. "Effect of Smoking and Other Economic Variables on Wages in the Euro Area." MPRA Discussion Paper No. 69230, 2016.

Bunn III, William B., Gregg M. Stave, Kristen E. Downs, Jose Ma. J. Alvir, and Riad Dirani. "Effect of Smoking Status on Productivity Loss." *Journal of Occupational and Environmental Medicine*, vol. 48, no. 10, Oct. 2006, pp. 1099–1108.

Centers for Disease Control and Prevention (CDC) "Cigarette Smoking – Attributable Morbidity --- United States, 2000." *Morbidity and Mortality Weekly Report*, vol. 52, no. 35, Sep. 2003, pp. 842-844.

Centers for Disease Control and Prevention (CDC) "Current Cigarette Smoking Among Adults --- United States, 2006." *Morbidity and Mortality Weekly Report*, vol. 56, no. 44, Nov. 2006, pp. 1157-1161.

Grafova, Irina B., and Frank P. Stafford. "The Wage Effects of Personal Smoking History." *Industrial and Labor Relations Review*, vol. 62, no. 3, Apr. 2009, pp. 381–393.

Halpern, Michael T, Richard Shikiar, Anne M. Rentz, and Zeba M. Khan. "Impact of Smoking Status on Workplace Absenteeism and Productivity." *Tobacco Control*, vol. 10, no. 3, 2001, pp. 233–238.

Haubner, Jill. "Education, Smoking, and Earnings." Miami University Honors Thesis, Spring 2003

Jamal, Ahmed, Elyse Phillips, Andrea S. Gentzke, David M. Homa, Stephen D. Babb, Brian A. King, and Linda J. Neff. "Current Cigarette Smoking Among Adults – United Sates, 2016." *Morbidity and Mortality Weekly Report*, vol. 67, no. 2, 19 Jan. 2018, pp. 53–59.

Jha, Prabhat, Victoria Landsman, and Robert N. Anderson. "21st-Century Hazards of Smoking and Benefits of Cessation in the United States." *The New England Journal of Medicine*, vol. 368, no. 5, 24 Jan. 2013, pp. 341–350.

Levine, Phillip, Tara A. Gustafson, and Ann D. Velenchik. "More Bad News for Smokers? The Effects of Cigarette Smoking on Labor Market Outcomes." *Industrial and Labor Relations Review*, vol. 50, no. 3, Apr. 1997, pp. 493–509.

Suárez-Bonel MP, Villaverde-Royo MV, Nerín I, Sanz-Andrés C, Mezquida-Arno J, Córdoba-García R. "Costes derivados del uso de servicios sanitarios y bajas laborales en pacientes fumadores: estudio en una comunidad urbana. *Arch Bronconeumol*. vol. 51, no. 12, 2015, pp. 615-620.

Substance Abuse and Mental Health Services Administration. "Results from the 2006 National Survey on Drug Use and Health: National Findings." Office of Applied Studies, NSDUH Series H-32, DHHS Publication No. SMA 07-4293, Substance Abuse and Mental Health Services Administration, 2007.

Substance Abuse and Mental Health Services Administration, "Results from the 2011 National Survey on Drug Use and Health: Summary of National Findings." NSDUH Series H-44, HHS Publication No. (SMA) 12-4713. Substance Abuse and Mental Health Services Administration, 2012.

U.S. Department of Health and Human Services (HHS). *A Report of the Surgeon General: How Tobacco Smoke Causes Disease: What It Means to You.* U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2010, pp. 670-676.

U.S. Department of Health and Human Services (HHS). *The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General.* Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Jan. 2014.

Wilson, Valerie and William M. Rodgers III. "Black-White Wage Gaps Expand with Rising Wage Inequality." Economic Policy Institute, Sep. 2016.

Xu, Xin, Ellen E. Bishop, Sara M. Kennedy, Sean A. Simpson, and Terry F. Pechacek. "Annual Healthcare Spending Attributable to Cigarette Smoking." *American Journal of Preventative Medicine*, vol. 48, no. 3, 2015, pp. 326-333

Yuda, Michio. "The Effect of Habitual Smoking on Labour Productivity." *Applied Economics Letters*, vol. 18, no. 12, 2011, pp. 1125–1132.

Appendix

Table 1: Mean Characteristics of the 2006 and 2011 Cross-Section Samples, by Smoking Status

Characteristic	2006 Non-Smoker	2006 Smoker	2011 Non-Smoker	2011 Smoker
Wage	9.93***	10.7	14.19***	17.05
Log Wage	2.275***	2.153	2.718***	2.537
Years of Education	13.93***	12.783	14.75***	13.279
Years of Work Experience	6.573*	6.423	10.633	10.545
Age	24.164	24.114	28.988	29.003
% Female	49.041***	40.858	45.745**	41.347
% White	66.307***	77.73	68.182***	77.236
% Living in Urban Area	79.161	77.349	77.468	77.23
% Married	29.352***	21.268	46.124***	33.023
Number of Children	0.353	0.399	0.715	0.714
Household Size	2.896	2.966	2.932**	3.058
% Have Health Insurance	80.024***	65.884	81.946***	72.59
Cigarettes Smoked per Day	0***	10.42	0***	9.257
Dad's Years of Education	9.871	9.562	10.282	9.904
Mom's Years of Education	12.271	12.099	12.504**	12.094
Sample Size	1668	979	2068	861

*Difference from the mean of Smokers is significant at 10% level
**Significant at 5% level
***Significant at 1% level

Table 2: Mean Differences of the Sibling Data in 2006 and 2011, by Sibling Smoking Differences[1]

	2006			2011		
Variable	Younger Sibling Smokes	No Difference in Smoking Status	Older Sibling Smokes	Younger Sibling Smokes	No Difference in Smoking Status	Older Sibling Smokes
Difference in (D.i.) Wage	2.11	2.83	0.75	0.65	3	-3.34**
Difference in Log Wage	0.054	0.35	0.127	0.038	0.209	–.513***
D.i. Years of Education	1.963***	0.566	–.64**	1.192	0.444	0.067
D.i. Years of Work Experience	1.724	1.524	0.723*	1.432	1.591	1.675
D.i. Age	2.148	2.142	1.846	1.935	2.158	2.125
D.i. % Female	37.037**	2.655	3.846	25.806	3.797	–9.375
D.i. % Living in Urban Area	15.385*	–.943	–15.385	6.452	–.641	0
D.i. % Married	44.444**	15.044	15.385	19.355	15.924	–3.125*
D.i. Number of Children	0.037	0.159	0.423	0.645	0.411	0.156
D.i. Household Size	–.444	–.018	0.231	0.129	0.323	0.438
D.i. % Have Health Insurance	7.407	8.85	–19.231**	6.452	6.962	–3.125
Sample Size	27	113	26	31	158	32

[1]Difference in terms of older sibling - younger sibling

*Mean difference between siblings is statistically different at 10% level from the mean difference between siblings with no change in smoking status

**Significant at 5% level

***Significant at 1% level

Table 3: OLS Regression Results, by Cross-Section Year[1] (equation 1)

	2006				2011			
Dependent Variable	1	2	3	4	5	6	7	8
Smoker	−.1213*** *(.0333)*	−.0810** *(.0333)*	−.0620* *(.0329)*	−.0621* *(.0329)*	−.1816*** *(.0282)*	−.0918*** *(.0298)*	−.0775*** *(.0295)*	−.0775*** *(.0294)*
Years of Education		.0368*** *(.0070)*	.0075 *(.0076)*	.0066* *(.0076)*		.0520*** *(.0050)*	.0346*** *(.0052)*	.0343*** *(.0052)*
Years of Work Experience			.0795*** *(.0109)*	.1870*** *(.0614)*			.0451*** *(.0064)*	.1095*** *(.0397)*
Age			.0537*** *(.0130)*	.0563*** *(.0130)*			−.0011 *(.0110)*	.0020 *(.0110)*
Female			−.2035*** *(.0317)*	−.2042*** *(.0316)*			−.1623*** *(.0253)*	−.1640*** *(.0254)*
White			.0132 *(.0388)*	.0128 *(.0388)*			.0267 *(.0301)*	.0279 *(.0301)*
Lives in Urban Area			.0097 *(.0371)*	.0033 *(.0374)*			.0755** *(.0357)*	.0737** *(.0355)*
Married			.0129 *(.0354)*	.0127 *(.0354)*			.1358*** *(.0286)*	.1359*** *(.0286)*
Number of Children			−.0244 *(.0295)*	−.0251 *(.0295)*			.0126 *(.0180)*	.0130 *(.0178)*
Household Size			−.0341*** *(.0130)*	−.0341*** *(.0130)*			−.0347*** *(.0123)*	−.0349*** *(.0123)*
Has Health Insurance			.2587*** *(.0424)*	.2576*** *(.0422)*			.3658*** *(.0441)*	.3670*** *(.0441)*
Dad's Years of Education			.0040 *(.0027)*	.0042 *(.0027)*			.0059*** *(.0023)*	.0059*** *(.0023)*
Mom's Years of Education			.0058 *(.0041)*	.0060 *(.0041)*			.0043 *(.0034)*	.0043 *(.0034)*
Years of Work Experience Squared				−.0084* *(.0045)*				−.0032* *(.0019)*
Constant	2.2745*** *(.0181)*	1.7603*** *(.0989)*	.1734 *(.2955)*	−.1882 *(.3500)*	2.7182*** *(.0156)*	1.937*** *(.0778)*	1.345*** *(.2988)*	.957** *(.3850)*
	N = 2676 R^2 = .0044	N = 2676 R^2 = .0220	N = 2427 R^2 = .1586	N = 2427 R^2 = .1612	N = 3310 R^2 = .0106	N = 3088 R^2 = .0691	N = 2894 R^2 = .1968	N = 2894 R^2 = .1981

[1]Std. errors in parentheses
N = sample size
* Significant at 10% level
**Significant at 5% level
***Significant at 1% level

Table 4: OLS Interaction Term Results[1] (equation 1)

	2006	2011
	1	2
Dependent Variable		
Smoker	.9204 *(.5793)*	1.0803* *(.5725)*
Smoker * Education	.0101 *(.0140)*	-.0119 *(.0096)*
Smoker * Experience	.0099 *(.0218)*	.0085 *(.0128)*
Smoker * Age	–.0488* *(.0263)*	–.0370* *(.0213)*
Individual Variables	Yes	Yes
Family Background Variables	Yes	Yes
Constant	-.0772 *(.3269)*	.8022*** *(.2934)*
	N = 2427 R^2 = .1602	N = 2894 R^2 = .1983

[1]Std. errors in parentheses
N = sample size
* Significant at 10% level
**Significant at 5% level
***Significant at 1% level

Table 5: Individual First Difference Results[1] (equation 4)

Dependent Variable	1	2
Difference in (D.i.) Smoker	–.0848 *(.0547)*	–.0822 *(.0536)*
D.i. Years of Education	.0094 *(.0105)*	.0050 *(.0107)*
D.i. Years of Work Experience	.2475*** *(.0674)*	.4897*** *(.0794)*
D.i. Lives in Urban Area	.0942* *(.0500)*	.0975** *(.0493)*
D.i. in Marital Status	.0959* *(.0509)*	.0812* *(.0483)*
D.i. in Number of Children	.0232 *(.0326)*	.0339 *(.0318)*
D.i. Household Size	–.0413** *(.0162)*	–.0358** *(.0161)*
D.i. Has Health Insurance	.0113 *(.0532)*	.0092 *(.0536)*
D.i. Years of Work Experience Squared		–.0089*** *(.0013)*
Constant	–.7606** *(.3301)*	–1.1287*** *(.3396)*
	N = 1400 R^2 = .3201	N = 1400 R^2 = .3515

[1]Std. errors in parentheses

N = sample size

* Significant at 10% level

**Significant at 5% level

***Significant at 1% level

Table 6: Sibling Differences Results[12]

Dependent Variable	Sibling Differences Method (equation 7)	Individual + Sibling Differences Method (equation 9)
	1	2
Difference in (D.i.) Smoker	–.1177 *(.1002)*	–.2296 *(.1844)*
D.i. Years of Education	.0155 *(.0967)*	.0207 *(.0312)*
D.i. Years of Work Experience	.0967*** *(.0196)*	.0580 *(.1592)*
D.i. Lives in Urban Area	.0150 *(.100)*	.0589 *(.2140)*
D.i. in Marital Status	.2179** *(.0839)*	.0215 *(.1512)*
D.i. in Number of Children	-.0062 *(.0574)*	.0830 *(.2122)*
D.i. Household Size	–.0259 *(.0458)*	.0294 *(.1220)*
D.i. Has Health Insurance	.3222*** *(.0958)*	.4226* *(.2124)*
D.i. in Gender	–.2279*** *(.0633)*	
Constant	–.0635 *(.0609)*	–.1663 *(.1646)*
	N = 336 R^2 = .1660	N = 56 R^2 = .1279

[1]Std. errors in parentheses

[2]Difference in terms of older sibling - younger sibling

N = sample size

* Significant at 10% level

**Significant at 5% level

***Significant at 1% level

Table 7: Robustness Results, by Subsample[1]

Subgroup	OLS With 2006 Cross-Section (equation 1)		OLS With 2011 Cross-Section (equation 1)		Individual FD Method (equation 4)		Sibling Differences With Pooled Data[2] (equation 7)		Indivual + Sibling Difference Method[2] (equation 9)	
	1		2		3		4		5	
Male	–.0395 (.0381)	N = 1337 R^2 = .1445	–.0250 *(.0353)*	N = 1548 R^2 = .1850	–.0649 *(.0515)*	N = 821 R^2 = .3280	–.0749 *(.1652)*	N = 117 R^2 = .1781	–.2930 *(.2682)*	N = 30 R^2 = .3276
Female	–.0842* (.0511)	N = 1090 R^2 = .1720	–.1121** *(.0395)*	N = 1346 R^2 = .2180	-.1391 *(.0515)*	N = 579 R^2 = .3210	–.4869* *(.2578)*	N = 64 R^2 = .2855	X	N = 9
Higher Education (Years Education >12)	–.0740* (.0403)	N = 1052 R^2 = .1652	–.0581* *(.0343)*	N = 1927 R^2 = .1376	–.0737 *(.0782)*	N = 804 R^2 = .3786	.2361* *(.1391)*	N = 55 R^2 = .3833	X	N = 7
Lower Education (Years Education < or = 12)	–.0593 (.0480)	N = 1375 R^2 = .1616	–.05907 *(.0417)*	N = 967 R^2 = .2230	-.0621 *(.0782)*	N = 453 R^2 = .2126	–.1578 *(.1793)*	N = 154 R^2 = .1029	-.4254 (.6874)	N = 18 R^2 = .4716

[1]Std. errors in parentheses
[2]Difference in terms of older sibling - younger sibling
N = sample size
* Significant at 10% level
**Significant at 5% level
***Significant at 1% level
X : not enough observations to obtain
meaningful regression results