THE DYNAMICS OF THE SMOKING WAGE PENALTY

Michael E. Darden
Julie L. Hotchkiss
M. Melinda Pitts

The Dynamics of the Smoking Wage Penalty
Michael E. Darden, Julie L. Hotchkiss, and M. Melinda Pitts
NBER Working Paper No. 27567
July 2020
JEL No. I10,I12

ABSTRACT

Cigarette smokers earn significantly less than nonsmokers, but the magnitude of the smoking wage gap and the pathways by which it originates are unclear. Proposed mechanisms often focus on spot differences in employee productivity or employer preferences, neglecting the dynamic nature of human capital development and addiction. In this paper, we formulate a dynamic model of young workers as they transition from schooling to the labor market, a period in which the lifetime trajectory of wages is being developed. We estimate the model with data from the National Longitudinal Survey of Youth, 1997 Cohort, and we simulate the model under counterfactual scenarios that isolate the contemporaneous effects of smoking from dynamic differences in human capital accumulation and occupational selection. Results from our preferred model, which accounts for unobserved heterogeneity in the joint determination of smoking, human capital, labor supply, and wages, suggest that continued heavy smoking in young adulthood results in a wage penalty at age 30 of 14.8% and 9.3% for women and men, respectively. These differences are less than half of the raw difference in means in wages at age 30. We show that the contemporaneous effect of heavy smoking net of any life-cycle effects explains roughly 67% of the female smoking wage gap but only 11% of the male smoking wage gap.

Michael E. Darden
Carey School of Business Johns
Hopkins University
100 International Dr. Baltimore,
MD 21202
and NBER
mdarden4@jhu.edu

Julie L. Hotchkiss
Federal Reserve Bank of Atlanta
Research Department
1000 Peachtree St. NE Atlanta,
GA 30309-4470
and Georgia State University
Julie.L.Hotchkiss@atl.frb.org

M. Melinda Pitts
Federal Reserve Bank of Altanta
1000 Peachtree Street NE
Atlanta, GA 30309-4470
melinda.pitts@atl.frb.org

A data appendix is available at http://www.nber.org/data-appendix/w27567

1 Introduction

Smoking cigarettes is associated with well-documented expected longevity (Doll et al., 2004; Darden et al., 2018) and health care expenditure (Sloan et al., 2006; Xu et al., 2015) effects. In addition, smokers earn between 2% and 24% lower wages than nonsmokers (Auld, 2005; Grafova & Stafford, 2009; van Ours, 2004), but neither the magnitude nor the mechanisms behind the smoking wage gap are well understood. A simple set of explanations is that smokers are less productive due to higher rates of absenteeism and illness, and they generate higher health care costs, which may be particularly important to firms that self-insure.[1] Similarly, employers may have preferences (increasingly over time) for nonsmokers, and the smoking wage gap may represent a form of discrimination in which labor market opportunities are less forthcoming to smokers (Hotchkiss & Pitts, July 2013). While the literature has focused on these types of static mechanisms, differences in wages are inherently a dynamic process: smoking behavior and human capital formation begin at young ages, and these endogenous states influence decisions regarding education, labor supply, smoking, and occupation. Failing to account for the joint determination of these histories will lead to biased estimates of the smoking wage gap. Furthermore, there is little empirical evidence on the importance of these life-cycle mechanisms relative to static explanations, especially in early-career workers for whom human capital is still developing.

We estimate a dynamic model of human capital accumulation, wage determination, and smoking behavior to decompose static and life-cycle mechanisms behind the smoking wage gap. In our model, state variables that capture past smoking behavior, work decisions, and human capital accumulation influence the joint decisions to work, smoke, and seek education. Conditional on working, we observe wages and occupational task requirements, which we define as factors that characterize occupations on the basis of mental reasoning and physical strenuousness.[2] While task requirements shed light on productivity differences as a mechanism for the smoking wage gap, explicitly modeling the sequence of school enrollment decisions characterizes the extent to which those who choose to smoke accumulate lower human capital. An important feature of our dynamic empirical model is the treatment of unobserved heterogeneity, which we

---

[1]Berman et al. (2014) estimate the excess health care cost to be $5,816 per worker.

[2]These continuous occupational factors stem from principal component analysis and are highly correlated with the discrete occupational classifications used in the literature (Autor, 2019).

allow to be flexibly correlated across behavioral and outcome equations. If, for example, unobserved factors cause some individuals to be more likely to select into physically demanding jobs and to be more likely to start (and continue) smoking, we allow for this correlation. Finally, given that there are approximately 55 million *former* smokers in the United States, our model allows us to investigate the extent to which wages of former smokers converge to those of nonsmokers (Creamer et al., 2018).

We estimate our model with geocoded data from the National Longitudinal Survey of Youth, 1997 Cohort, which allows us to (a) evaluate competing mechanisms for the wage gap among a modern cohort of workers and (b) track the initial wage trajectories for smokers and nonsmokers, beginning with schooling and proceeding through a person's early career. In our raw data, the smoking wage gap between nonsmokers and near-daily smokers emerges by age 23 and grows to roughly 40% for women and 25% for men by age 30. To demonstrate the importance of accounting for observed and unobserved heterogeneity, we simulate our estimated model in which we impose different patterns of smoking behavior from age 16 through age 30 while updating state variables that capture past smoking, work, and education decisions. These simulations suggest that the selectivity-corrected age 30 smoking wage gap is 14.8% and 9.3% for women and men, respectively. The dramatic reduction between our simulated smoking results and those from the raw data is due to the strong positive correlation in factors that drive abstinence from smoking and human capital accumulation. These findings are broadly consistent with the literature (Auld, 2005; Grafova & Stafford, 2009; van Ours, 2004), but they stem from a model-based approach which is conducive to isolating mechanisms.

Our simulated results reflect the total marginal effect of smoking on wages – the combination of static and life-style mechanisms. To separate the static mechanisms of smoking on wages, we simulate our model, updating schooling, employment, and occupation decisions/outcomes as though an individual does not smoke while imposing different smoking behaviors in the determination of wages. Thus, wages progress as a function of nonsmoker employment, education, and occupation paths, but, in some simulations, as though individuals were smoking when wages are determined. We call this the contemporaneous effect of smoking on wages, and our simulation results imply that this contemporaneous effect explains 67% the total female smoking wage gap but only 11% of the male gap. That is, the majority of the smoking wage gap

for men stems from factors that are pre-determined when wages are set; for women, on the other hand, the contemporaneous effect explains the majority of the smoking wage gap. This finding - that mechanisms behind the smoking wage gap are very different for men versus women - extends the literature that the smoking wage gap differs by gender (van Ours, 2004).

Part of the difference in contemporaneous versus life-cycle effects by gender is due to occupation. For women at age 30, the smoking wage gap is increasing in the amount of mental reasoning required in an occupation - the smoking wage penalty is approximately 40% larger for women in the most mentally taxing occupations relative to the least - but invariant to its physical strenuousness. For men, the smoking wage gap is nearly 100% larger in the most mentally taxing jobs; however, the gap is *becomes smaller* in the degree of physical strenuousness. The smoking wage gap is roughly 36% smaller for men in the most physically demanding occupations relative to the least. This creates an incentive for male smokers to select into more physically strenuous occupations, where the penalty for smoking is lower. As educational attainment translates to more mentally rigorous and less physically strenuous occupations, the opportunity cost of an additional year of schooling for male smokers is higher than for female smokers. Our findings with respect to physical strenuousness also suggest that productivity and health differences, which should be exacerbated as an occupation becomes more physically demanding, are unlikely to drive the smoking wage gap for either men or women.

Both the magnitude and the mechanisms behind the smoking wage gap have significant policy relevance. Not only does smoking have implications for individual lifetime earnings, but the external implications for lower income tax revenues and productivity could be significant. Food and Drug Administration (FDA) cost-benefit and cost-effectiveness analyses already take these consequences into account (e.g., MacMonegle et al. (2018)), and our results may inform future regulatory action, including regulatory action with respect to substitute nicotine delivery products such as e-cigarettes (Cotti et al., 2020). Accounting for selection into smoking and individual heterogeneity, our analysis provides a more accurate estimate of the direct impact of smoking on wages, and we find that while the smoking wage gap is larger for women than for men, the returns to quitting are also more immediate and significant for women.

2   Background

The debate over regulating tobacco use in the U.S. dates back at least to 1964 when then U.S. Surgeon General at the time, Dr. Luther Terry, released the first government report concluding that cigarette smoking causes lung cancer, in addition to other types of cancer. The report also concludes, while there is no one characteristic that is uniquely identifiable with smokers, "The overwhelming evidence points to the conclusion that smoking—it's beginning, habituation, and occasional discontinuation—is to a large extent psychologically and socially determined." The report also makes a point to say that this conclusion "does not rule out...the existence of predisposing constitutional or hereditary factors."[3] The question, then, for public health officials, was whether smoking is a behavior that can be modified through social and psychological intervention, or whether it's just that smokers will be smokers and efforts to modify behavior will be ineffectual.

There is significant evidence that public health campaigns, smoking restrictions, and taxation can affect smoking behavior, indicating that the decision to smoke has more to do with psychological and social influences than hereditary factors. For example, Bauer et al. (2005) found that individuals in smoke-free work environments were 1.9 time more likely to stop smoking than those whose work places were not restricted (also see Fichtenberg & Glantz (2002)). Additionally, youth living in towns with strict restaurant smoking regulations were significantly less likely to progress from experimental to established smoking behavior than youth in locations with weak restaurant smoking regulations Siegel et al. (2009).

Regarding public health campaigns, "The Real Cost!" anti-smoking campaign was estimated to have prevented about 350,000 youth from starting smoking between 2014 and 2016 (MacMonegle et al. (2018)). However, of course, some campaigns are found to be more effective than others (for example, see Davis & Carpenter (2009)). The combined effectiveness of anti-smoking campaigns and regulations during the past half-century can be seen in the significant decline in the number of per capita smokers in the U.S. After rising steadily during the first half of the twentieth century, the rise in smokers per capita plateaued in the 1950s-1970s, then embarked on a dramatic decline through today (Centers for Disease Control and

---

[3]See page 40. HHS (1964)

Prevention, 2019; National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and, 2014).

Another effective deterrent to smoking, to varying degrees, has been found to be the imposition of excise taxes on cigarettes. While the general consensus is that imposing or raising taxes will reduce smoking (Chaloupka et al., 2012), a number of studies have found that a very large increase in taxes is required to reduce smoking by any meaningful amount (Callison & Kaestner, 2014). And while most evidence points to youth smoking behavior being more sensitive to taxes (i.e., price) (for example, see Carpenter & Cook (2008)), Hansen et al. (2017) found that the impact on taxes on youth smoking has all but disappeared more recently.

The health benefits of reduced smoking are clear. The CDC estimates that smoking related illnesses in the United States cost nearly \$300 billion each year (U.S. Department of Health and Human Services, 2014), and in 2010, nearly nine percent of healthcare expenditures in the U.S. were attributed to smoking cigarettes (Xu et al. (2015)). However, efforts to restrict, cajole, or incentivize people to not start or to quit smoking come at a cost, in terms of resources expended for enforcement, advertisement, and lost consumer surplus (Lemieux (2019)), not to mention the cost to the tobacco companies and tobacco farmers from reduced demand. By Presidential executive order, all government agencies are required, "in choosing among alternative regulatory approaches, those approaches that maximize net benefits (including potential economic, environmental, public health and safety, and other advantages; distributive impacts; and equity)" (White House, 2011, p. 3821). Policy makers have so far concluded that the benefits in terms of fewer deaths, fewer chronic smoking-related illnesses, and lower overall healthcare expenditures are worth the costs of regulating tobacco consumption. And, indeed, in the case of taxation, there is a bonus of increased revenue (Congressional Budget Office, 2012). Finally, it's clear that consideration of costs versus benefits of further regulation are explicit in FDA deliberations: "FDA also recognizes that, if FDA were to proceed to the stage of proposing a rule in this area, potential costs and benefits from a possible nicotine tobacco product standard would be estimated and considered in an accompanying preliminary impact analysis, including the potential impacts on growers of tobacco and current users of potentially regulated products." (Tobacco Product Standard for Nicotine Level of Combusted Cigarettes (Proposed Rules), 2018, p. 11820).

While pro-tobacco and libertarian influences have successfully convinced FDA to weigh the loss of consumer surplus associated with restricting access or making smoking costlier against the health benefits of further regulation, the economic benefit through higher wages, higher income tax revenue, and, potentially, higher productivity should also explicitly be used as motivation for regulation that reduces smoking initiation or induces quitting. This paper considers that direct benefit by providing a model-based estimate of the wage penalty smokers experience during crucial early years of human capital accumulation and work experience. Furthermore, we show how smoking cessation affects wages; we decompose our results by gender; and we investigate mechanisms that drive wage differences.

3   Evidence from NLSY

The National Longitudinal Survey of Youth, 1997 Cohort (NLSY97) offers a unique look at human capital accumulation and early-career wage progression among a nationally representative sample of the United States. The data contain rich, longitudinal information on wages, occupation, employment, education, and smoking for up to 18 years, beginning when respondents were between the ages of 12 and 18. Our data include geocoded information that, in some cases, is down to the Metropolitan Statistical Area (MSA) level, which allows us to characterize local labor market conditions and local tobacco control policies. In addition to local area unemployment rates that are included with NLSY data, we merge information on local economic conditions at the county level from Bartik et al. (2017). These data include shares of manufacturing and service occupations, as well as detailed labor market characteristics. Furthermore, we merge local area tobacco control measures, including state, county, and in some cases, MSA indoor smoking bans in restaurants, bars, and workplaces, and cigarette price levels and excise taxes from Americans for Nonsmokers' Rights.[4] Finally, because early career occupational selection may help to explain the smoking wage gap, we merge occupational task requirements from O*NET at the 3-digit occupational code level. Thus, our data neatly characterize the environment in which teenagers make initial decisions regarding smoking and human capital formation and in which young adults realize early career wage outcomes over up to 18 years. While the NLSY 1979 cohort offers a longer panel, smoking information is only sparsely

---

[4]Data are available at https://no-smoke.org/materials-services/lists-maps/

available, and the 1997 cohort offers a much more modern perspective on the smoking wage gap.

We restrict our attention to those individuals observed between 1997 and 2015 with no missing information on location, education, labor supply, and daily smoking.[5] Importantly, we do not drop individuals observed to attrit from the sample nor do we drop individuals who miss one or more NLSY waves; instead, we jointly model attrition with other endogenous variables, and we allow the error structure in the attrition equation to be correlated with all other equations. Table 1 provides precise details about how we construct our sample. The unit of time in our data is age, and we structure the data such that the first of up to 17 waves of information is at age 16.[6] To study wage differences between smokers, former smokers, and never smokers, we construct a panel of 5,529 individuals who generate 93,993 person/year observations and who facilitate the estimation of the dynamic model presented in Section 4. Because of the restrictions necessary to generate our estimation sample, we do not employ survey weights provided by NLSY. Instead, in all models in which we condition on covariates, we also condition on the survey weight. We split all of our empirical work by gender such that we have separate samples of 2,675 women and 2,854 men.

<div align="center">Table 1 Goes Here</div>

Figure 1 demonstrates the smoking wage gap. We define smoking based on the number of days in which a person reports smoking over the month preceding the NLSY interview. We define light and heavy smoking as smoking on fewer than 20 of the past 30 days and smoking on 20 or more of the last 30 days, respectively. By focusing on number of days smoked, rather than defining smoking based on the number of cigarettes per day, we are able to access two additional waves of the NLSY 1997 (2013 and 2015); our results are robust to alternative smoking definitions and classifications.[7] Log wages are defined as the natural logarithm of reported hourly cents per hour from an individual's main job. All dollar figures are converted to real, 2015 dollars using the Consumer Price Index. For both men and women, Figure 1 demonstrates a significant wage gap that emerges by age 23 and grows over time. By age 30, the gap between heavy smokers and

---

[5]We also impose that an individual must have been in school in 1997 as every state had compulsory education attendance through at least 16 years of age in 1997. https://nces.ed.gov/programs/digest/d97/d97t152.asp

[6]Because of the initial variation in age in 1997, we do not observe all individuals at all ages. Thus, we focus our descriptive and econometric results on ages 16 though 30.

[7]These results are available by request.

nonsmokers is 40% (25%) for women (men). Interestingly, the smoking wage gap for both genders is mainly a difference between heavy smokers (i.e., near-daily smokers) relative to nonsmokers. Indeed, for both men and women, intermittent smoking is not associated with a significant wage gap at age 30.

<div align="center">FIGURE 1 GOES HERE</div>

To investigate the trends in Figure 1, Table 2 presents summary statistics of all exogenous and pre-determined variables in our dynamic model by smoking status and gender. Over our sample, 17.6% of women and 22.5% of men report heavy smoking. Among the exogenous variables, there is a notable gap in Armed Services Vocational Aptitude Battery (ASVAB) scores between heavy and nonsmokers for men (37.3 vs. 41.1), but no such difference for women. For both men and women, Blacks and Hispanics are less likely to report heavy smoking. Table 2 demonstrates the richness of exogenous characteristics in our data. Employment share data are at the county level, as are the local area cigarette taxes. Furthermore, our data include a rich set of county and state level indoor smoking ban information for workplaces. We define these based on their geographic coverage (i.e., partial or complete within geographic area) and whether they are qualified or complete. Pre-determined variables are those variables determined by our model but which represent a fixed state when decisions are being made. These variables highlight the importance of selection. For example, heavy smokers have, on average more than one fewer years of schooling than never smokers and are much less likely to be college graduates relative to nonsmokers. Furthermore, heavy smokers have lower mean tenure (working within the same occupation) than never smokers.

<div align="center">TABLE 2 GOES HERE</div>

Because early career occupational selection may be an important driver of wage differences between smokers and nonsmokers, we merge O*Net data at the 2010 3-digit occupation code. These data include a plethora of measurements that characterize an occupation by the tasks required on a daily basis. We follow the literature and aggregate these measurements using Principal Component Analysis (PCA). Specifically, we construct two latent factors, and Table 3 shows the measurements that enter our PCA for each factor, as well

<div align="center">8</div>

as the weight associated with the first principal component for each factor. The first factor, which captures roughly 53% of the variation in measurements, loads strongly on communication skills, comprehension, and deductive and inductive reasoning. We name this factor "mental reasoning." The second factor, which explains roughly 63% of the variation in measurements, loads heavily on measurements that capture the physical exposure that workers place on their bodies. We name this factor "physical strenuousness." Each factor is normalized to be between 0 and 1, where 1 represents the most task complexity. Factors for mental reasoning and physical strenuousness are highly correlated with factors from Autor (2019) for non-routine interactive and routine manual occupations, respectively. We focus on latent factors generated by PCA because they provide continuous metrics for task intensity, which we use to investigate heterogeneity in the smoking wage gap.[8]

<div align="center">Table 3 Goes Here</div>

Table 4, which presents the means for the model's endogenous variables by smoking status, demonstrates potential sources of the smoking wage gap. Specifically, heavy smokers are significantly less likely to be enrolled in school and more likely to be working full-time. This point highlights the dynamics of the smoking wage gap - smokers fail to invest in human capital at young ages; their initial wages are higher because they are working full-time, but over time never smokers earn more. Conditional on working, heavy smokers work in occupations that require lower mental reasoning, a factor that is strongly associated with wages, and higher physical strenuousness. When averaged over the entire sample period, female heavy smokers earn approximately 21% lower wages than nonsmokers, and that gap for men is roughly 6%.

<div align="center">Table 4 Goes Here</div>

To demonstrate the dynamic importance of the endogenous variables in Table 4, Figure 2 presents the age profile of school enrollment and occupation variables by smoking status separately for women and men.

---

[8]We have estimated our model with a binary equation for nonroutine interactive occupations instead the two continuous latent factors, and our results are very similar. These results are available upon request.

Recall that our data are constructed such that all sample individuals are enrolled in school at age 16. By age 17, for both genders, heavy smokers drop school enrollment much faster than nonsmokers. By age 19, the proportion of heavy smokers enrolled in school is roughly 27 percentage points lower than non-smokers for both men and women. The figure also shows how mental reasoning and physical strenuousness of one's occupation evolve by smoking status. By age 23, a significant gap in the mental reasoning required in an occupation emerges between heavy smokers and nonsmokers. For both men and women, heavy smokers work in more physically strenuous occupations, a difference that is relatively constant over the ages we study. Figure 2 demonstrates the dynamic nature of the smoking wage gap. Smokers leave school earlier, accumulating less human capital; they enter the workforce earlier and work in occupations that are more physically strenuous and less mentally rigorous, all of which are associated with lower wages.

Given the descriptive evidence in Tables 2 and 4 and Figure 2, the wage trends in Figure 1 are difficult to interpret for several reasons. First, smoking behavior is not time invariant. Individuals start and stop smoking at different ages, and it remains an open question if those who quit smoking see convergence in wages. Second, the sample changes over the age profile through selective attrition. To maximize sample size, we include those leaving the sample prior to age 30, so the fact that the wage gap at age 30 is larger than at age 25 is a function of both the underlying wage dynamics and changes in sample composition due to attrition. Finally, wages at age 30 are a function of complex interactions between past smoking, education, and employment behaviors and occupational outcomes, as illustrated in Figure 2. The remainder of this paper develops a dynamic model of these outcomes and behaviors that helps to explain the mechanisms that drive the wage trends in Figure 1.

4   Dynamic Model

Recognizing the limitations of the statistics above, we propose a dynamic system of equations estimator that approximates a more general dynamic stochastic structural model.[9] The empirical model accounts for selection into both smoking and employment by jointly estimating behavioral equations for enrollment in school, employment, and smoking, as well as outcomes for the mental and physical strenuousness of

---

[9]Examples of similar estimators include Gilleskie <u>et al.</u> (2017) and Darden <u>et al.</u> (2018).

occupations and wages. To account for dynamic selection caused by non-random sample attrition, the model also captures the endogenous probability of leaving the sample. Finally, we allow initial conditions (age 16) for smoking and employment to shift the distribution of permanent unobserved heterogeneity that flexibly affects each behavior and outcome equation. In what follows, we outline (rather than solve) the dynamic structural model; we explain how the structural model informs the system of demand and outcome equations which we jointly estimate; and we discuss the exclusion restrictions and functional form assumptions which guarantee identification of the system.

4.1   Structural Equations

The structural model follows from standard labor supply theory in which an individual chooses whether and how much to work based on a wage/job task draw. Because our goal is to decompose the mechanisms that lead smokers to realize lower wages, we augment this model to allow for endogenous smoking and school enrollment decisions. Consistent with the data in Section 3, we start the model at age 16, when all individuals are assumed to be enrolled in school. The timeline picture below documents a representative age t. Individual i enters representative age t with vector of state variables Θ_t, which include past work history (both labor supply and occupation choices), human capital accumulation, and past smoking behavior.[10] Conditional on their state vector, an individual receives a draw from the joint distribution of wages and task complexity, which characterize the requirements of the occupation on both mental and physical dimensions. Conditional on this offer and on their state variables, they simultaneously decide whether to enroll in school (a_t), whether to work (e_t), and to what extent to smoke (s_t).[11] If they decide to work, wages (w_t) and occupational requirements, both mental (m_t) and physical (ph_t), are observed. At the end of the period, the individual decides whether or not to leave the sample through attrition (L_t) as a function of their state variables and their age t behaviors. As age transitions to $t+1$, the state vector Θ_{t+1} is updated given age t behavior and outcomes conditional on remaining in the sample.

---

[10]For ease of notation, we omit the individual i subscript where possible.
[11]Working does not preclude school attendance.

$$
\begin{array}{ccccc}
 & s_t & w_t & & \\
 & e_t & ph_t & & \\
\Theta_t & a_t & m_t & \Theta_{t+1} & \\
\end{array}
$$
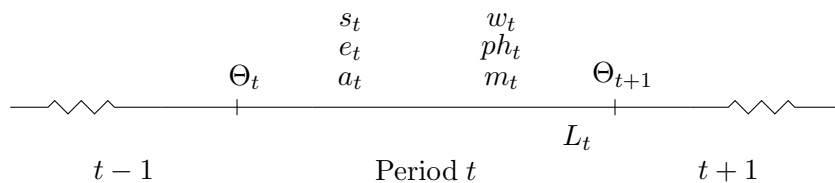
Implicit in the timing of the model is an important identification assumption. Theory suggests that

employment, education, and smoking should be chosen simultaneously as a function of the same variables.

Furthermore, because wages and tasks are only observed conditional on working, outcomes for wages and

tasks are also a function of the same state variables and exogenous characteristics as the behaviors. Put

differently, because the wage/task offer is made as a function of the state vector and exogenous character-

istics, behavior and outcomes are jointly determined. We discuss further identification assumptions (i.e.,

exclusion restrictions) below. Rather than estimate the fully structural, forward-looking model, we estimate

an approximation, which captures the dynamics of the system. Specifically, solution to the structural model

would yield demand equations for smoking, schooling, and employment:

$$p(s_t = s) = s(\Theta_t, X_t, P_t, \mu^s, \epsilon_t^s) \tag{1}$$

$$p(a_t = a) = a(\Theta_t, X_t, P_t, \mu^a, \epsilon_t^a) \tag{2}$$

$$p(e_t = e) = e(\Theta_t, X_t, P_t, \mu^e, \epsilon_t^e) \tag{3}$$

where, s takes values 0, 1, and 2, corresponding to none, light, and heavy smoking, respectively, and

Equation 2 captures the binary decision to attend school. Equation 3 is our model of employment, where e

takes values of 0, 1, and 2, representing none, part-time, and full-time work. The state vector is denoted Θ_t,

and X_t represents exogenous individual characteristics. P_t is a price vector that captures cigarette prices

and tobacco control policies, as well as local labor market conditions that may shift the decisions to seeking

schooling and/or work. The μ and ϵ terms represent the error structure of each equation, which we define

below.

Conditional on working, Equations 4 and 5 describe observations of mental and physical task complexity,

$$m_t = m(\Theta_t, X_t, P_t^e, \mu^m, \epsilon_t^m | e_t > 0) \tag{4}$$

$$ph_t = ph(\Theta_t, X_t, P_t^e, \mu^{ph}, \epsilon_t^{ph} | e_t > 0) \tag{5}$$

both of which are continuous, linear equations, where m_t and ph_t are defined between 0 and 1. Also conditional on working, Equation 6 describes the (log) wage conditional on the state vector:

$$w_t = w(\Theta_t, X_t, P_t^e, \mu^w, \epsilon_t^w | e_t > 0) \tag{6}$$

Aside the error structure, the only difference between behavioral Equations 1-3 and outcome Equations 4-6 is the the price vector P_t^e, which represents only a subset of the variables in P_t. We discuss these exclusions below. Finally, Equation 7 describes the probability of sample attrition:

$$p(L_t = l) = l(s_t, e_t, a_t, m_t, ph_t, w_t, \Theta_t, X_t, \mu^L, \epsilon_t^L) \tag{7}$$

Equations 1-7 represent the per-period system of dynamic equations that we estimate via full-information maximum likelihood. In what follows, we discuss the error structure of the model (i.e., μ and ϵ), the initial conditions, exclusion restrictions, and the form of the likelihood function.

4.2   Unobserved Heterogeneity, Initial Conditions, and Identification

We handle unobserved heterogeneity (e.g., unobserved variables that shift both the probability of smoking and the realization of wages) by decomposing the error structure of each equation into a permanent component (i.e., μ) and an i.i.d. component (i.e., ϵ). We take a semi-parametric approach in the sense that while the ϵ term is assumed to follow an extreme value type 1 distribution for all discrete outcomes and a normal distribution for all continuous outcomes, we make no distributional assumption on the joint distribution of permanent unobserved heterogeneity μ. Instead, we estimate a step-function of discrete support points that are allowed to vary by equation. Intuitively, our discrete approach implies a set of unobserved "types"

which jointly determine all outcomes and behaviors of the model.[12] We allow each instance of the perma-

nent disturbance $\mu^{\cdot}$ to take one of four points: $\mu \in \{\mu_1^{\cdot}, \ldots, \mu_4^{\cdot}\}$, and we estimate the μ parameters along

with the probability that an individual's behavior and outcomes are consistent with each of the four types

jointly with other parameters in the model.[13] For example, if a group of individuals have a particularly low

discount rate, which is unobserved in our data, we would expect to estimate one type where the value of

μ in the smoking equation was negative and the value of μ in the school enrollment equation was positive.

We select four points of support because a fifth does not improve the likelihood function value and in some

cases the estimation algorithm fails to differentiate types within some equations when considering a fifth

point.

Equations 1-7 are dynamic in the sense that each dependent variable depends on past behaviors and

outcomes; however, at age 16, our data do not include information on these lagged variables. We solve this

initial condition problem in two ways. First, we construct our data such that all individuals are enrolled

in school at age 16, thus eliminating variation in the school enrollment initial condition. Second, following

Keane & Wolpin (1997), we allow initial conditions for light (heavy) smoking and part-time (full-time) labor

supply to affect the permanent unobserved component μ. Specifically, we let the probability that individual

i exhibits behavior consistent with type k be given as:

$$\tau_k = ln(\frac{P(\mu_i = \mu^k)}{P(\mu_i = \mu^1)}) = \psi_0^k + \psi_1^k 1[s_{i,t=16} = 1] + \psi_2^k 1[s_{i,t=16} = 2] + \psi_3^k 1[e_{i,t=16} = 1] + \psi_4^k 1[e_{i,t=16} = 2] \quad (8)$$

Thus, age 16 smoking and labor market behavior affect the probability of being each type. If, for example,

type 2 individuals are more likely to smoke cigarettes between ages 17 and 30, then we would expect ψ^2

to be positive because individuals observed to be smoking heavily at age 16 should be more likely to be

observed smoking heavily at older ages.

At this point, the model is technically identified. By allowing the initial conditions to affect future

behavior and outcomes through μ, and by assuming that some equations exhibit non-linear functional

---

[12]See Heckman & Singer (1984); Mroz (1999); Darden <u>et al.</u> (2018) for further details. This approach is designed to capture between-individual permanent unobserved heterogeneity while still allowing for the estimation of coefficients on time-invariant individual specific variables.

[13]Estimation of these points is possible up to the normalization that the first point of support $\mu_1 = 0$ for all equations.

forms through assumptions place on ϵ, the likelihood function is maximized by a unique set of parameters. To estimate the system, we maximize the log-likelihood function with respect to the parameters that dictate the distribution of unobserved heterogeneity (i.e., $\mu^2 \ldots \mu^4$ for each equation plus the ψ parameters, which dictate the share of individuals consistent with each type) and per-period behavior and outcomes. The latent factor μ is integrated out of the likelihood function by constructing the μ-type specific contribution to the likelihood function and then weighting each type specific likelihood value by the probability that a given person's behavior is consistent with that type. The likelihood function value for period i is given as:

$$
\begin{aligned}
L_i(\Theta) \;\; = \;\; & \sum_{k=1}^{4} \tau_k \Bigg\{ \prod_{t=2}^{T(i)} \Big\{ \prod_{s=0}^{2} P(s_t = s|\mu_k)^{1[s_t=s]} \prod_{a=0}^{1} P(a_t = a|\mu_k)^{1[a_t=a]} \prod_{e=0}^{1} P(e_t = e|\mu_k)^{1[e_t=e]} \times \\
& \times \phi(w_t - (\hat{w}_t|e_t = 1, \mu_k))\phi(m_t - (\hat{m}_t|e_t = 1, \mu_k))\phi(ph_t - (\hat{ph}_t|e_t = 1, \mu_k)) \times \\
& \times \prod_{l=0}^{1} P(l_t = l|\mu_k)^{1[l_t=l]} \Big\} \Bigg\}
\end{aligned}
$$

$$(9)$$

$$(10)$$

We estimate the model separately for men and women, and we evaluate the estimated models via simulation.[14]

While uniqueness of the parameters is guaranteed by our functional form assumptions, our identification argument is also informed by economic theory via exclusion restrictions. As noted above, the timing of the model implies that schooling, employment, and smoking decisions are made simultaneously as a function of the same variables. Included in Equations 1-3 is a vector P_t, which includes local area cigarette prices, excise taxes, and the presence of local indoor smoking bans in restaurants, bars, and workplaces. P_t also includes local area labor demand variables such as the manufacturing share of employment, and the unemployment rate. Importantly, variables that dictate local labor market conditions are included in P_t^e, which shift equations for mental reasoning, physical strenuousness, and log wages, but smoking policy and price variables in P_t are *excluded* from P_t^e. Specifically, we decompose the price vector $P_t = \{P_t^s, P_t^e\}$

---

[14]The models are slightly different for women and men because we control for lagged pregnancy in the female model. Clearly pregnancy is endogenous in that fertility may be chosen jointly with smoking and employment decisions. We have estimated a model in which we treat fertility as endogenous and it does not change our results. These results are available upon request.

such that the vector of tobacco prices P_t^s is excluded from the outcome equations. The assumption is that cigarette prices, taxes, and regulations (e.g., indoor smoking bans) only affect wages through their effect on contemporaneous smoking behavior. Furthermore, the entire vector P_{t-1} is excluded from period t behavior and outcomes, which assumes that lagged prices, policies, and labor market conditions only affect period t behavior and outcomes through lagged behavior and outcomes. As with any exclusion restriction, the assumptions laid out here are untestable; however, elements of P_t^s are jointly predictive of smoking behavior.[15]

Finally, in addition to theoretically motivated exclusion restrictions, the dynamics of the system along with our treatment of unobserved heterogeneity handle a variety of competing explanations for patterns in the data. For example, similar to Farrell & Fuchs (1982), we find a significantly negative correlation between smoking behavior and years of schooling. One possibility is that there is something transformative about education that causes individuals to choose not to smoke. If so, our model captures this explanation because current period smoking behavior is a function of both the lagged school enrollment decision as well as the cumulative years of schooling up to the current period. Alternatively, the negative correlation between smoking and schooling may be the result of unobserved "third" variable(s) that drive both behaviors. Indeed, Farrell & Fuchs (1982) demonstrate that nearly all of the negative correlation between years of schooling and smoking at age 24 is explained by smoking behavior at age 17, and those authors argue that unobserved variables (e.g., rates of time preference) drive both smoking and schooling choices. Our model also allows for this possibility because we allow the permanent unobserved component (i.e., μ^a) of the school enrollment equation to be correlated with the permanent unobserved component (i.e., μ^s) of the smoking equation.

---

[15]The F-statistic on the test that all coefficients on cigarette prices, taxes, and regulations are zero in the smoking equation is 55.47 (p-value=0.000) for men and 43.45 (p-value=0.0173) for women.

5   Results

5.1   Parameter Estimates

Table 5 presents parameter estimates from the wage equation. Columns labeled as "separately" present coefficients and standard errors from estimation of the wage equation separately from other equations in the model, and columns labeled "jointly" present estimates from our preferred model in which all equations are estimated together with correlated errors. Hence, columns labeled jointly present estimates of $\mu_2^w \ldots \mu_4^w$. In our preferred system model, the parameter estimates themselves are difficult to interpret because of the complex dynamics. For example, smoking affects wages both directly, through direct effects of lagged smoking and its interaction with exogenous age and gender, but also through the stock of years of smoking and smoking cessation. Furthermore, interactions between lagged smoking and a quadratic of age allow for a flexible age profile by smoking status, but they make inferring marginal effects difficult. Thus, we wait to interpret the role of smoking until the simulations of the model are discussed.[16] All results that follow are from estimation of the joint dynamic system.

<div align="center">Table 5 Goes Here</div>

While interpretation of individual parameters within the system is difficult, the estimated parameters of the unobserved heterogeneity distribution demonstrate the importance of allowing for correlation in the errors across equations. Table 6 presents these parameters along with estimates of the ψ parameters, which relate initial conditions to unobserved type. For example, for women, light smoking at age 16 implies that type 3 behavioral and outcomes patterns are most likely ($\psi_1^3 = 0.561$). Relative to type 1 individuals, from ages 17-30, type 3 individuals are statistically more likely to smoke heavily, less likely to be enrolled in school, less likely to be working part-time or full-time, and yet they earn higher wages while working in less mentally rigorous and more physically strenuous occupations.

<div align="center">Table 6 Goes Here</div>

---

[16]Parameter estimates from other equations are presented in the Appendix.

Rather than rely on interpretation of the parameter estimates, we simulate the estimated model, using the exogenous characteristics of our sample while updating the endogenous variables. Formally, we expand the longitudinal records of the 2,675 women and 2,854 men by 50, and we endow each person with a complete set of error draws, including a draw from the unobserved "type" distribution based on their age 16 smoking and labor supply behavior. We simulate behavior and outcomes forward for each gender separately, taking care to update the state vector based on the simulated behaviors and outcomes. For example, if an individual is simulated to smoke in year t, we update the years of smoking state variable in $t + 1$ to reflect this behavior, regardless of if the person in question actually smoked in t. In the event that a sample individual leaves the sample through attrition but we do not simulate that person to leave, we assume that the exogenous characteristics stay fixed to the last actual observation.

<div align="center">

FIGURE 3 GOES HERE

</div>

Figure 3 demonstrates that our estimated models capture the age profiles of smoking and wages for both men and women. Here, we compare the simulated behaviors and outcomes among those simulated to remain in the sample and the observed behaviors and outcomes of those who have yet to leave through attrition. These simulations represent the baseline model. All behaviors and outcomes are determined by gender-specific estimated models, and they demonstrate that the combination of individual exogenous characteristics plus the estimated models can generate the same patterns of endogenous variables that we observed in our data.

5.2   Simulations

5.2.1   Smokers vs. Never Smokers

To evaluate the smoking wage gap, we impose different smoking patterns and simulate the corresponding behaviors and outcomes. That is, we conduct our simulations just as those in Figure 3, but rather than allow smoking behavior to evolve according to the model, we impose different smoking patterns. As a natural starting point, we simulate our model imposing smoking behavior at every age and in every equation.[17] This

---

[17]We do not impose smoking behavior on the initial conditions in any simulation, as such a change would shift the estimated distribution of unobserved heterogeneity. Instead, age 16 smoking behavior stays fixed, and it informs the probabilities of each type for each individual.

simulation captures the total marginal effect of smoking on wages in the sense that smoking affects wages both directly in the wage equation and indirectly through its effects on school enrollment and employment decisions and occupation outcomes. Figures 4a and 4b presents the resulting differences in simulated wages for those assumed to always smoke lightly (heavily) relative to those simulated to never smoke. In stark contrast to the descriptive evidence presented in Figure 1, the smoking wage gap for women (men) at age 30 is only 14.8% (9.3%) for heavy smoking. The considerable attenuation of these results reflects the importance of unobserved heterogeneity in the joint determination of smoking and labor market outcomes. Furthermore, the contrast between results in Figures 1 and 4(a and b) is only evident for heavy smoking - the results are consistent for light smoking.

<div align="center">Figure 4 Goes Here</div>

5.2.2   Smoking Cessation

Figures 4c and 4d explore how smoking cessation affects wages. Specifically, the figures present simulation results relative to a simulation in which no one smokes for heavy smokers who continue smoking (identical to those in Figure 4a and 4b) and heavy smokers who quit at age 24. For women, the effect of cessation is immediate: wages increase significantly in the year of cessation, which suggests that the smoking wage gap for women may depend on spot mechanisms when wages are determined. For men, wages increase only gradually, suggesting that life-cycle effects are more important.

5.2.3   Contemporaneous Impact of Smoking

Next, following Gilleskie et al. (2017), we isolate contemporaneous mechanisms that drive wage differences from mechanisms that originate prior to wage outcomes being determined. This is done by comparing simulated wages in an environment where no one ever smokes to simulated wages assuming everyone smokes only in the wage equation. In other words, the dynamic contribution of smoking in the determination of schooling, employment, occupation, and attrition is shut down. Thus, smoking is only allowed to enter the model contemporaneously in the determination of wages. Figure 5a and 5b present the differences in log

wages between these simulations for women and men, respectively. For women, the majority of the total

smoking wage gap (i.e., 14.8%) remains at age 30 - wages are roughly 10% lower for heavy smokers relative

to nonsmokers. This result implies that contemporaneous mechanisms - those mechanisms that occur at the

time wages are determined - explain 67% of the smoking wage gap for women. In contrast, contemporaneous

mechanisms explain only 11% of the male gap. By age 30, the penalty for heavy smoking among men is

only about 1% when only considering contemporaneous mechanisms. These results are consistent with

the smoking cessation results in Figures 4c and 4d, which demonstrate that upon quitting, female wages

increase much faster than do male wages.

<div align="center">Figure 5 Goes Here</div>

5.2.4 Smoking Wage Gap and Productivity

Finally, we explore heterogeneity in both the total marginal effect and the contemporaneous effect by the

degree of task intensity within occupations. If, for example, the contemporaneous effect of smoking on

wages increases (i.e., the wage gap increases) with the degree of physical strenuousness on the job, then

differences in productivity conditional on human capital may explain the smoking wage gap. Alternatively,

if only the total effect of smoking on wages increases in the degree of physical strenuousness, then smoking

may affect the growth in occupational task complexity, which drives current wage differences.

<div align="center">Figure 6 Goes Here</div>

Figure 6 presents estimates of the smoking wage gap at age 30 at each of 20 bins of the distributions of

mental reasoning and physical strenuousness. In each figure, we present both the total effect (dashed line)

of smoking and the contemporaneous effect (solid line). For women, the smoking wage gap increases as

occupations become more mentally taxing (panel a), but the gap is constant across different levels of physical

strenuousness (panel c). For men, similar to women, the total wage gap grows strongly in the degree of

mental reasoning. Interestingly, however, the total effect of smoking on wages actually shrinks for men

in the degree of physical strenuousness. Both the flat wage gap for women and *decreasing* wage gap for men as physical strenuous increases is inconsistent with a productivity explanation for the smoking wage gap. Results in Figure 6 also shed light on the differential explanatory power of life-cycle mechanisms by gender. For men, the decreasing smoking wage gap in physical strenuousness creates an incentive for male smokers to select into physically strenuous occupations. Because physically strenuous occupations require less formal schooling, results in Figure 6d also create a disincentive for male smokers to enroll in costly schooling.

6   Discussion and Conclusion

This research employs a dynamic model of human capital accumulation, wage determination, occupational outcomes, and smoking behavior that allows us to decompose the contemporaneous and life-cycle effects of the smoking wage gap. The model is dynamic in the sense that current period outcomes of wages and occupational task requirements are a function of both current period and past work, smoking, and education decisions. We allow errors that dictate these behavioral and outcome equations to be flexibly correlated, which helps to adjust for permanent unobserved heterogeneity in their joint determination.

We find that the total selectivity corrected penalty for heavy smoking is 14.8% for women and 9.3% for men at age 30. For women, two-thirds of this gap is due to contemporaneous effects, perhaps due to employer preferences, discrimination, or lower productivity; although productivity effects appear to be concentrated in jobs with higher mental reasoning. For men, the contemporaneous difference, about a 1% pay gap, is only 11% of the total penalty associated with smoking. This differential in the mechanisms behind the smoking gap has implications for the benefits of smoking cessation. For women, smoking cessation at age 24 results in an immediate increase in wages. However, for men the attenuation of the penalty is much smaller as the differential investment in human capital and occupational outcomes drive the smoking penalty. Our differential results by gender are consistent with differences the obesity literature, which has found larger spot penalties associated with being obese in women's wages relative to men's (Cawley, 2013; Mason, 2012; Gilleskie et al., 2017).

These differential mechanisms by gender also highlight the importance of dynamics and institutional

characteristics when studying disparities. The spot wage differences by smoking, which is itself a choice, are not just a function of one's current employer's preferences, but, rather, they stem from the lived experience over many years. Smokers make different choices at different points in their labor market and educational experiences and these life-cycle effects compound over time, especially for men.

The smaller life-cycle effect for women has implications for the overall gender gap. In general, the literature has found that women have a greater return to education than men, with some evidence that education appears to dampen the negative effect of discrimination, tastes, and circumstances (DTC) on wages for women (Dougherty, 2005). Given that the negative impact of smoking on school enrollment is relatively small for women, the results suggest that smoking removes at least some of the protection from DTC on wages that is provided to women from education.

Life-cycle effects for both men and women have implications for smoking policies. It appears that the decision to smoke is negatively correlated with the decision to obtain education and to enter lower paying occupations for both men, and to a lesser degree, women. This suggests that the return to youth smoking prevention programs will have much larger returns than have been shown in the literature. For example, the impact of smoking on lifetime earnings, even with cessation at a relatively early age, suggests that the return to youth prevention programs such as "The Real Cost" are under-counted if they do not take into account the lost tax revenue and productivity associated with smoking over the lifetime of the youth (MacMonegle et al., 2018).

References

Auld, Christopher. 2005. Smoking, Drinking, and Income. Journal of Human Resources, **40**(2), 505–518.

Autor, David H. 2019. Work of the Past, Work of the Future. American Economic Association: Papers and Proceeding, **109**(5), 1–32.

Bartik, Timothy, Biddle, Stephen, Hershbein, Brad, & Sotherland, Nathan. 2017. WholeData: Unsuppressed County Business Patterns Data: Version 1.0. Tech. rept. W. E. Upjohn Institute for Employment Research.

Bauer, J., Hyland, A., Li, Qiang, Steger, C., & Cummings, M. 2005. A Longitudinal Assessment of the Impact of Smoke-Free Worksite Policies on Tobacco Use. American Journal of Public Health, **95**(6), 1024–1029.

Berman, M, Crane, R, Seiber, E, & Munur, M. 2014. Estimating the cost of a smoking employee. Tobacco Control, **23**(5).

Callison, Kevin, & Kaestner, Robert. 2014. Do Higher Tobacco Taxes Reduce Adult Smoking? New Evidence of the Effect of Recnt Cigarette Tax Increases on Adult Smoking. Economic Inquiry, **52**(1), 155–172.

Carpenter, C., & Cook, P. 2008. Cigarette taxes and youth smoking: New evidence from national, state, and local Youth Risk Behavior Surveys. Journal of Health Economics, **27**(2), 287–299.

Cawley, J. 2013. The economics of obesity. NBER Reporter, 7–10.

Chaloupka, Frank, Yurekli, A., & Fong, G. 2012. Tobacco taxes as a tobacco control strategy. Tobacco Control, **21**, 172–180.

Cotti, Chad, Courtemanche, Charles, Maclean, Catherine, Nesson, Erik, Pesko, Michael, & Tefft, Nathan. 2020. THE EFFECTS OF E-CIGARETTE TAXES ON E-CIGARETTE PRICES AND TOBACCO PRODUCT SALES: EVIDENCE FROM RETAIL PANEL DATA. NBER Working Paper.

Creamer, MR, Wang, TW, Babb, S, & et al. 2018. Tobacco Product Use and Cessation Indicators Among Adults — United States. MMWR Morb Mortal Wkly Rep. MMWR Morb Mortal Wkly Rep.

Darden, Michael, Gilleskie, Donna, & Strumpf, Koleman. 2018. Smoking and Mortality: New Evidence from a Long Panel. International Economic Review, **59**(3), 1571–1619.

Davis, B., & Carpenter, C. 2009. Proximity of fast-food restaurants to schools and adolescent obesity. Journal Information, **99**(3).

Doll, R., Peto, R., Boreham, J., Gray, R., & Sutherland, I. 2004. Mortality in Relation to Smoking: 50 Years' Observations on Male British Doctors. British Medical Journal, **328**, 1519–1528.

Dougherty, Christopher. 2005. Why Are the Returns to Schooling Higher for Women than for Men? Journal of Human Resources, **40**(4), 969–988.

Farrell, Phillip, & Fuchs, Victor R. 1982. Schooling and Health: The Cigarette Connection. Journal of Health Economics, **1**(3), 217–230.

Fichtenberg, C., & Glantz, S. 2002. Effect of smoke-free workplaces on smoking behaviour: systematic review. BMJ, **325**(7357), 188.

Gilleskie, Donna, Han, Euna, & Norton, Edward. 2017. Disentangling the contemporaneous and dynamic effects of human and health capital on wages over the life cycle. Review of Economic Dynamics, **25**, 350–383.

Grafova, Irina, & Stafford, Frank. 2009. The Wage Effects of Personal Smoking History. Industrial Labor Relations Review, **62**(3), 381.

Hansen, B., Sabia, J., & Rees, D. 2017. Have Cigarette Taxes Lost Their Bite? New Estimates of the Relationship between Cigarette Taxes and Youth Smoking. American Journal of Health Economics, **3**(1).

Heckman, James J., & Singer, Burton. 1984. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. Econometrica, **52**(2), 271–320.

HHS. 1964. Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service. Report. US Department of Health, Education, and Welfare.

Hotchkiss, Julie L., & Pitts, M. Melinda. July 2013. Even One Is Too Much: The Economic Consequences of Being a Smoker. Working Paper 2013-3. Federal Reserve Bank of Atlanta.

Keane, Michael, & Wolpin, Kenneth. 1997. The Career Decisions of Young Men. Journal of Political Economy, **105**(3), 473–522.

Lemieux, P. 2019. CONSUMER SURPLUS IN THE FDA'S TOBACCO REGULATIONS WITH APPLICATIONS TO NICOTINE REDUCTION AND E-CIGARETTE FLAVORS. Report. Reason Foundation.

MacMonegle, A.J., Nonnemaker, J., Duke, J.C., Farrelly, M.C., X., Zhao, Delahanty, J.C., Smith, A.A., Rao, P., & J.A., Allen. 2018. Cost-Effectiveness Analysis of the Real Cost Campaign's Effect on Smoking Prevention. American Journal of Preventative Medicine, **55**(3), 319–325.

Mason, Katherine. 2012. The Unequal Weight of Discrimination: Gender, Body Size, and Income Inequality. Journal of Human Resources, **59**(3), 411–435.

Mroz, Thomas. 1999. Discrete Factor Approximations in Simultaneous Equation Models: Estimating the Impact of a Dummy Endogenous Variable on a Continuous Outcome. Journal of Econometrics, **92**(2), 233–274.

Siegel, M., Albers, A., Cheng, D., W., Hamilton, & Biner, L. 2009. Local Restaurant Smoking Regulations and the Adolescent Smoking Initiation Process. Archives of Pediatric Adolescent Medicine, **162**(5), 477–483.

Sloan, Frank, Ostermann, Jan, Conover, Christopher, Taylor, Donald, & Picone, Gabriel. 2006. The Price of Smoking. MIT press.

van Ours, Jan. 2004. A pint a day raises a man's pay; but smoking blows that gain away. Journal of Health Economics, **23**, 863–886.

Xu, Xin, Bishop, Ellen, Kennedy, Sara, Simpson, Sean, & Pechacek, Terry. 2015. Annual Healthcare Spending Attributable to Cigarette Smoking: An Update. American Jouranal of Preventive Medicine, **48**(3), 326–333.

A   Main Tables

Table 1: Sample Construction

# Ind.	Inclusion Criteria
8,984	Base Sample of NLSY 1997.
8,804	Complete school enrollment information
8,181	16 or younger in 1997
7,590	Complete smoking information
6,919	Complete employment information
6,755	Never in the military
6,520	Complete location (county and state) information
5,529	Enrolled in School and not pregnant at age 16

Notes: The 5,529 individuals we study generate 93,993 person/year observations. The sample includes 2,675 women and 2,854 men.

Table 2: Exogenous and Pre-Determined Variables by Smoking

	Women, n=45,475			Men, n=48,518		
	Non-Smokers	Light	Heavy	Non-Smokers	Light	Heavy
	0.738	0.086	0.176	0.663	0.112	0.225
	Exogenous Characteristics					
Age	23.523	22.508	23.269	23.437	22.850	23.337
Asvab	41.935	45.968	41.298	41.104	39.332	37.323
Black	0.284	0.168	0.149	0.247	0.204	0.209
Hispanic	0.233	0.233	0.134	0.224	0.288	0.120
MSA Unemployment Rate *10	64.386	60.376	62.085	62.574	61.337	61.350
Population	0.381	0.400	0.365	0.380	0.378	0.377
Manufacturing Share	0.120	0.121	0.137	0.120	0.120	0.134
Service Share	0.511	0.511	0.492	0.510	0.505	0.500
Computing Share	0.059	0.060	0.061	0.059	0.060	0.059
Earnings per Employee	37.378	37.336	34.842	37.303	36.158	35.605
Share Missing	0.037	0.029	0.034	0.036	0.044	0.035
Father's Education Low	0.176	0.125	0.160	0.173	0.190	0.183
Father's Education Med.	0.313	0.296	0.358	0.304	0.280	0.366
Father's Education High	0.335	0.426	0.284	0.348	0.342	0.295
Mother's Education Low	0.219	0.142	0.174	0.184	0.214	0.165
Mother's Education Med.	0.321	0.308	0.386	0.349	0.323	0.387
Mother's Education High	0.409	0.509	0.388	0.398	0.376	0.384
Mother Education Missing	0.051	0.041	0.053	0.069	0.087	0.064
Father Education Missing	0.176	0.153	0.198	0.175	0.189	0.157
Pack Sales Per Capita (within State)	60.568	61.928	64.437	59.498	59.989	64.165
Gross Cigarette Tax Rev. (m $)	60.325	62.110	54.445	61.861	59.132	54.777
State Cigarette Tax	1.047	1.045	1.081	1.096	1.040	1.065
Local Area Cigarette Tax	0.063	0.055	0.052	0.065	0.050	0.044
Mean Cost per Pack	5.228	5.153	5.252	5.298	5.170	5.223
Workplace Indoor Smoking Bans						
Complete Ban, Partial County	0.308	0.330	0.223	0.327	0.319	0.213
Partial Ban, Partial County	0.629	0.652	0.573	0.632	0.645	0.548
Comlete Ban, Complete County	0.008	0.004	0.007	0.011	0.014	0.018
Partial Ban, Complete County	0.032	0.025	0.033	0.028	0.031	0.032
Statewide Complete Ban	0.233	0.179	0.210	0.212	0.188	0.235
Statewide Partial Ban	0.754	0.796	0.770	0.781	0.767	0.735
Missing Ban Information	0.003	0.003	0.003	0.003	0.004	0.010
	Pre-Determined Characteristics					
Job Experience (Years)	4.991	4.326	4.662	4.859	4.315	4.602
Job Tenure (Years)	0.991	0.734	0.710	1.062	0.855	0.790
Duration Smoking (Years)	1.492	4.635	7.650	1.780	4.906	7.388
Cessation Smoking (Years)	1.792	0.764	0.197	1.769	0.700	0.189
Years of Schooling	13.402	12.987	12.119	13.006	12.677	11.895
Years≥12	0.702	0.662	0.482	0.640	0.611	0.464
Years≥16	0.213	0.152	0.085	0.172	0.127	0.063

Notes: Table presents exogenous and pre-determined variables by smoking status. Light smoking is defined as smoking on between 1 and 20 of the previous 30 days from each NLSY wave. Heavy smoking is defined as smoking on 20 or more of the previous 30 days from each NLSY wave. Data are from our main estimation sample, where the person/age sample sizes for women and men are 45,475 and 48,518, respectively. All dollar figures are in 2015 dollars.

Table 3: Factor Construction from O*NET

Mental Reasoning Factor		Physical Strenuousness Factor	
Measurement	Weight	Measurement	Weight
Oral Expression	0.357	Spend Time Climbing	0.398
Written Expression	0.357	Outdoors Under Cover	0.346
Writing	0.349	Outdoors No Cover	0.336
Oral Comprehension	0.309	Exposed Whole Body	0.301
Complex Problem Solving	0.309	Exposed Extreme Light	0.265
Written Comprehension	0.296	Spend Time Balancing	0.263
Reading Comprehension	0.291	Exposed Extreme Temp	0.251
Fluency of Ideas	0.237	Confined Space Work	0.248
Deductive Reasoning	0.233	Spend Time Kneel Crouch	0.234
Inductive Reasoning	0.228	Indoors No AC	0.198
Science	0.173	Dynamic Flexibility	0.188
Memorization	0.152	Gross Body Equilibrium	0.165
Analytical Thinking	0.121	Exposed To Hazardous	0.137
Category Flexibility	0.114	Exposed Loud Noises	0.109
Mathematical Reasoning	0.046	Exposed To Nicks	0.089
Critical Thinking	0.043	TimeBending Twisting Body	0.078
Speed of Closure	0.011	Extent Flexibility	0.059
Number Facility	0.010	Speed Of Limb Movement	0.048
Mathematics	-0.055	Gross Body Coordination	0.028
Flexibility of Closure	-0.105	Spend Time Walk Run	0.023
		Spend Time Sitting	0.022
		Stamina	0.010
		Dynamic Strength	0.008
		Spend Time Repetitive Motions	-0.001
		Multi-limb Coordination	-0.008
		Trunk Strength	-0.015
		Response Orientation	-0.020
		Static Strength	-0.024
		Reaction Time	-0.025
		Rate Control	-0.025
		Explosive Strength	-0.031
		Spend Time Standing	-0.044
		Manual Dexterity	-0.073
		Indoors AC	-0.187

Notes: Table demonstrates the 2010 occupation code O*NET measurements and the associated weights from the first principal component of principal component analysis. The first principal component of the first column measurements explains 53.08% of the variation, whereas the first principal component of the third column explains 63.22 % of the variation.

Table 4: Endogenous Variables by Smoking

	Women, n=45,475			Men, n=48,518		
	Non-Smokers	Light	Heavy	Non-Smokers	Light	Heavy
	0.738	0.086	0.176	0.663	0.112	0.225
School Enrollment	0.346	0.395	0.230	0.312	0.301	0.167
Part-Time Work	0.220	0.250	0.192	0.153	0.143	0.115
Full-Time Work	0.513	0.486	0.510	0.586	0.563	0.593
Log Wage | Working	7.176	7.117	6.969	7.295	7.231	7.165
Mental Reasoning | Working	0.459	0.435	0.394	0.417	0.397	0.358
Physical Strenuousness | Working	0.209	0.217	0.238	0.360	0.373	0.433

Notes: Table presents endogenous variables by smoking status. Mental reasoning and physical strenuousness are occupation-specific measures of task requirements constructed from ONET data (see Table 3). Light smoking is defined as smoking on between 1 and 20 of the previous 30 days from the NLSY 1997 interview. Heavy smoking is defined as smoking on 20 or more of the previous 30 days from the NLSY 1997 interview. Data are from our main estimation sample, where the person/age sample sizes for women and men are 45,475 and 48,518, respectively.

Table 5: Log Wage Equation Parameters

	Women				Men			
	Separately		Jointly		Separately		Jointly	
	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.
L. Light Smoking	-0.021	0.031	-0.022	0.033	-0.003	0.032	0.032	0.034
L. Heavy Smoking	-0.032	0.027	-0.012	0.027	0.023	0.026	0.041	0.026
Age*L. Light Smoking	0.005	0.009	0.001	0.010	-0.008	0.009	-0.016	0.009
Age Squared/100 * L. Light Smoking	-0.037	0.057	0.008	0.062	0.069	0.054	0.112	0.055
Age*L. Heavy Smoking	-0.014	0.007	-0.010	0.007	-0.019	0.007	-0.011	0.007
Age Squared/100 * L. Heavy Smoking	0.051	0.044	0.027	0.042	0.097	0.042	0.053	0.040
L. Pregnancy	0.003	0.009	0.011	0.009	0.000	.	0.000	.
L. School Enrollment	-0.073	0.008	-0.081	0.008	-0.087	0.009	-0.098	0.008
L. Mental Reasoning	0.655	0.019	0.329	0.016	0.599	0.019	0.221	0.015
L. Physical Strenuousness	-0.266	0.023	-0.108	0.022	0.194	0.014	0.120	0.013
L. Child in the Home	-0.026	0.008	-0.046	0.007	0.087	0.009	0.067	0.006
L. Part-Time Work	-0.081	0.010	-0.040	0.009	-0.115	0.011	-0.048	0.010
L. Full-Time Work	0.024	0.010	0.028	0.010	0.007	0.010	0.021	0.010
L. Mental Reasoning Missing	0.285	0.031	0.159	0.024	0.353	0.028	0.194	0.018
L. Years of Schooling≥12	-0.057	0.010	-0.075	0.009	-0.030	0.010	-0.022	0.007
L. Years of Schooling≥16	0.042	0.013	0.060	0.011	-0.014	0.015	0.020	0.010
L. Years of Schooling	0.034	0.003	0.047	0.003	0.029	0.003	0.037	0.002
Job Tenure Years	0.018	0.002	0.009	0.002	0.020	0.002	0.010	0.002
Job Experience Years	0.016	0.002	0.019	0.001	0.010	0.002	0.016	0.001
Smoking Duration	0.003	0.001	0.002	0.001	-0.001	0.001	-0.001	0.001
Smoking Cessation	0.000	0.001	0.001	0.001	0.004	0.001	0.003	0.001
Age	0.003	0.004	0.009	0.004	0.014	0.004	0.021	0.004
Age Squared/100	-0.046	0.022	-0.063	0.020	-0.078	0.023	-0.100	0.020
Black	-0.013	0.013	0.008	0.012	-0.089	0.013	-0.139	0.011
Hispanic	0.012	0.013	0.036	0.013	-0.010	0.012	-0.036	0.010
Asvab Score	0.001	0.000	0.002	0.000	0.001	0.000	0.000	0.000
Asvab Score Missing	0.039	0.010	0.038	0.010	0.000	0.010	-0.053	0.008
Survey Weight	-0.003	0.005	0.016	0.005	0.007	0.006	-0.014	0.005
Mother's Education Med.	-0.004	0.009	0.010	0.009	-0.006	0.009	0.004	0.008
Mother's Education High	-0.011	0.009	-0.015	0.010	-0.004	0.010	0.008	0.009
Mother's Education Missing	0.048	0.016	0.056	0.014	-0.019	0.014	0.008	0.012
Father's Education Med.	-0.012	0.009	-0.015	0.008	0.008	0.009	0.002	0.008
Father's Education High	0.012	0.010	0.042	0.009	0.016	0.010	-0.002	0.008
Father's Education Missing	-0.016	0.010	0.000	0.010	-0.007	0.011	-0.030	0.009
MSA UR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Population	-0.252	0.028	-0.151	0.023	-0.316	0.029	-0.233	0.024
Man. Share	-0.234	0.051	-0.164	0.044	-0.158	0.053	-0.172	0.042
Service Share	-0.198	0.053	-0.191	0.042	-0.272	0.053	-0.285	0.040
Comptuer Share	0.173	0.124	0.241	0.098	0.374	0.126	0.102	0.096
Earnings	0.011	0.000	0.010	0.000	0.010	0.000	0.009	0.000
Economics Variables Missing	-0.025	0.018	0.008	0.016	-0.014	0.018	-0.021	0.016
Constant	6.222	0.052	5.935	0.045	6.360	0.052	6.147	0.043
μ_2^w			0.334	0.006			0.757	0.008
μ_3^w			0.163	0.009			0.310	0.007
μ_4^w			-0.442	0.008			0.479	0.008

Notes: Estimates of the log wage equation by gender. Columns labeled "separate" present results from estimation of the log wage equation separately from other equations that dictate behavior and outcomes. Columns labeled "jointly" present results from our preferred model, which estimates all equations jointly and allows for unobserved heterogeneity. The person/age sample sizes for women and men are 45,475 and 48,518, respectively.

Table 6: Unobserved Heterogeneity Parameters

	L. Smoking		H. Smoking		S. Enrollment		Part-Time		Full-Time	
Women, $n = 45,475$										
	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.
Mass Points										
μ_2	0.060	0.060	-0.248	0.073	-0.133	0.051	-0.226	0.059	0.093	0.054
μ_3	-0.075	0.104	0.315	0.098	-0.347	0.081	-0.242	0.082	-0.159	0.077
μ_4	0.129	0.093	0.412	0.091	0.030	0.073	0.198	0.084	-0.217	0.083
	Attrition		Log Wages		M. Reasoning		P. Strenuousness			
μ_2	0.194	0.164	0.334	0.006	0.048	0.002	-0.015	0.002		
μ_3	-0.310	0.306	0.163	0.009	-0.027	0.003	0.095	0.003		
μ_4	-0.546	0.269	-0.442	0.008	-0.057	0.002	0.030	0.003		
	Type 2		Type 3		Type 4					
Age 16 Initial Conditions										
Constant ψ_0	-0.235	0.194	-0.040	0.304	-0.389	0.333				
Light Smoking ψ_1	-0.272	0.222	0.561	0.260	0.444	0.248				
Heavy Smoking ψ_2	0.072	0.127	-0.152	0.203	-0.041	0.196				
Part-Time Work ψ_3	0.232	0.307	0.339	0.442	-0.169	0.514				
Full-Time Work ψ_4	-0.519	0.093	-1.645	0.144	-1.722	0.137				
Men, $n = 48,518$										
	L. Smoking		H. Smoking		S. Enrollment		Part-Time		Full-Time	
Mass Points										
	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.
μ_2	0.055	0.083	-0.628	0.088	-0.131	0.080	-0.300	0.085	0.369	0.077
μ_3	0.181	0.068	-0.178	0.064	0.069	0.059	-0.099	0.068	0.354	0.059
μ_4	0.044	0.076	-0.273	0.072	-0.619	0.076	-0.447	0.083	0.443	0.068
	Attrition		Log Wages		M. Reasoning		P. Strenuousness			
μ_2	0.224	0.258	0.757	0.008	0.110	0.003	-0.040	0.004		
μ_3	0.482	0.217	0.310	0.007	0.066	0.002	-0.037	0.003		
μ_4	1.110	0.217	0.479	0.008	0.045	0.003	0.146	0.003		
	Type 2		Type 3		Type 4					
Age 16 Initial Conditions										
Constant ψ_0	-0.212	0.273	-0.090	0.219	0.118	0.232				
Light Smoking ψ_1	-0.780	0.295	-0.348	0.209	-0.102	0.217				
Heavy Smoking ψ_2	-0.089	0.192	0.290	0.152	0.062	0.166				
Part-Time Work ψ_3	0.260	0.333	-0.039	0.285	0.319	0.283				
Full-Time Work ψ_4	-0.287	0.125	0.726	0.109	-0.027	0.116				

Notes: Estimates and their standard errors of the distribution of unobserved heterogeneity by gender.

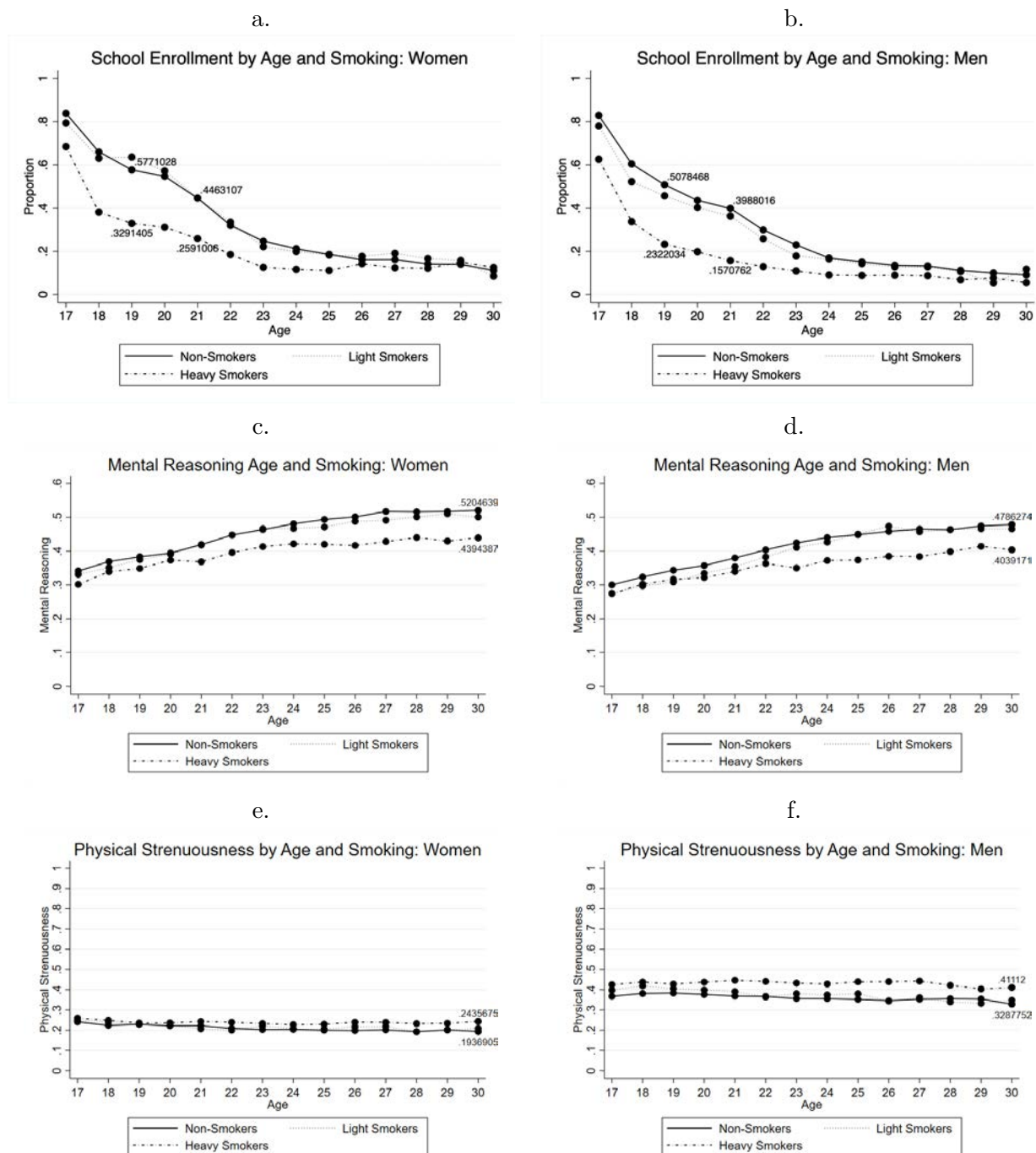**Main Figures**

Figure 1

a.   b.
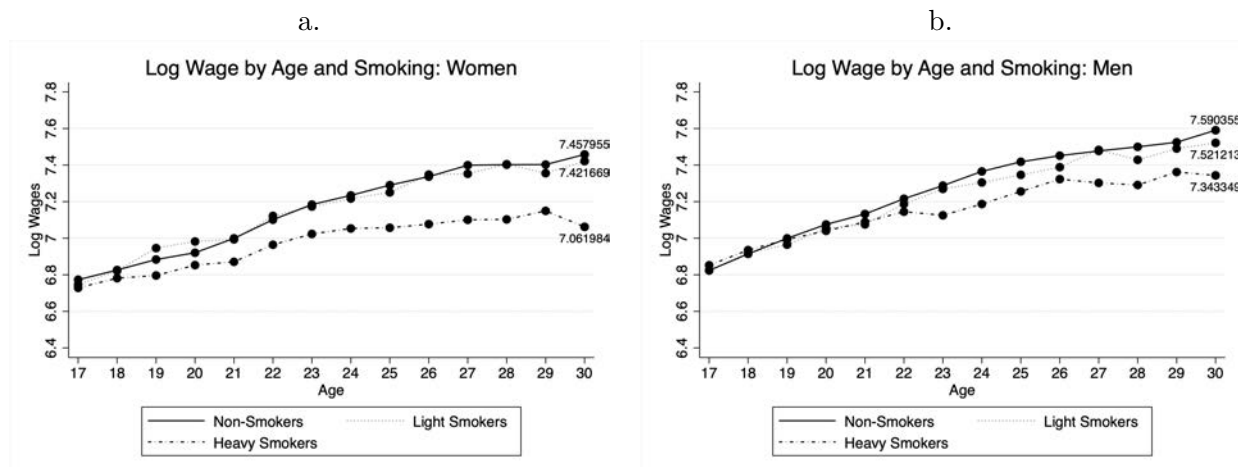
Notes: The figure presents the mean log wage conditional on working for women (a.) and men (b.) by current smoking status and by age. Light smoking is defined as smoking on between 1 and 20 of the previous 30 days from each NLSY wave. Heavy smoking is defined as smoking on 20 or more of the previous 30 days from each NLSY wave. Data are from our main estimation sample, where the person/age sample sizes for women and men are 45,475 and 48,518, respectively.

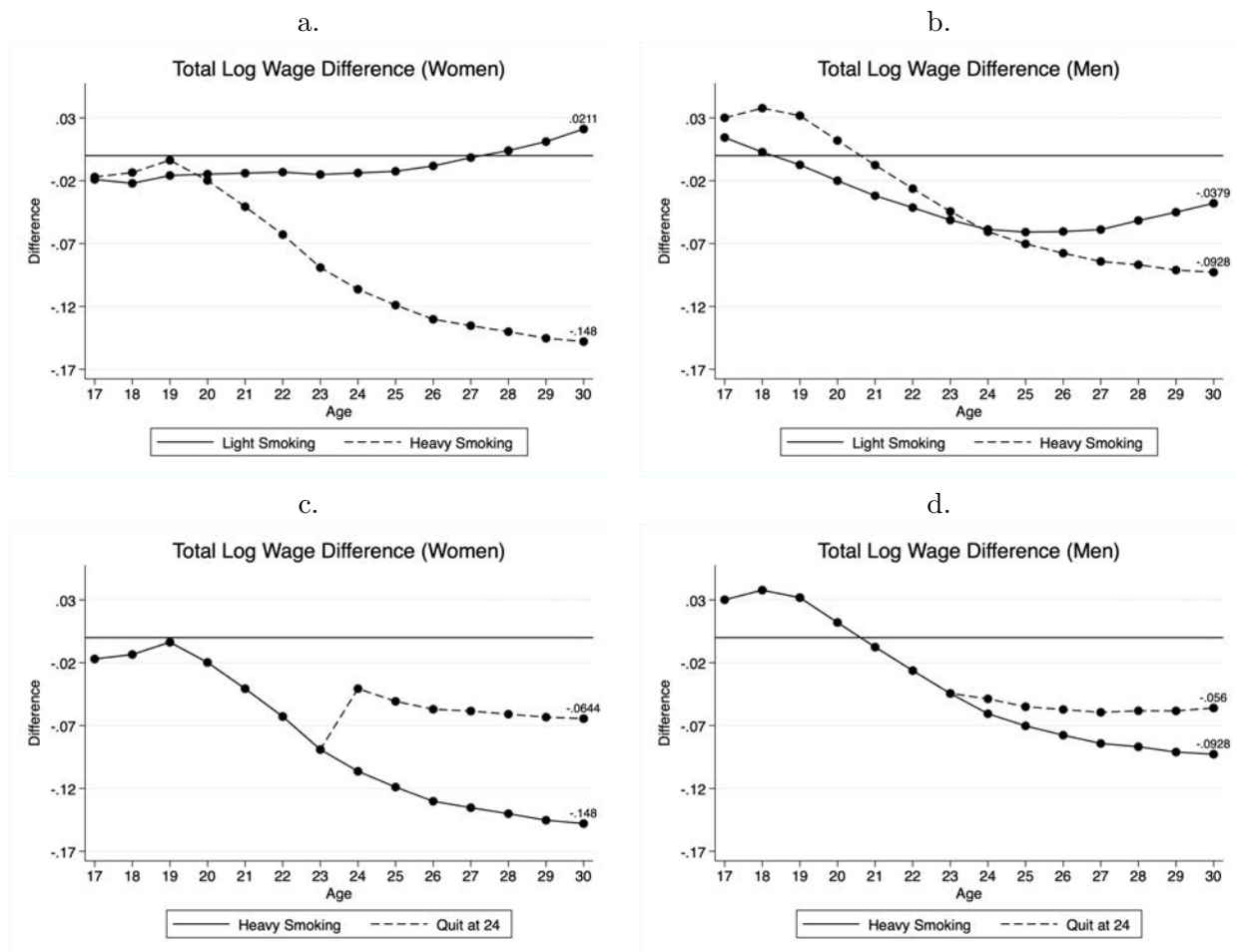Figure 2

a.

b.

c.

d.

e.

f.

Notes: The figure presents the mean level of school enrollment and, conditional on working, the mean level of occupational mental reasoning and physical strenuousness for women (a., c., and, e.) and men (b., d., and f.) by current smoking status and by age. Light smoking is defined as smoking on between 1 and 20 of the previous 30 days from each NLSY wave. Heavy smoking is defined as smoking on 20 or more of the previous 30 days from each NLSY wave. Data are from our main estimation sample, where the person/age sample sizes for women and men are 45,475 and 48,518, respectively.
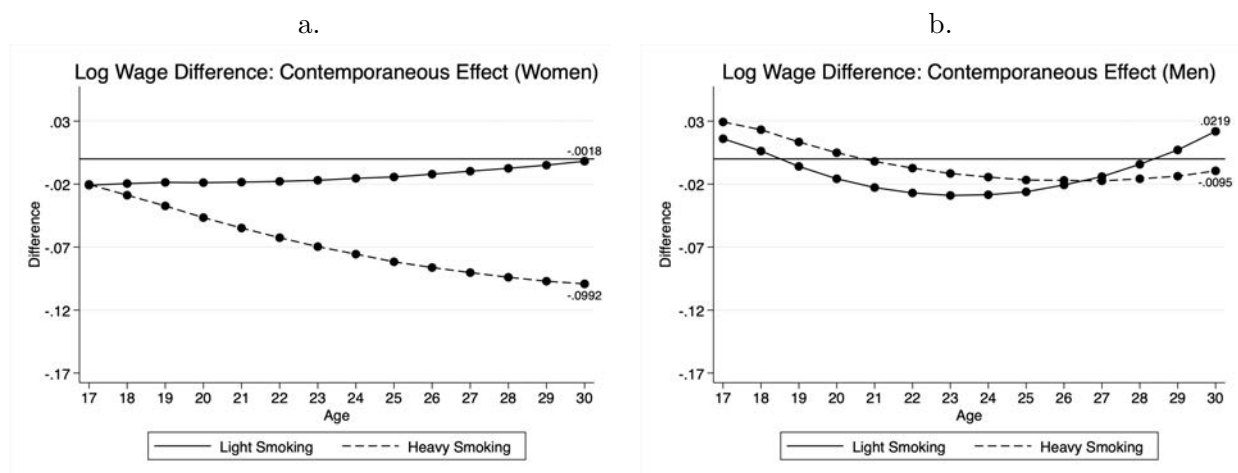
Figure 3

a.

b.

c.

d.
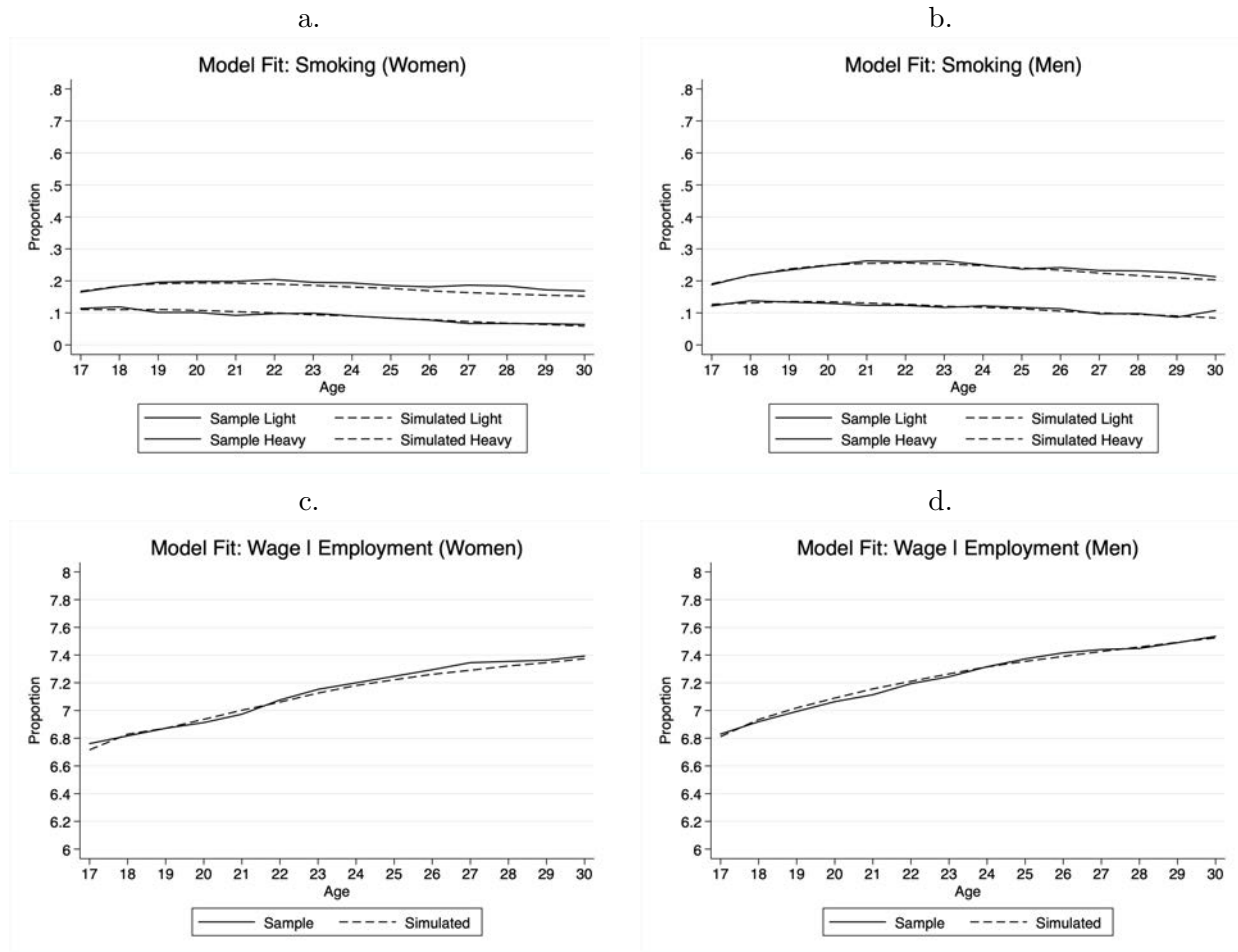
Notes: The figures represent the age path of each behavior or outcome in the data and from the baseline simulation of our model for women (a. and c.) and men (b. and d.). Simulation results represent those simulated to remain in the sample (i.e., not attrit). In cases in which a sample individual has left the sample through attrition but we simulate them to remain, their exogenous characteristics are fixed at their last observation and their endogenous characteristics are updated given other simulated behaviors and outcomes. Light smoking is defined as smoking on between 1 and 20 of the previous 30 days from each NLSY wave. Heavy smoking is defined as smoking on 20 or more of the previous 30 days from each NLSY wave. Data are from our main estimation sample, where the person/age sample sizes for women and men are 45,475 and 48,518, respectively.
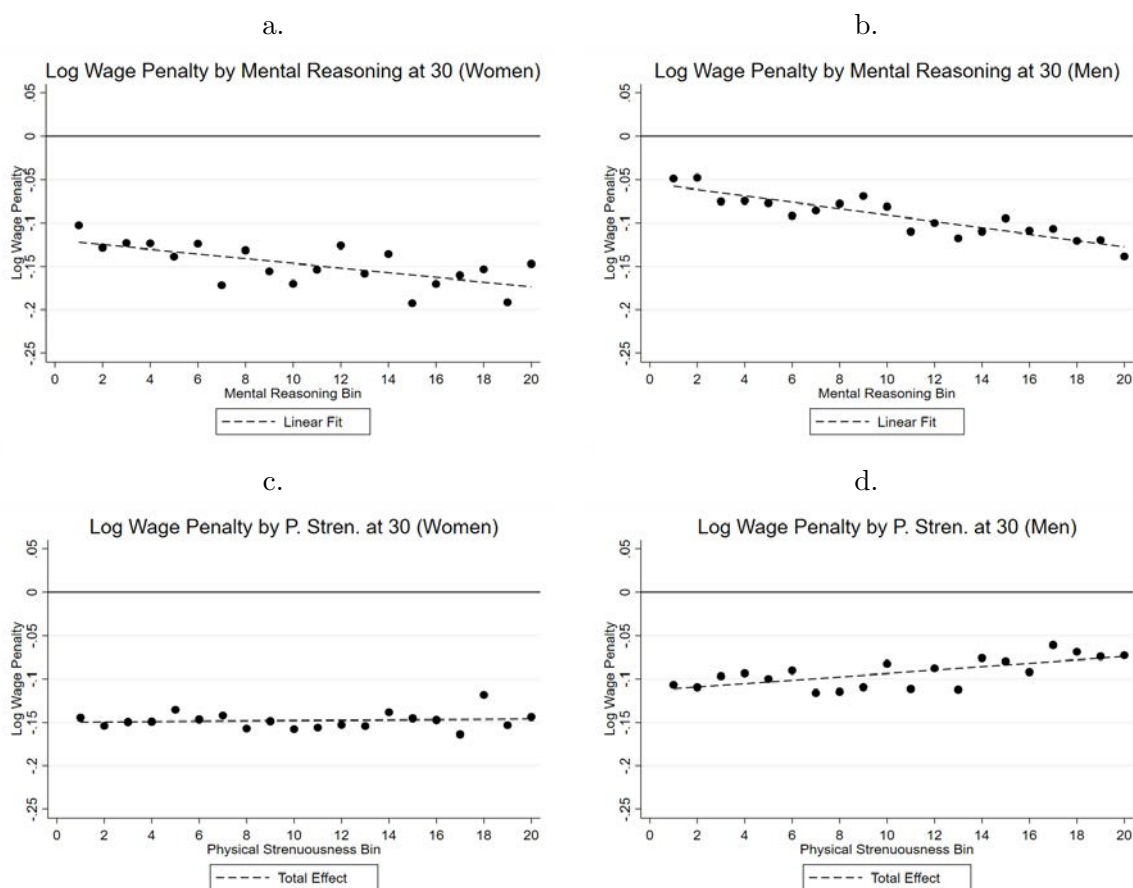
Figure 4

a.

b.

c.

d.

Notes: The figure represents the simulated difference in log wages by age for different smoking behaviors relative to non-smoking for women (a.) and men (b.). Smoking behavior is imposed on all simulated individuals from ages 17 through 30. Light smoking is defined as smoking on between 1 and 20 of the previous 30 days from each NLSY wave. Heavy smoking is defined as smoking on 20 or more of the previous 30 days from each NLSY wave. Data are from our main estimation sample, where the person/age sample sizes for women and men are 45,475 and 48,518, respectively.

Figure 5

a.   b.

Notes: Figures a. and b. represent the difference in log wages between simulations in which individuals are assumed to not smoke in equations that affect school enrollment, employment, job task requirements, and attrition, while various smoking behaviors are imposed in the wage equation. Both lines are relative to the case in which a person is assumed to not smoke contemporaneously in the wage equation. Light smoking is defined as smoking on between 1 and 20 of the previous 30 days from each NLSY wave. Heavy smoking is defined as smoking on 20 or more of the previous 30 days from each NLSY wave. Data are from our main estimation sample, where the person/age sample sizes for women and men are 45,475 and 48,518, respectively.

Figure 6

a.

Log Wage Penalty by Mental Reasoning at 30 (Women)

b.

Log Wage Penalty by Mental Reasoning at 30 (Men)

c.

Log Wage Penalty by P. Stren. at 30 (Women)

d.

Log Wage Penalty by P. Stren. at 30 (Men)

Notes: Each figure represents the log wage penalty, both light and heavy smoking, by the bin occupational task bin. For each type of occupational task, we divide the simulated distribution into 20(gender-specific) bins at age 30 and calculate the wage penalty associated with always smoking heavily relative to not smoking. Heavy smoking is defined as smoking on 20 or more of the previous 30 days from each NLSY wave. Data are from our main estimation sample, where the person/age sample sizes for women and men are 45,475 and 48,518, respectively.