Smoking and Mortality: New Evidence from a Long Panel Online Appendix*

Michael Darden[1], Donna B. Gilleskie[2], and Koleman Strumpf[3]

[1] *Department of Economics, Tulane University*
[2] *Department of Economics, University of North Carolina at Chapel Hill*
[3] *School of Business, University of Kansas*

June 2015

*The Framingham Heart Study (FHS) is conducted and supported by the NHLBI in collaboration with the FHS Study Investigators. This manuscript was prepared using a limited access dataset obtained from the NHLBI and does not necessarily reflect the opinions or views of the FHS or the NHLBI. We appreciate the comments of seminar participants at Indiana University, Tulane University, the University of York, Lund University, the Health Econometrics Symposium at Leeds University, the 5th (Spain) and 10th (Dublin) iHEA World Congress, the Econometric Society Summer Meetings, and the Triangle Health Economics Workshop. Financial support from the National Institutes for Health (grant #1RO1HD42256-01) is gratefully acknowledged. Correspondence: mdarden1@tulane.edu; dgill@email.unc.edu; cigar@ku.edu.

A Other Tables

Table A1: Summary of Equation Specifications

Outcome	Estimator	Explanatory Variables		
		Endogenous	Exogenous	Unobs'd Het
Smoking behavior, S_t				
Smoke at t	logit	H_t^S, H_t^D	X_t, P_t	$\mu^S, \nu_t^S, \epsilon_t^S$
Morbidity outcomes, D_t^d				
CVD events at t	mlogit	H_t^S, H_t^D	X_t	$\mu^{D^1}, \nu_t^{D^1}, \epsilon_t^{D^1}$
Cancer diagnosis at t | no cancer up to t	logit	H_t^S, H_t^D	X_t	$\mu^{D^2}, \nu_t^{D^2}, \epsilon_t^{D^2}$
Diabetes diagnosis at t | no diabetes up to t	logit	H_t^S, H_t^D	X_t	$\mu^{D^3}, \nu_t^{D^3}, \epsilon_t^{D^3}$
Body Mass Index at t	OLS	H_t^S, H_t^D	X_t	$\mu^{D^4}, \nu_t^{D^4}, \epsilon_t^{D^4}$
Mortality outcomes, M_{t+1}				
Death hazard by end of t	logit	H_{t+1}^S, H_{t+1}^D	X_t	$\mu^{M^1}, \nu_t^{M^1}, \epsilon_t^{M^1}$
Cause of death in t | death in t	mlogit	H_{t+1}^S, H_{t+1}^D	X_t	$\mu^{M^2}, \nu_t^{M^2}, \epsilon_t^{M^2}$
Initial conditions, $I_2^{S'}, I_2^{D'}$				
Never smoked up to $t = 2$	logit		X_1, P_1, Z_1	$\mu^{I^{S_1}}, \epsilon_1^{I^{S_1}}$
Current smoker in $t = 1$ | ever smoked	logit		X_1, P_1, Z_1	$\mu^{I^{S_2}}, \epsilon_1^{I^{S_2}}$
Years of smoking up to $t = 2$ | smoker in $t = 1$	OLS		X_1, P_1, Z_1	$\mu^{I^{S_3}}, \epsilon_1^{I^{S_3}}$
Any CVD up to $t = 2$	logit		X_1, P_1, Z_1	$\mu^{I^{D_1}}, \epsilon_1^{I^{D_1}}$
Body Mass Index in $t = 1$	OLS		X_1, P_1, Z_1	$\mu^{I^{D_2}}, \epsilon_1^{I^{D_2}}$

Table A2: Estimation Results: Unobserved Heterogeneity Distributions

Mass point	Smoking	Mortality	Cause of death CVD	Cause of death Cancer	CVD=1	CVD>1	Cancer diagnosis	Diabetes diagnosis	BMI	Mass point probability
Time-invariant unobserved heterogeneity										
μ_2	-0.521	-0.179	1.037	0.132	-0.144	0.511	0.126	0.020	0.010	0.032
	(0.454)	(0.476)	(0.837)	(0.998)	(0.411)	(0.637)	(0.638)	(0.626)	(0.018)	
μ_3	-2.205 ***	0.089	1.312 **	-0.304	0.252	0.846	0.332	0.201	0.072 ***	0.147
	(0.641)	(0.354)	(0.588)	(0.703)	(0.24)	(0.534)	(0.542)	(0.511)	(0.009)	
μ_4	-2.095 ***	0.010	-0.020	-1.180	0.861 **	0.386	0.127	0.707	0.171 ***	0.034
	(0.691)	(0.435)	(0.732)	(0.846)	(0.354)	(0.961)	(0.767)	(0.67)	(0.018)	
μ_5	-2.799 ***	-0.588	1.031 *	0.612	0.045	0.562	0.505	-2.024 **	-0.040 ***	0.297
	(0.357)	(0.375)	(0.543)	(0.623)	(0.255)	(0.475)	(0.475)	(0.861)	(0.008)	
μ_6	-3.061 ***	-0.198	1.384 **	0.508	0.168	0.513	0.521	-0.755 *	0.007	0.353
	(0.353)	(0.476)	(0.596)	(0.713)	(0.213)	(0.526)	(0.495)	(0.387)	(0.007)	
Time-varying unobserved heterogeneity										
ν_{t2}	-0.566	-0.531	-0.444	-0.126	-0.896	-3.761 ***	-0.505	-0.760	0.976 ***	0.023
	(1.674)	(0.677)	(1.257)	(1.431)	(0.654)	(1.195)	(1.288)	(1.134)	(0.084)	
ν_{t3}	4.508 **	-0.099	1.356	-0.109	-1.606 **	-3.353 ***	-0.953	-2.800	0.689 ***	0.548
	(2.05)	(0.738)	(0.845)	(1.126)	(0.644)	(0.626)	(1.19)	(2.863)	(0.06)	
ν_{t4}	0.291	-1.572 *	1.822	1.512	-0.704	-2.707 ***	0.037	-0.143	0.703 ***	0.391
	(1.692)	(0.868)	(1.201)	(1.293)	(0.58)	(0.667)	(1.13)	(1.011)	(0.062)	
ν_{t5}	3.578	-0.310	1.338 *	0.429	0.380	-0.403	0.706	1.040	0.414 ***	0.034
	(2.196)	(0.699)	(0.811)	(1.171)	(0.606)	(0.555)	(1.169)	(0.998)	(0.044)	
ν_{t6}	0.044	0.280	0.329	-1.862	-0.945	-2.339 ***	-1.461	-2.616 **	1.586 ***	0.001
	(2.292)	(1.051)	(1.611)	(1.926)	(1.4)	(0.622)	(1.166)	(1.128)	(0.171)	

Note: First mass point vector values normalized to zero for both distrubutions, and occur with probabilities 0.137 and 0.003. Standard errors are in parentheses. *** indicates joint significance at the 1% level; ** 5% level; * 10% level.

Table A3: Parameter Estimates: Initial Conditions

	Never Smoked			Current Smoker			Smoking Duration			CVD			BMI		
Age (years)	0.067	0.023	***	-0.117	0.035	***	0.110	0.010	***	0.132	0.061	**	0.003	0.003	
Educ: grade school	-0.123	0.199		-0.534	0.299	*	0.058	0.054		-0.647	0.479		-0.019	0.031	
Educ: some high school	-0.524	0.222	**	0.412	0.356		0.027	0.046		-0.693	0.589		-0.028	0.030	
Educ: some college	-0.291	0.272		-1.143	0.395	***	-0.128	0.061	**	-0.656	0.683		-0.025	0.036	
Educ: college degree	0.371	0.248		-0.628	0.385		-0.179	0.057	***	-0.578	0.623		-0.012	0.029	
Educ: post college	0.635	0.249	**	-0.216	0.440		-0.280	0.078	***	0.288	0.574		0.048	0.032	
Born outside U.S.	0.040	0.197		-0.231	0.293		-0.072	0.062		-0.200	0.459		-0.007	0.026	
Italian ancestry	0.351	0.176	**	-0.219	0.259		-0.016	0.042		-0.510	0.560		0.142	0.025	***
Older cohort	-0.288	0.269		-0.144	0.392		-0.207	0.099	**	-0.016	0.627		-0.033	0.037	
Ad expenditure at age 10-14	0.177	0.111		-0.428	0.172	**				0.221	0.335		-0.009	0.011	
Ad expenditure at age 15-18							0.065	0.033	**						
Number of siblings	0.019	0.032		-0.001	0.046		0.006	0.009		-0.006	0.088		-0.007	0.003	**
Only child	-0.869	0.441	**	1.238	0.806		0.106	0.065		-0.464	1.092		-0.013	0.037	
Sibling information missing	-0.568	0.272	**	1.849	0.441	***	0.149	0.065	**	-0.323	0.769		0.008	0.036	
Birth order (up to 5th)	-0.008	0.058		0.024	0.091		-0.012	0.016		-0.383	0.212	*	0.014	0.009	*
First born child										-0.739	0.655		0.005	0.026	
Constant	-5.556	1.354	***	6.958	1.721	***	-2.288	0.490	***	-7.307	3.187	**	2.469	0.141	***
Time-invariant unobserved heterogeneity															
μ_2	1.823	0.844	**	0.553	0.619		0.023	0.109		0.795	0.743		0.722	0.062	***
μ_3	2.086	0.854	**	1.355	0.640	**	-0.066	0.075		-2.477	1.168	**	0.040	0.063	
μ_4	-1.060	5.472		2.757	0.674	***	-1.991	0.224	***	-2.666	1.069	**	0.025	0.110	
μ_5	0.719	0.812		2.042	0.458	***	-0.049	0.072		-1.495	0.607	**	-0.309	0.046	***
μ_6	1.307	0.892		0.611	0.440		-0.118	0.112		-2.275	1.720		0.322	0.056	***

Note: Standard errors are in parentheses. ** indicates joint significance at the 5% level; * 10% level.

Table A4: Selected Parameter Estimates: Cause of Death conditional on Death by end of period

Variable	Cardiovascular disease						Cancer					
	Without UH			With UH			Without UH			With UH		
	Estimate	Std. error		Estimate	Std. error		Estimate	Std. error		Estimate	Std. error	
Smoker in t, $\mathtt{S}_t$	0.537	0.317	*	0.974	0.747		0.381	0.337		1.306	0.588	**
Years of cessation, $\mathtt{C}_t$	0.002	0.007		0.008	0.008		0.012	0.008		0.018	0.009	*
Years of duration, $\mathtt{D}_t$	-0.006	0.007		-0.005	0.014		0.000	0.007		-0.011	0.011	
Years of experience, $\mathtt{E}_t$	-0.003	0.005		-0.005	0.005		0.001	0.005		-0.002	0.006	
$1[\mathtt{CVD}_t = 1]$	0.967	0.254	***	1.067	0.371	***	-0.188	0.333		-0.340	0.408	
$1[\mathtt{CVD}_t > 1]$	1.428	0.463	***	1.373	0.646	**	0.094	0.639		-0.083	0.787	
$\mathtt{CAN}_t$	1.664	0.652	**	1.643	0.647	**	3.156	0.626	***	3.023	0.676	***
$\mathtt{DIA}_t$	-0.797	0.659		-0.851	0.985		0.355	0.687		0.148	0.970	
$1[\mathtt{CVD}_{t-1} = 1]$	0.014	0.284		-0.016	0.313		-0.489	0.388		-0.557	0.400	
$1[\mathtt{CVD}_{t-1} > 1]$	0.024	0.434		0.047	0.475		-1.014	0.735		-1.017	0.778	
$\mathtt{CAN}_{t-1}$	0.498	0.653		0.460	0.727		1.269	0.563	**	1.299	0.667	*
$\mathtt{DIA}_{t-1}$	-0.332	0.631		-0.217	0.570		0.679	0.684		0.874	0.669	
$\mathtt{E\_CVD}_{t-1}$	0.460	0.225	**	0.607	0.236	***	-0.137	0.299		-0.107	0.358	
$\mathtt{N\_CVD}_{t-1}$	0.122	0.095		0.100	0.100		-0.099	0.146		-0.098	0.181	
$\mathtt{E\_CAN}_{t-1}$	-0.086	0.286		-0.149	0.310		1.639	0.257	***	1.661	0.294	***
$\mathtt{E\_DIA}_{t-1}$	0.256	0.230		0.397	0.267		-0.180	0.300		0.004	0.353	
$\mathtt{BMI}_t$	0.047	0.024	**	0.060	0.041		0.006	0.026		0.075	0.049	
$\mathtt{SBP}_t$	0.005	0.004		0.005	0.005		0.002	0.005		0.002	0.006	
$\mathtt{DBP}_t$	-0.003	0.008		-0.002	0.009		-0.009	0.010		-0.009	0.010	
$\mathtt{CHO}_t$	0.010	0.003	***	0.011	0.003	***	0.002	0.003		0.002	0.003	
BMI, SBP, DBP missing	1.399	0.766	*	1.704	1.147		-0.578	0.877		1.072	1.380	
CHO missing	1.677	0.581	***	1.849	0.594	***	0.218	0.650		0.174	0.668	
$\mathtt{ART}_t$	-0.334	0.153	**	-0.348	0.161	**	-0.393	0.174	**	-0.404	0.190	**
Constant	-8.689	2.947	***	-12.363	2.463	***	-9.692	3.530	***	-14.383	1.919	***

Note: Specifications also include controls for age, education, ancestry, origin, cohort, and year trends.
Standard errors are in parentheses. *** indicates joint significance at the 1% level; ** 5% level; * 10% level.

Table A5: Selected Parameter Estimates: Cardiovascular Disease Event in period

Variable	CVD=1 (relative to none)						CVD>1 (relative to none)					
	Without UH			With UH			Without UH			With UH		
	Estimate	Std. error		Estimate	Std. error		Estimate	Std. error		Estimate	Std. error	
Smoker in $t-1$, $\mathtt{S}_{t-1}$	0.184	0.251		0.279	0.275		0.007	0.535		0.227	0.603	
Years of cessation, $\mathtt{C}_t$	-0.004	0.011		-0.002	0.013		0.015	0.024		0.027	0.026	
$\mathtt{C}_t^2/100$	0.018	0.028		0.013	0.030		-0.059	0.065		-0.088	0.066	
Years of duration, $\mathtt{D}_t$	0.035	0.017	**	0.032	0.017	*	0.081	0.035	**	0.081	0.036	**
$\mathtt{D}_t^2/100$	-0.067	0.030	**	-0.064	0.030	**	-0.139	0.058	**	-0.145	0.057	**
Years of experience, $\mathtt{E}_t$	0.002	0.010		0.005	0.012		-0.026	0.018		-0.036	0.019	*
$\mathtt{E}_t^2/100$	-0.001	0.020		-0.005	0.024		0.048	0.037		0.065	0.037	*
$\mathtt{1[CVD}_{t-1} = 1]$	0.441	0.128	***	0.474	0.149	***	0.540	0.231	**	0.664	0.251	***
$\mathtt{1[CVD}_{t-1} > 1]$	0.694	0.205	***	0.744	0.224	***	0.857	0.308	***	0.952	0.319	***
$\mathtt{CAN}_{t-1}$	-0.173	0.322		-0.164	0.358		-0.144	0.664		-0.044	0.662	
$\mathtt{DIA}_{t-1}$	0.018	0.274		-0.076	0.284		-1.189	0.748		-1.432	0.727	**
$\mathtt{E\_CVD}_{t-1}$	0.529	0.130	***	0.499	0.143	***	0.164	0.236		0.114	0.250	
$\mathtt{N\_CVD}_{t-1}$	0.174	0.049	***	0.177	0.059	***	0.345	0.068	***	0.383	0.065	***
$\mathtt{E\_CAN}_{t-1}$	0.349	0.179	*	0.323	0.188	*	0.161	0.360		0.076	0.376	
$\mathtt{E\_DIA}_{t-1}$	0.376	0.139	***	0.324	0.166	*	0.869	0.223	***	0.881	0.247	***
$\mathtt{E\_CVD}_{t-1} * \mathtt{S}_{t-1}$	-0.269	0.164		-0.050	0.348		-0.265	0.303		-0.082	0.664	
$\mathtt{E\_CAN}_{t-1} * \mathtt{S}_{t-1}$	-0.025	0.347		0.817	0.281	***	-0.004	0.694		-0.629	0.792	
$\mathtt{E\_DIA}_{t-1} * \mathtt{S}_{t-1}$	0.755	0.258	***	-0.269	0.185		-0.734	0.651		-0.321	0.325	
$\mathtt{BMI}_{t-1}$	-0.064	0.078		0.026	0.108		0.247	0.201		0.379	0.112	***
$\mathtt{BMI}_{t-1}^2/100$	0.170	0.135		-0.077	0.181		-0.361	0.358		-0.711	0.247	***
$\mathtt{SBP}_{t-1}$	0.009	0.014		0.007	0.015		0.030	0.024		0.027	0.024	
$\mathtt{SBP}_{t-1}^2/100$	0.000	0.004		0.001	0.005		-0.005	0.008		-0.004	0.008	
$\mathtt{DBP}_{t-1}$	-0.036	0.024		-0.034	0.025		-0.042	0.042		-0.040	0.043	
$\mathtt{DBP}_{t-1}^2/100$	0.022	0.014		0.021	0.015		0.031	0.024		0.031	0.025	
$\mathtt{CHO}_{t-1}$	0.002	0.006		0.005	0.005		-0.005	0.007		0.001	0.006	
$\mathtt{CHO}_{t-1}^2/100$	0.000	0.001		0.000	0.001		0.002	0.001		0.001	0.001	
BMI, SBP, DBP missing	-0.997	1.260		-0.342	1.699		4.386	2.913		5.352	1.471	***
CHO missing	0.721	0.670		1.105	0.623	*	-0.208	0.940		0.472	0.838	
$\mathtt{ART}_{t-1}$	0.124	0.078		0.108	0.081		-0.226	0.151		-0.246	0.166	
Constant	-11.303	1.876	***	-12.022	2.261	***	-17.530	3.843	***	-17.436	1.311	***

Note: Specifications also include controls for age, education, ancestry, origin, cohort, and year trends.
Standard errors are in parentheses. *** indicates joint significance at the 1% level; ** 5% level; * 10% level.

Table A6: Selected Parameter Estimates: Cancer Diagnosis and Diabetes Diagnosis in period

Variable	Cancer diagnosis | No previous diagnosis						Diabetes diagnosis | No previous diagnosis					
	Without UH			With UH			Without UH			With UH		
	Estimate	Std. error		Estimate	Std. error		Estimate	Std. error		Estimate	Std. error	
Smoker in $t-1$, $\texttt{S}_{t-1}$	-0.001	0.004		0.387	0.433		0.543	0.333		0.197	0.362	
Years of cessation, $\texttt{C}_t$	0.000	0.000		0.009	0.006		-0.007	0.008		-0.012	0.008	
Years of duration, $\texttt{D}_t$	0.000	0.000		0.004	0.009		-0.021	0.009	**	-0.024	0.010	**
Years of experience, $\texttt{E}_t$	0.000	0.000		0.001	0.004		-0.005	0.005		0.003	0.006	
$\texttt{E\_CVD}_{t-1}$	0.003	0.004		0.197	0.247		0.063	0.259		-0.110	0.277	
$\texttt{N\_CVD}_{t-1}$	-0.002	0.002		-0.122	0.125		0.341	0.111	***	0.355	0.115	***
$\texttt{E\_CAN}_{t-1}$							-0.991	0.519	*	-1.132	0.522	**
$\texttt{E\_DIA}_{t-1}$	-0.005	0.004		-0.218	0.246							
$\texttt{BMI}_{t-1}$	0.000	0.000		-0.004	0.035		0.143	0.018	***	-0.018	0.060	
$\texttt{SBP}_{t-1}$	0.000	0.000		0.001	0.004		0.009	0.004	**	0.009	0.004	**
$\texttt{DBP}_{t-1}$	0.000	0.000		-0.006	0.008		-0.012	0.008		-0.015	0.008	*
$\texttt{CHO}_{t-1}$	0.000	0.000	***	-0.006	0.002	***	-0.002	0.002		-0.002	0.002	
BMI, SBP, DBP missing	-0.015	0.009		-0.787	1.075		2.179	0.973	**	-2.255	1.826	
CHO missing	-0.005	0.006		-0.992	0.472	**	-0.028	0.490		-0.078	0.532	
$\texttt{ART}_{t-1}$	0.001	0.002		0.083	0.134		0.043	0.155		0.012	0.163	
Constant	1.014	0.021	***	-20.791	3.585	***	-22.242	2.882	***	-17.731	3.854	***

Note: Specifications also include controls for age, education, ancestry, origin, cohort, and year trends.
Standard errors are in parentheses. ** indicates joint significance at the 5% level; * 10% level.

Table A7: Selected Parameter Estimates: Body Mass Index in period

	Without UH			With UH		
Variable	Estimate	Std. error		Estimate	Std. error	
Smoker in $t-1$, $\mathtt{S}_{t-1}$	0.005	0.005		0.001	0.005	
Years of cessation, $\mathtt{C}_t$	0.000	0.000		-0.001	0.000	**
$\mathtt{C}_t^2/100$	0.001	0.000	*	0.001	0.001	**
Years of duration, $\mathtt{D}_t$	-0.001	0.000	***	-0.001	0.000	***
$\mathtt{D}_t^2/100$	0.002	0.001	**	0.002	0.001	***
Years of experience, $\mathtt{E}_t$	0.000	0.000	*	0.001	0.000	***
$\mathtt{E}_t^2/100$	-0.001	0.001	**	-0.002	0.001	***
$\mathtt{1[CVD}_{t-1} = 1\mathtt{]}$	0.004	0.005		-0.001	0.005	
$\mathtt{1[CVD}_{t-1} > 1\mathtt{]}$	0.017	0.009	*	0.014	0.009	
$\mathtt{CAN}_{t-1}$	0.020	0.011	*	0.007	0.011	
$\mathtt{DIA}_{t-1}$	-0.018	0.010	*	0.004	0.012	
$\mathtt{E\_CVD}_{t-1}$	0.003	0.005		0.005	0.006	
$\mathtt{N\_CVD}_{t-1}$	-0.003	0.002		-0.006	0.004	
$\mathtt{E\_CAN}_{t-1}$	-0.009	0.007		-0.001	0.007	
$\mathtt{E\_DIA}_{t-1}$	0.000	0.005		-0.021	0.007	***
$\mathtt{E\_CVD}_{t-1} * \mathtt{S}_{t-1}$	-0.003	0.006		-0.019	0.016	
$\mathtt{E\_CAN}_{t-1} * \mathtt{S}_{t-1}$	-0.029	0.013	**	-0.006	0.013	
$\mathtt{E\_DIA}_{t-1} * \mathtt{S}_{t-1}$	-0.010	0.011		0.002	0.006	
$\mathtt{BMI}_{t-1}$	0.102	0.002	***	0.110	0.011	***
$\mathtt{BMI}_{t-1}^2/100$	-0.016	0.004	***	-0.044	0.022	**
$\mathtt{SBP}_{t-1}$	-0.001	0.000		-0.001	0.000	
$\mathtt{SBP}_{t-1}^2/100$	0.000	0.000		0.000	0.000	
$\mathtt{DBP}_{t-1}$	0.002	0.001	**	0.002	0.001	**
$\mathtt{DBP}_{t-1}^2/100$	-0.001	0.000	**	-0.001	0.000	**
$\mathtt{CHO}_{t-1}$	0.000	0.000		0.000	0.000	
$\mathtt{CHO}_{t-1}^2/100$	0.000	0.000		0.000	0.000	
BMI, SBP, DBP missing	2.682	0.043	***	2.637	0.190	***
CHO missing	-0.014	0.019		0.000	0.020	
$\mathtt{ART}_{t-1}$	0.000	0.002		-0.001	0.002	
Constant	-0.003	0.048		-0.734	0.150	***

Note: Specifications also include controls for age, education, ancestry, origin, cohort, and year trends. Standard errors are in parentheses. *** indicates joint significance at the 1% level; ** 5% level; * 10% level.

B The Role of Unobserved Heterogeneity

While the rich observed health heterogeneity of individuals plays an important role in explaining the mortality rates of individuals with different lifetime smoking patterns by addressing concern over confounding (using observable data), individual UH also plays an important role. Its main function is to capture the correlation, through unobservables, among the modeled behaviors and outcomes that would otherwise bias estimated impacts of the smoking and health histories. Our jointly estimated model allows for UH that is likely to be common across a lifespan (such as genetics, risk-aversion, time preference or self-esteem, for example) as well as differences that may vary over time (such as unobserved stress or health, for example). We model these two types of UH using discretized distributions characterized by mass point vectors that describe the impact of each type of heterogeneity on the outcomes of interest. Appendix Table A6 displays the estimated coefficients and standard errors that capture the distributions. Estimated probability weights of each mass point are listed in the last column.

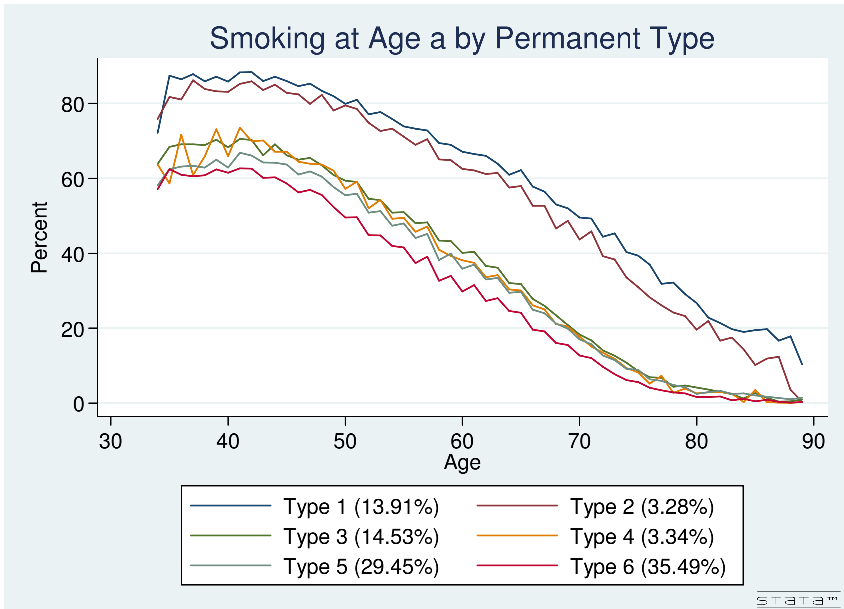
Recall that we replicate, R times, the exogenous characteristics of all individuals in our estimation sample, N. For each replication, we use the estimated heterogeneity distributions to draw a permanent "type" that is common for that replicated individual across all time periods and a second "type" that varies each time period. Using the estimated model to simulate lifetime behavior and outcomes (from an individual's observed initial age through age 100), we show that lifetime smoking probabilities differ by these unobserved types. While it is difficult to depict the differences associated with the time-varying UH, we can condition on (simulated) permanent UH type and plot the resulting smoking rates by age (Figure B1). We order the "types" by the simulated smoking probability at age 40 (i.e., highest to lowest).[1]

The distribution of permanent UH suggests that about 17 percent of the sample (types 1 and 2) are as much as 20 percentage points more likely to smoke at any given age than the other 83 percent. The figure correctly shows that this time-invariant unobserved determinant of smoking shifts smoking probabilities uniformly, unconditional on smoking and health histories, at each age.[2] Additionally, the model allows for behavior and outcome shifters each two-year period based on a draw from the distribution capturing time-varying unobservables. A likelihood ratio test comparing the goodness of fit of the nested models with and without UH suggests that the model with UH fits

[1]The figures in parentheses are the simulated proportions of each permanent UH type. Appendix Table A2 displays the estimated proportions.

[2]The permanent UH is linearly added to the operand of the linear (OLS) and non-linear (LOGIT or MLOGIT) operators. For the latter, this does not translate into an intercept shift but also depends on the level of the product of observed variables and their coefficients.

Figure B1: Smoking Probabilities by Permanent Unobserved Heterogeneity



Note: The colored lines indicate each heterogeneous type, with probabilities in parentheses.

significantly better.

Using the same ordering based on smoking propensity, we also report the mortality outcomes for each permanent UH type in Table B8. We see that those individuals with an UH type that makes them more likely to smoke also experience the shortest lifespan with an average age of death of 71.8. They also have the highest proportion of cancer deaths. Types 5 and 6, who are the least likely to smoke, have the highest mean (at 75.1 and 73.9, respectively) and percentile ages of death. Type 3 individuals are much more likely to have a CVD-related death than any of the other types. Finally, note that type 2 captures individuals with high smoking rates yet longer than average expected lifetimes.

These conditional (on UH type) death distributions reflect 1.) differences in lifespan due to unobserved permanent factors (like genetics or time preferences) as well as 2.) differences in smoking behavior (as illustrated in Figure B1). While we cannot say exactly what the UH captures, knowing these different smoking and mortality patterns by type gives us insight into both the estimation results and policy recommendation. First, the (unconditional on type) death distribution would be different if UH were ignored. (We see this in the biased coefficients of the model without UH.) It is not simply that inclusion of UH improves precision by reducing important selection, endogeneity, and measurement error biases, but it allows different lifetime smoking patterns which

Table B8: Age and Cause of Death by Permanent Unobserved Heterogeneity

Permanent UH Type	Simulated Percent	Mean	Age of death distribution (percentile) 10th	25th	50th	75th	90th	Cause of death CVD	Cancer	Other
1	13.9	71.8	57	65	72	79	86	28.1	29.9	42.0
2	3.3	73.4	59	66	74	81	87	43.0	24.3	32.7
3	14.5	72.5	57	65	73	80	86	56.7	15.6	27.6
4	3.3	72.9	58	66	74	80	87	44.6	13.4	42.0
5	29.5	75.1	60	68	76	83	89	35.2	28.4	36.3
6	35.5	73.9	59	67	75	82	88	46.4	25.1	28.5

in turn have non-linear feedback effects (on both health and subsequent smoking) via the dynamic system of equations. Second, policy evaluation should be more sensitive to distributional issues knowing there is heterogeneity in the population in terms of smoking initiation rates, quit rates, relapse rates, and mortality rates. We find, for example, that some individuals are more predisposed to smoke, but only some of these have shorter expected life spans.

C Historical Data

In this appendix we discuss the cigarette advertising and price data used to construct important cigarette market variables over the 19th and 20th centuries. For each variable (i.e., average advertising expenditure and average price), we first provide a justification of its use as an instrument for cigarette smoking and then discuss details associated with construction of the advertising expenditure and price time series. Some of the discussion focuses on the state of Massachusetts, since the FHS data are from the town of Framingham.

C.1 Advertising and Cigarette Consumption

We use industry-wide advertising spending to instrument for cigarette initiation during the years 1895-1939 and for smoking behavior over the years 1950-1996. There are two key conditions needed for identification: smoking initiation must be responsive to advertising and trends in advertising spending must be aimed at market expansion rather than brand switching. We deal with each of these issues in turn making reference to the literature.

The first condition is that firm advertising impacts smoking behavior. There are several channels by which this could occur. For example, during the pre-World War II period cigarette advertising increased social acceptability of smoking (particularly for women for whom it had been considered taboo), promoted the image of smokers as independent and glamorous, and listed health benefits such as hunger suppression (Brandt, 2007). There is empirical evidence linking advertising to youth smoking initiation (and almost all smokers in our data begin smoking by the teen years). In their survey of the economics of smoking, Chaloupka and Warner (2000) note that advertising has a positive and significant impact on teen smoking initiation in studies using individual-level data. Borden (1942), Tennant (1950), and Pierce and Gilpin (1977) note that cigarette advertising during our study period was primarily targeted to groups, such as female youths, which had not smoked previously, and that these groups experienced greater increases in smoking initiation rates at those times. Telser (1962) provides estimates which show that firm-level cigarette advertising increased overall smoking levels during 1925-1939. (Participants in the FHS original cohort were born between 1886 and 1918 and were in their teens between 1900 and 1932.)

The second condition deals with the intentions underlying the decision to advertise. Advertising can both increase demand (the focus here) and also lead to brand switching (which might not increase smoking initiation). The main threat to identification would be if the latter effect predominates or if it changes in importance over time. In the period through 1912 this is not a major concern since cigarettes and all other forms of tobacco were sold by a monopolist, the American Tobacco Company, also referred to as the Tobacco Trust. Since there was limited variation of prices and market segmentation at this time, there would be little advertising related to brand-names. In the post-Trust period, the industry largely moved in lock-step. The main cigarette manufacturers were convicted in 1941 of violating the Sherman Act, both Section 1 (restraint of trade) and Section 2 (monopolization). For example, the wholesale prices of all leading brands were identical from 1928 to 1946 and virtually identical prior to that with manufacturers changing prices within days of one another. In such an environment of likely tacit collusion, an important feature of advertising was to increase smoking overall as much as to promote individual brands. Echoing the goals of smoking advertising in the last paragraph, George Washington Hill, president of American Tobacco, testified at the 1941 anti-trust trial: "The impetus of those great advertising campaigns not only built this for ourselves, but built the cigarette business as well ... You don't benefit yourself most, I mean, altogether ... you help the whole industry if you do a good job" (p. 137, Tennant (1950)). There were two periods of relatively strong competition: the period immediately following the dissolution of the Tobacco Trust and the 1930s with a short-lived rise of economy cigarettes. Counter to what would be expected under brand-switching, advertising moved erratically in the first period and decreased during the latter period (see Figure C2). Also Telser (1962) shows that advertising at the brand-level was market expanding and that brand-stealing effects are small in magnitude during the 1920s and 1930s.

C.2 Construction of Advertising Expenditures Time Series

Annual nominal advertising spending on cigarettes, exclusive of free goods (e.g., give-aways of cigarettes) and other non-traditional advertising, comes from a variety of sources. Spending for the years 1893-1913 are from United States Department of Commerce (1915), which lists advertising spending per cigarette and also total cigarette sales. These totals include the entire cigarette business of the American Tobacco Company (the Tobacco Trust), exclusive of exports and foreign manufacturing business as well as Turkish cigarettes. Spending for the years 1893-1910 and the spending by the Trust's successor companies for 1912-1913 are government assembled totals completed in the wake of the the Supreme Court's break-up of the Trust in 1911. (No data are available for 1911 and spending is interpolated for that year).

Advertising expenditure for the years 1914-1928 are based on Nicholls (1951). Nicholls lists R.J. Reynolds Tobacco Company's cost of advertising, exclusive of gratis goods. Largely due to its Camel brand, over most of this period Reynolds was the leading cigarette producer and it annually sold between a third and almost half of all cigarettes. The aggregate spending on cigarettes is approximated by dividing this total by Reynolds' share of total cigarettes and multiplying this by the share of cigarettes among all tobacco products.

Expenditures for the years 1929-1949 are drawn directly from Fujii (1980). He uses a variety of primary and secondary sources to create an index of corporate cigarette advertising. Expenditures for the years 1950-1962 come from Schneider *et al.* (1981). They credit their series to a telephone interview with Television Bureau of Advertising, Inc. Both of these sources list real spending.

Advertising expenditure values from 1963 onwards are from the Federal Trade Commission (2013). Starting in this year the FTC began collecting information on cigarette spending across a variety of media including TV, radio, print and others. In all cases we net out totals related to price promotion, promotional allowances, and other specific channels which were added in later years.

We consider a variety of robustness checks to ensure that differences between these sources do not create artificial variation. Several of the series overlap and the patterns discussed below remain when we use values for the other series. These overlaps include United States Department of Commerce (1915) and Nicholls (1951) which both include data for 1913; Nicholls (1951) and Fujii (1980) which both include data for 1929-1949 (Nicholls' data are for Reynolds' total traceable advertising expenditures over 1939-1949); Fujii (1980) and Schneider *et al.* (1981) which both include data for 1950-1973; Schneider *et al.* (1981) and Federal Trade Commission (2013) which both include data for 1963-1978. A second check was to include additional company's advertising spending during 1914-1928. Nicholls (1951) includes data for American Tobacco for 1925-1928, and aggregate spending on cigarettes is not sharply changed when the same approach described earlier is used. Advertising costs for American for 1929-1939 and Ligett & Myers for 1935-1939 is also available and is used to compare Nicholls (1951) and Fujii (1980) in the first robustness check. Finally, as a robustness check we compare these assembled values to other data sources. Borden (1942) includes various measure of total advertising over 1929-1939 for Camels and for all brands that are comparable to the values in Nicholls (1951) and Fujii (1980). Tennant (1950) presents several series that are identical or follow a similar pattern as United States Department of Commerce (1915) and Nicholls (1951).

Additionally, our assembled time series cigarette advertising expenditure data are converted into per capita terms using the United States population figures from United States Census Bureau (2000), United States Census Bureau (2011), and United States Census Bureau (2012). In all cases, terms are converted to year 2000 dollars using Bureau of Labor Statistics (2013) for 1913 onwards and Sahr (2013) for earlier years.

Figures A2a and A2b graph the resulting time series in levels and per capita terms, respectively. A few common features are present in both series. There is a run-up in advertising after the break-up of the Tobacco Trust (i.e., annual spending tripled within three years) as well as a reduction in advertising during each of the World Wars and The Depression. There was another steady increase in the post-war period (i.e., annual spending went up almost eight fold from 1945 to 1967), and then fell starting in 1967 with the FCC's ruling in that year that the fairness doctrine required anti-smoking ads on TV and radio and the 1971 ban in ads on those media. Advertising again climbed in the mid-1970s to the mid-1980s, after which it steadily declined.

C.3 Prices and Cigarette Consumption

Standard price theory suggests that own prices should impact cigarette consumption. The important distinction is that, as highlighted in the theoretical foundation section, smoking decisions are inherently dynamic: smoking impacts future health and future utility (via preferences that capture addiction). The rational addiction literature shows forward-looking agents alter their smoking behavior based on both current and expected

Figure C2: Annual Real Aggregate Cigarette Advertising Expenditure

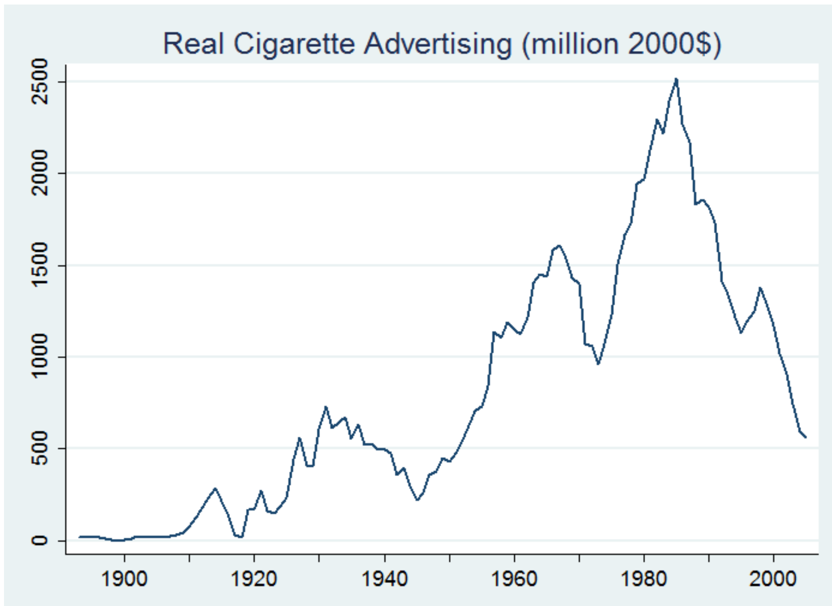


Figure C2a.

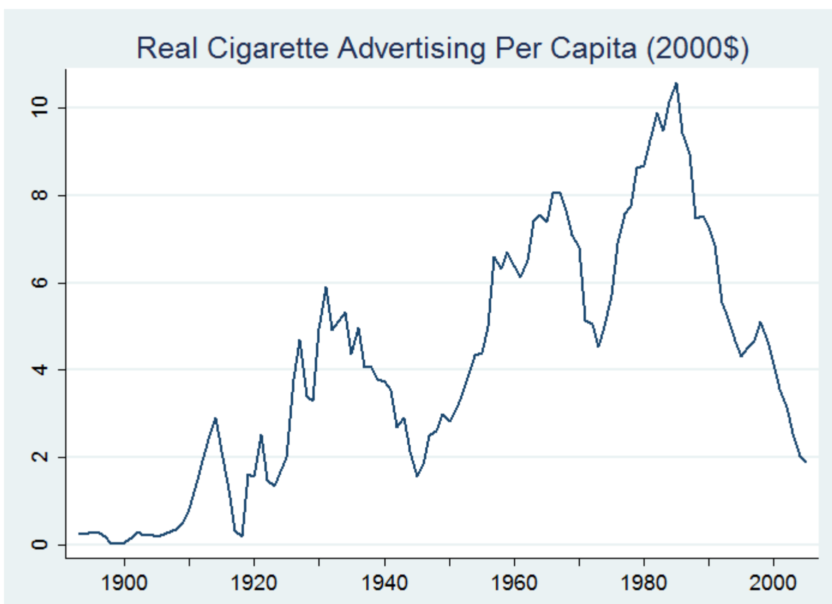


Figure C2b.

future cigarette prices (Becker and Murphy (1988); Gruber and Koszegi (2001)). In the analysis below we focus on contemporaneous prices.

There is a large literature documenting economically and statistically significant effects of prices (Chaloupka and Warner, 2000). For youth the price elasticity is -0.5 to -1.5, reflecting the responsiveness of smoking initiation (which determines the initial conditions in our estimation framework). For adults the price elasticity is -0.2 to -0.5, reflecting responsiveness of quits, relapse, and conditional intensity (captured by the contemporaneous smoking equations in our estimation framework). These results come mainly from analyses of recent data. It is noted that our data involve an earlier period with different technologies (e.g., non-filtered cigarettes were not introduced until the 1950s) and social mores with regards to smoking.

For identification purposes, we argue that our price data are exogenous. There are two main threats to this argument. The first issue is that firms might set prices strategically in response to consumer demand, and such reverse causality will lead to bias. While tobacco companies have some market power, it is important to remember that other factors shape consumer prices. The additional factors are federal and state excise taxes on cigarettes, state sales taxes, and state-imposed price regulations. These factors change for reasons that are largely exogenous to cigarette demand: the introduction and subsequent increases in Massachusetts cigarette excise taxes through at least 1950 were instituted as emergency measures related to budget shortfalls; Massachusetts implemented a minimum cigarette price law in 1945 and over time continued to tinker with its formula (e.g., the mark-up rate, whether the state excise tax is included, differential treatment of less expensive brands, differential prices for non-chain stores). We show in the next subsection that taxes comprise on average half of the consumer price, and this share varies substantially over time. The minimum price rule makes it difficult for cigarette manufacturers to set final consumer prices; while the minimum price is based on wholesale prices, the specific formula continually changes (Annotated Laws of Massachusetts, 2007).

The second concern is that consumers buy cigarettes in other states that have lower prices. (There is far more price variation between- rather than within-states due to the role of state taxes and regulations.) Merriman (2010) shows that large tax differences lead to substantial cross-border shopping particularly over short distances. If this is true then observed prices at the state level would not reflect the true price that consumers face, and the extent of the mismeasurement would vary based on the size of the price differential. In the case of Framingham Massachusetts, the nearby states are Connecticut, New Hampshire, Rhode Island, and Vermont. For the years 1955-2011 Connecticut and Rhode Island have comparable prices as Massachusetts (Orzechowski and Walker, 2011), so cross-border shopping is not an issue. New Hampshire and Vermont both have lower prices over at least a portion of this period. Still, it is unlikely that cross-border shopping was a big issue over much of our sample, due to the relatively high cost of inter-state transportation until at least the 1950s and 1960s. There is also indirect evidence against cross-state traffic: the price differential grows over time, so if there is more inter-state purchases then sales between the states should become more lopsided over time. Per capita sales in New Hampshire and Vermont increased relative to Massachusetts when the price differential first started to become significant (i.e., the 1960s for New Hampshire and 1970s for Vermont). But, in the next decade as the price differential continued to grow, sales stopped shifting to the other states or even shifted back to Massachusetts.

C.4 Construction of Prices Time Series

This subsection discusses construction of the cigarette price time series for the period 1901-2011. The series is for Massachusetts, the smallest area for which we could collect prices. (We argue this is reasonable given the relatively small size of Massachusetts). In all cases prices are per one thousand cigarettes (an industry standard), include all state and federal taxes, and are converted to year 2000 dollars using Bureau of Labor Statistics (2013) for 1913-2011 and Sahr (2013) for earlier years. We generate various summary statistics including the unweighted-average, minimum across brands, and these values exclusive of generics/economy brands. (Generic/economy brands were prominent for three periods: 1901-1910, 1931-1950 and 1991-2011.)

Prices through 1950 come from a variety of sources that list price at the brand-level and at the national level. (Taxes and other Massachusetts-specific factors are discussed below.) Prices for 1901-1911 are from United States Department of Commerce (1915). Prices are available annually for the principal brands of the American Tobacco Company (the Tobacco Trust), exclusive of Turkish cigarettes. The principal brands comprised a majority of sales, with one brand accounting for three-fourths of domestic sales in the beginning of the period. These are government assembled totals completed in the wake of the the Supreme Court's break-up of the Trust in 1911. (No data are available for 1911 and prices are interpolated for that year). There is also data from this source for 1912-1913 for the Trust's successor companies, which is combined with the sources listed below.

Prices for 1912-1950 are based on Nicholls (1951). This source lists the date and level of all list price changes for the main brands. The market was quite concentrated during this period and just the three leading brands (Lucky Strikes, Camel, Chesterfield) accounted for almost all sales until the 1930s and two-thirds of sales through 1950 (Maxwell, various years; Nicholls, 1951). The data include prices for all brands, including economy/generics, which account for virtually all domestic cigarette sales. The price level and date of change were checked against Tennant (1950) and there are only a few and relatively minor discrepancies. Massachusetts cigarette excise taxes (a per unit tax) were first introduced 11 August 1939 and are added onto these prices. (Federal excise taxes are included in the list price.)

No data are available for 1951-1954 and interpolation is used. The only change in taxes during this time was a one cent per pack increase in the federal excise tax on cigarettes on 01 November 1951.

Data for 1955-2011 are from Orzechowski and Walker (2011), which lists average retail price by state. Prices are the market share-weighted average of prices of all brands based on surveyed consumer prices in Massachusetts for fiscal years ending 30 June. A separate series, which includes generics brands, is included starting in 1991. Prices include state and federal cigarette excise taxes but do not include sales taxes. Prices were adjusted to reflect the sales tax after Massachusetts removed the exemption for cigarettes on 01 July 1988.

Figure C3 graphs the resulting price time series. This figure uses the minimum price across brands and omits generics. In our main analysis we focus on the average price exclusive of generics. The omission of generics is relatively innocuous since for the years of our main model (1952-1996) the different summary statistics (of prices with and without generics) are virtually identical (i.e., generic/economy brands were prominent for three periods — 1901-1910, 1931-1950, and after 1990 — which only overlap with the very end of our observation period). Figure A3a. shows prices over the century. While they appear relatively stable, note the wide-range of the vertical axis. (In fact, the post-2000 period is omitted from the graph since prices continue to rise and this would further obscure the variation.) Prices collapse almost in half after the dissolution of the Tobacco Trust in 1911. Prices then rise and fall repeatedly during the 1920s , 1930s, and 1940s. Prices then steadily rise for the next two decades, dip again, and finally increase sharply in the 1990s. Figure A3b. shows that, on average, half of this price is composed of taxes (i.e., state and federal excise taxes on cigarettes as well as state sales tax). This information is helpful for identification since the tax share is one of the main sources of variation in prices, and it oscillates for non-demand reasons (e.g., taxes rise during World War 2 when fiscal demands necessitated the creation of the state excise tax, initially an emergency measure, and increases in the federal tax).

Figure C3: Annual Real Cigarette Prices and Taxes

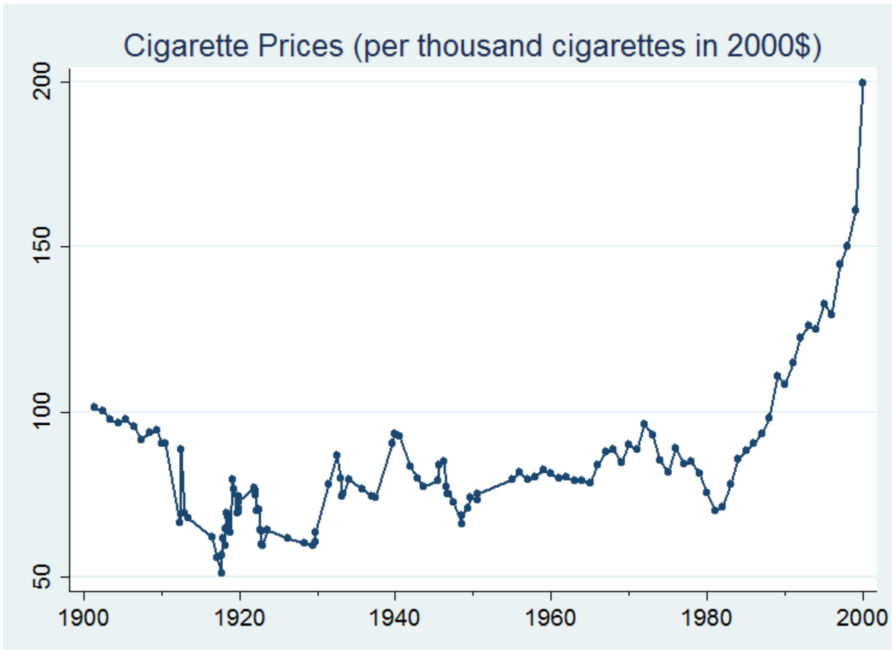


Figure C3a.

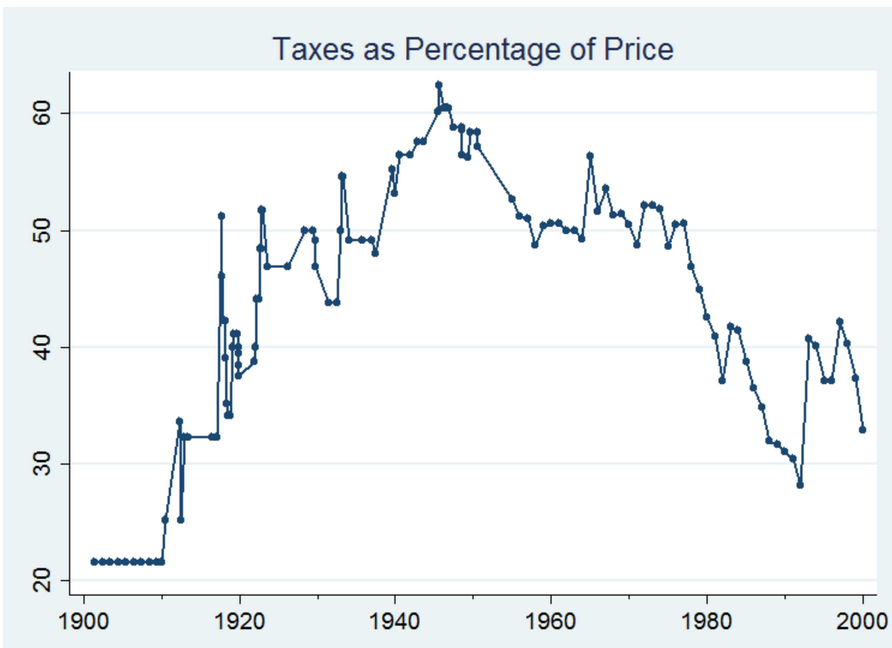


Figure C3b.

References

Annotated Laws of Massachusetts (2007). *PART I ADMINISTRATION OF THE GOVERNMENT. TITLE IX TAXATION. Chapter 64C Cigarette Excise.* Tech. rep., General Court of the Commonwealth of Massachusetts, https://malegislature.gov/Laws/GeneralLaws/PartI/TitleIX/Chapter64C.

Becker, G. S. and Murphy, K. (1988). A Theory of Rational Addiction. *Journal of Political Economy*, **96** (4), 675–700.

Borden, N. (1942). *The Economic Effects of Advertising.* Richard D. Irwin, Inc.: Chicago.

Brandt, A. (2007). *The Cigarette Century.* Basic Books: New York.

Bureau of Labor Statistics (2013). *Consumer Price Index, All Urban Consumers - (CPI-U), U.S. city average, All items.* Tech. rep., Bureau of Labor Statistics, http://www.bls.gov/cpi/.

Chaloupka, F. J. and Warner, K. E. (2000). The Economics of Smoking. *Handbook of Health Economics*, **1B**, 1539–1627.

Federal Trade Commission (2013). *Cigarette Report for 2011.* Tech. rep., Federal Trade Commission, http://www.ftc.gov/reports/federaltradecommission-cigarettereport2011.

Fujii, E. (1980). The Demand for Cigarettes: Further Empirical Evidence and Its Implications for Public Policy. *Applied Economics*, **12**, 479–489.

Gruber, J. and Koszegi, B. (2001). Is Addiction "Rational"? Theory and Evidence. *The Quarterly Journal of Economics*, **116** (4).

Maxwell, J. C., Jr. (various years). *Historical Sales Trends in the Cigarette Industry.* Tech. rep., Wheat, First Securities, Inc.

Merriman, D. (2010). The micro-geography of tax avoidance: Evidence from littered cigarette packs in chicago. *American Economic Journal: Economic Policy*, **2**, 61–84.

Nicholls, W. (1951). *Price Policies in the Cigarette Industry: A Study of 'Concerted Action' and Its Social Control, 1911-50.* Nashville: Vanderbilt University Press.

ORZECHOWSKI and WALKER (2011). *Tax Burden on Tobacco: Historical Compilation, vol 46*. Tech. rep., Orzechowski and Walker Consulting Firm, http://www.taxadmin.org/fta/tobacco/papers/Tax_Burden_2011.pdf.

PIERCE, J. and GILPIN, E. (1977). A Historical Analysis of Tobacco Marketing and the Uptake of Smoking by Youth in the United States: 1890–1977. *Health Psychology*, **14**, 1–9.

SAHR, R. (2013). *Inflation Conversion Factors for Years 1774 to estimated 2018*. Tech. rep., Oregon State University, http://oregonstate.edu/cla/polisci/faculty-research/sahr/infcf17742007.pdf1.

SCHNEIDER, L., KLEIN, B. and MURPHY, K. (1981). Governmental Regulation of Cigarette Health Information. *Journal of Law and Economics*, **23**, 575–612.

TELSER, L. (1962). Advertising and Cigarettes. *Journal of Political Economy*, **70**, 471–499.

TENNANT, R. (1950). *The American Cigarette Industry: A Study in Economic Analysis and Public Policy.*. New Haven, Yale University Press.

UNITED STATES CENSUS BUREAU (2000). *Historical National Population Estimates: July 1, 1900 to July 1, 1999* . Tech. rep., United States Census Bureau, http://www.census.gov/popest/data/national/totals/pre-1980/tables/popclockest.txt.

UNITED STATES CENSUS BUREAU (2011). *Table 1. Intercensal Estimates of the Resident Population by Sex and Age for the United States: April 1, 2000 to July 1, 2010 (US-EST00INT-01)*. Tech. rep., United States Census Bureau, http://www.census.gov/popest/data/intercensal/national/tables/US--ST00INT01.csv.

UNITED STATES CENSUS BUREAU (2012). *Historical Statistics of the United States, Colonial Times to 1970*. Tech. rep., United States Census Bureau.

UNITED STATES DEPARTMENT OF COMMERCE (1915). *Report of the Commissioner of Corporations On the Tobacco Industry, Part III: Prices, Costs and Profits.* Tech. rep., Department of Commerce, Washington: Government Printing Office.