APPENDIX 1. Estimation and Forecast of Smoking Prevalence by Intensity

This Appendix describes the detailed methodology for estimating the effect of per capita California tobacco control funding on the prevalence of current smoking by intensity (light, moderate, and heavy), and to project the prevalence rates of current smoking by intensity in future years.

Estimation of Prevalence by Intensity of Smoking

We used data from the 1984-2000 BRFSS and the CPS-TUS (1992-1993, 1995-1996, 1998-1999, 2001-2003, and 2006-2007) to estimate four dependent smoking prevalence variables: total current smoking prevalence and the proportion of current smokers by intensity (light, moderate, heavy). Each dependent variable is regressed on the difference in per capita tobacco control funding between California and the control states, and the difference in per capita income between California and the control states.[1,2] Prevalence of current smoking was calculated using equivalent definitions for both data sources. One regression was estimated for each dependent variable. It was assumed that the data were nonstationary and contained an autoregressive unit root, or showed high persistence, so a cointegrating regression specification was the most efficient approach for a small sample.[3] The long-run relationships between the prevalence of smoking by intensity were estimated with a static cointegrating regression using irrelevant instrumental variables, which are appropriate for nonstationary data,[4] or data with high persistence.[5]

The residuals of the irrelevant instrumental variables regression were not normally distributed and showed evidence of several influential observations, so the final estimates used median regression, which is more robust to possible outliers than least squares. There were no substantively significant differences between the irrelevant instrumental variables regression results and the median regression results. The predicted proportions from the three independently estimated regressions of light, moderate and heavy smoking prevalence as a proportion of total smoking prevalence summed to within one percentage point of 100 percent.

Definition of control states. The control states were used to model unobservable secular trends that may affect smoking prevalence apart from California tobacco control expenditures. . Thirteen states were chosen as control states because they had continuous data on state specific prevalence in the BRFSS survey since 1984: Idaho, Illinois, Indiana, Minnesota, Montana, North Carolina, Ohio, Rhode Island, South Carolina, Tennessee, Utah, West Virginia, and Wisconsin. Arizona also had continuous data in the BRFSS sicne 1984 but was not selected as a control state because it has a strong tobacco control program. Data on current smoking prevalence by intensity for the control states were not available; therefore total prevalence of smoking using BRFSS survey data was used for control states. As a sensitivity analysis, we also estimated the model using all the other states in the Western Census Region (except Arizona, which has a strong tobacco control program) as controls. Using the Western Census Region as controls, the coefficient of cumulative per capita tobacco control spending was: (1) not statistically significantly different from zero for the moderate smoking prevalence model although with the same sign as when using the 13 control states, and (2) not statistically significantly different from the coefficient estimated using the 13 control states for light and heavy smoking prevalence. Therefore, the conclusion that tobacco control funding reduces heavy smoking and increases light smoking prevalence remains valid.

Use of Different Data Sources. An indicator variable was included as an explanatory variable to model the effect of using different data sources on the dependent variable. Several interaction terms and nonlinear transformations using the indicator variable were used to model the effect of using data from different sources (CPS-TUS, and BRFSS surveys). The results were not sensitive to these variations and including the indicator variable in the regression was the simplest approach.

The specification of the four smoking prevalence regressions is:

$$ps_{CA,i,t} = \beta_0 + \beta_1 source_t + \beta_2 ps_{control,t} + \beta_3 (E_{CA,t} - E_{control,t}) + \beta_4 (y_{comtrol,t} - y_{CA,t}) + \varepsilon_t \qquad \text{(Eq. 1)}$$

where $ps_{CA,i,t}$ = the total current smoking prevalence *i* (*i* = total) and proportion for current smoking by intensity in *i* (= light, moderate or heavy) for California in year *t*

$source_t$ = an indicator variable for source of the dependent variable in year *t*, 0 indicates BRFSS data and 1 indicates CPS-TUS data, for *i* = light, moderate and heavy smoking prevalence.

$ps_{control,t}$ = the prevalence of total current smoking in the control state population in year *t*

$E_{CA,t}, E_{control,t}$ = cumulative real per capita tobacco control funding for California or control state population in year *t*

$y_{control,t}, y_{CA,t}$ = real per capita income for control state population or California in year *t*

ε_t = the regression error in year *t*.

Note that the variable $source_t$ was omitted from the regression for total current smoking prevalence because all data came from BRFSS.

Estimation and Forecast of Future Prevalence by Intensity of Smoking

The Vector Autoregression (VAR) estimation approach was used to forecast the explanatory variables in Equation 1. An automated algorithm, *Autometrics*, was used to determine the specification of the VAR equations.[6,7] *Autometrics* uses a formal 'General to Specific' approach to model specification, using a structured search over hierarchical trees of different possible specifications, and encompassing tests and cut-point significance tests to delete variables that do not belong in the regression. *Autometrics* is designed to minimize the pre-test bias and loss of control of overall significance level that occurs as a result of repeated tests and re-estimation. An automatic model selection algorithm was used to, as far as possible, take human judgment and bias out of the model selection process. *Autometrics* includes formal diagnostics of residuals (serial correlation, normality, homoskedasticity, and influential observations) as well as various stability tests for parameter constancy and for structural breaks. *Autometrics* presents a 'best model' which is chosen either because it is the unique minimal model that encompasses the original unrestricted model, or is chosen from several candidates based on the Schwarz information criterion. The specifications produced by *Autometrics* were checked for consistency with a long-run relationship by re-estimation of a static long-run cointegrating regression using irrelevant instrumental variables.

The models estimated to forecast the explanatory variables are as follows. The reduced form VAR model produced the most reasonable estimate of the per capita personal income variables. The model for control state personal income was a second order autoregressive time series model, and a constant mean was the best estimate for the difference between California and control state per capita income. The short-run dynamic reduced form VAR model estimate was used to forecast control state per capita income. Per capita personal income in control states will increase at about $2035 per year, and the difference between California and control states per capita income will remain constant, so Califonia per capita personal income will also grow at $2035 per year.

The best model as determined by *Autometrics* for control state smoking prevalence was a reduced from VAR equation in which control state prevalence is a function of lagged control state per capita income, proportion of the control state population that is elderly, and lagged per capita control state tobacco control expenditure. The long-run equilibrium solution was used because it followed the trend of historical control state smoking prevalence with less variance than the dynamic predictions calculated directly from the VAR. The selected model forecast that smoking prevalence in control states is reduced by 0.23 percentage points per year. *Autometrics* could not find an acceptable model fit for cumulative control state tobacco control funding. An acceptable model for annual control state funding was found, so forecasts of annual control state funding were used to construct a forecast of cumulative funding. The annual real cumulative expenditures were calculated by simple summing of the estimated annual funding. The best model for annual control state per capita tobacco control funding used past values of annual control state per capita funding and per capita personal income. The long-run equilibrium solution was used for the forecasts, but these were nearly identical to the short-run predictions that used the short-run VAR model. Annual per capita funding for the control states continues at $2.50 per year, so that the cumulative control state per capita funding increases at $2.50 per year.

It was assumed that some current funding of California tobacco control programs (interest income and the Proposition 10 backfill) in addition to the nominal 5 cents per pack provided by Proposition 99 will continue at average relative levels observed between 2004 and 2007. The annual projected total and adult resident California population and proportion of the population that is elderly for control states were interpolated from U.S. Census Bureau projections**.**

The forecast of smoking prevalence by intensity was made by multiplying the forecast of the proportion of current smokers (by intensity) by the forecast of total smoking prevalence. The dependence between the regression errors in the three prevalence proportion equations was not modeled, and therefore the variance of regression error is over estimated. However, the sum of the unrestricted estimates and forecasts of the prevalence proportions was very close to one, so the overestimate of the variance should be relatively small.

References for Appendix 1

1 Lightwood J, Glantz S. Effect of the Arizona tobacco control program on cigarette consumption and healthcare expenditures. *Soc Sci Med*. 2011;**72**(2):166-72.

2 Lightwood JM, Dinno A, Glantz SA. Effect of the California tobacco control program on personal health care expenditures. *PLoS Med*. 2008;**5**(8):1214-1222.

3 Hoover DS, Johansen S, Juselius K. Allowing the data to speak freely: the macroeconometrics of the cointegrated vector autoregression. *Amer Econ Rev* 2008;**98**(2): 251-255.

4 Phillips P. *Optimal estimation of cointegrated systems with irrelevant instruments*. New Haven, CT: Cowles Foundation, Yale University, 2006.

5 Kourogenis K, Panopoulou E, Pittis N. *Irrelevant but highly persistent instruments in stationary regressions with endogenous variables containing near-to-unit roots*. Piraeus, Greece: Department of Banking and Financial Management, University of Piraeus, 2005.

6 Doornik J, Hendry D. *Empirical econometric modelling, PcGive 13*. Vol. I. London, UK: Timberlake Consulting, Ltd., 2009.

7 Doornik J. Autometrics. In Castle J, Shephard N, eds. *The Methodology and Practice of Econometrics*. Oxford, UK: Oxford University Press 2009:88-121.

APPENDIX 2. The Econometric Models for Smoking-Attributable Healthcare Expenditures

This Appendix details the specification of econometric models that analyze the impact of smoking on healthcare expenditures . The models consist of 10 equations which were estimated separately for each of the 6 subgroups stratified by age (18-34, 35-64, 65+) and gender (female, male). Each equation or group of equations is described below in terms of the functional form used, the dependent variable, and the independent variables.

(1) **Propensity of Ever Having Any Smoking-Related Diseases Model** (estimated as a binomial probit function)

Having smoking-related diseases = f_1 (smoking status/intensity (current light, current moderate, current heavy, former), age, race/ethnicity, region, education, marital status)

The smoking-related diseases are those for which incidence is identified in the 2004 US Surgeon General Report as being causally related to cigarette smoking.[1] Included are cardiovascular diseases (ischemic heart disease, other heart disease, cerebrovascular disease, atherosclerosis, aortic aneurysm, other circulatory diseases), respiratory diseases (chronic bronchitis, emphysema, chronic airways obstruction), and cancer (of the urinary bladder; cervix uteri; esophagus; kidney and renal pelvis; larynx; stomach; lung, trachea, and bronchus; lip, oral cavity, and pharynx; pancreas; and acute myeloid leukemia).

(2) Propensity of Having Poor Health Model (measured by health status of excellent, very good, good, fair, or poor; estimated as an ordered probit function)

Health status = f_2 (smoking status/intensity, age, race/ethnicity, region, education, marital status, predicted propensity of ever having smoking-related diseases)

(3) – (6) **Propensity of Having Positive Annual Expenditures Model** (estimated separately for Hospitalization, Ambulatory Care, Prescription Drugs, and Home Health Care; estimated as a binomial probit function)

Having positive expenditures = f_i (smoking status/intensity, age, race/ethnicity, region, education, marital status, health insurance coverage, predicted propensity of having poor health)

(7) – (10) **Logarithmic Level of Annual Expenditures among Individuals with Positive Expenditures Model** (estimated separately for Hospitalization, Ambulatory Care, Prescription Drugs, and Home Health Care; estimated using ordinary least squares)

Logarithm of expenditures = f_j (smoking status/intensity, age, race/ethnicity, region, education, marital status, health insurance coverage, predicted propensity of having poor health)

Note that when taking the anti-log of the predicted values from Equations (7)–(10), a smearing coefficient by smoking status was calculated to correct for the smearing effect bias.

The table below summarizes the age and gender subgroups for which each of the 10-equation models were estimated.

Model	F 18-34	M 18-34	F 35-64	M 35-64	F 65 +	M 65 +
1 Smoking-Related Disease	✔	✔	✔	✔	✔	✔
2 Poor Health Status	✔	✔	✔	✔	✔	✔
3 Positive Hospitalization Expenditures	✔	✔	✔	✔	✔	✔
4 Positive Ambulatory Care Expenditures	✔	✔	✔	✔	✔	✔

5 Positive Prescription Drug Expenditures	✔	✔	✔	✔	✔	✔
6 Positive Home Health Expenditures	✔	✔	✔	✔	✔	✔
7 log (Hospitalization Expenditures)	✔	✔	✔	✔	✔	✔
8 log (Ambulatory Expenditures)	✔	✔	✔	✔	✔	✔
9 log (Prescription Drugs Expenditures)	✔	✔	✔	✔	✔	✔
10 log (Home Health Expenditures)	✔	✔	✔	✔	✔	✔

Reference for Appendix 2

1 U.S. Department of Health and Human Services. *The health consequences of smoking: a report of the Surgeon General*. Atlanta, Georgia: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2004.

APPENDIX 3. Estimation and Forecast of Smoking-Attributable Healthcare Expenditures

Estimation of the National Models

The econometric models for healthcare cost of smoking consist of 10 equations (see Appendix 2). Because data on both smoking status and healthcare expenditures are not available in any one dataset for California, the econometric models were estimated by using the national data first (i.e., national models). The national models were estimated separately for each of the 6 subgroups defined by gender and age (18-34, 35-64, 65+) among adults aged 18 and older using the 2002-2007 linked MEPS-NHIS data.

Estimation of Smoking-Attributable Fractions (SAF) and Relative Risk (RR) for California for 2007

After the national models were estimated, we applied the estimated parameters to the actual values of the independent variables from California's 2007 CHIS data in order to obtain California-specific smoking-attributable fractions (SAFs) and relative risks (RRs) of health expenditures for 2007. The "smoking-attributable fraction" indicates the proportion of healthcare expenditures that could be attributed to smoking. The relative risk of healthcare expenditures is defined as the ratio of mean healthcare expenditure for smokers relative to that for never smokers. Two sets of predicted expenditures were calculated for California: one for the factual case, and one for a counterfactual case. In the factual case, the set of predicted expenditures was calculated for current light, current moderate, and current heavy, and former smokers. In the counterfactual case, the set of predicted expenditures was calculated for "hypothetical nonsmoking" current and former smokers. "Hypothetical nonsmoking" current (or former) smokers are identical to current (or former) smokers in every way except that they are assumed to be never smokers. The difference between these two sets of predicted expenditures is the excess cost of smoking. The ratio of the excess cost of smoking to the total predicted expenditures from the first set of predictions is the SAF for healthcare expenditures. The ratio of the mean predicted expenditures for smokers to the mean predicted expenditures for never smokers is the RR for healthcare expenditures of smoking. RRs and SAFs for California for 2007 were calculated separately for each type of healthcare expenditures.

The relationship between the SAFs and RRs can be illustrated by the following equations. The SAF for the type of healthcare expenditure j can be decomposed into two components: one for current smokers, and one for former smokers as specified below:

$$SAF_j = SAF_{j,c} + SAF_{j,f} \qquad \text{(Eq. 1)}$$

$$SAF_{j,c} = \frac{\sum_{cl=1}^{Ncl}(EXP_{j,cl} - EXP_{j,cl\to n}) + \sum_{cm=1}^{Ncm}(EXP_{j,cm} - EXP_{j,cm\to n}) + \sum_{ch=1}^{Nch}(EXP_{j,ch} - EXP_{j,ch\to n})}{\sum_{n=1}^{Nn}(EXP_{j,n}) + \sum_{cl=1}^{Ncl}(EXP_{j,cl}) + \sum_{cm=1}^{Ncm}(EXP_{j,cm}) + \sum_{ch=1}^{Nch}(EXP_{j,ch}) + \sum_{f=1}^{Nf}(EXP_{j,f})} \qquad \text{(Eq. 2)}$$

$$SAF_{j,f} = \frac{\sum_{f=1}^{Nf}(EXP_{j,f} - EXP_{j,f\to n})}{\sum_{n=1}^{Nn}(EXP_{j,n}) + \sum_{cl=1}^{Ncl}(EXP_{j,cl}) + \sum_{cm=1}^{Ncm}(EXP_{j,cm}) + \sum_{ch=1}^{Nch}(EXP_{j,ch}) + \sum_{f=1}^{Nf}(EXP_{j,f})} \qquad \text{(Eq. 3)}$$

where $SAF_{j,c}$ = smoking attributable fraction for current smokers including current light, moderate, and heavy smokers for healthcare expenditure type j

$SAF_{j,f}$ = smoking attributable fraction for former smokers for healthcare expenditure type j

$EXP_{j,n}$, $EXP_{j,cl}$, $EXP_{j,cm}$, $EXP_{j,ch}$, $EXP_{j,f}$ = predicted expenditures for a never smoker n, current light smoker cl, current moderate smoker cm, current heavy smoker ch, or former smoker f for healthcare expenditure type j

$EXP_{j,cl\rightarrow n}$, $EXP_{j,cm\rightarrow n}$, $EXP_{j,ch\rightarrow n}$, $EXP_{j,f\rightarrow n}$ = predicted expenditures for a hypothetical "nonsmoking current" light smoker cl, moderate smoker cm, or heavy smoker ch, or a "nonsmoking former smoker" f who has the identical characteristics of a smoker except that he/she is assumed to be a never smoker for healthcare expenditure type j

N_n, N_{cl}, N_{cm}, N_{ch}, N_f = total number of never smokers, current light smokers, current moderate smokers, current heavy smokers, or former smokers

Then, we rewrite Equations (2-3) and express them in terms of the relative risk (RR) and smoking prevalence by first replacing each summation by the product of its mean value and its sample size, then normalizing each sample size with total population (i.e., sum of never smokers, current smokers, and former smokers), and normalizing each mean value by the mean predicted expenditure for never smokers. Therefore, the SAF equations can be transformed into:

$$SAF_{j,c} = \frac{(P_{cl})(RR_{j,cl} - RR_{j,cl\rightarrow n}) + (P_{cm})(RR_{j,cm} - RR_{j,cm\rightarrow n}) + (P_{ch})(RR_{j,ch} - RR_{j,ch\rightarrow n})}{(P_n) + (P_{cl})(RR_{j,cl}) + (P_{cm})(RR_{j,cm}) + (P_{ch})(RR_{j,ch}) + (P_f)(RR_{j,f})} \qquad \text{(Eq. 4)}$$

$$SAF_{j,f} = \frac{(P_f)(RR_{j,f} - RR_{j,f\rightarrow n})}{(P_n) + (P_{cl})(RR_{j,cl}) + (P_{cm})(RR_{j,cm}) + (P_{ch})(RR_{j,ch}) + (P_f)(RR_{j,f})} \qquad \text{(Eq. 5)}$$

where P_n = prevalence of never smokers (=N_n / total population)

P_{cl}, P_{cm}, P_{ch} = prevalence of current light, current moderate, or current heavy smokers

P_f = prevalence of former smokers (=N_f / total population)

$RR_{j,cl}$, $RR_{j,cm}$, $RR_{j,ch}$, $RR_{j,f}$ = relative risk of healthcare expenditure type j for current light, current moderate, current heavy, or former smokers relative to never smokers, defined by the ratio of the mean predicted expenditures for current light, current moderate, current heavy, or former smokers to the mean predicted expenditures for never smokers

$RR_{j,cl\rightarrow n}$, $RR_{j,cm\rightarrow n}$, $RR_{j,ch\rightarrow n}$, $RR_{j,f\rightarrow n}$ = relative risk of healthcare expenditure type j for hypothetical "nonsmoking current" light, moderate or heavy smokers, or hypothetical "nonsmoking former smokers" relative to never smokers, defined by the ratio of the mean predicted expenditure for "nonsmoking current" light, moderate or heavy smokers, or hypothetical "nonsmoking former smokers" to the mean predicted expenditure for never smokers

Projection of the SAFs for California for 2008-2016

To forecast the SAFs for the years beyond 2007, we applied Equations (4-5) to the estimated California-specific RRs for 2007 and the projected smoking prevalence rates for 2008-2016 under each of the four scenarios. We assumed that the RRs would remain constant over time from 2007 to 2016. Thus, once the prevalence rates for current light, current moderate, current heavy, former, and never smokers for each year are determined, the SAF for each year would be determined.

Projection of California Healthcare Expenditures by Type of Healthcare Services for Adults for 2007-2016

State-specific healthcare expenditures by type of healthcare services were available only until 2004. Thus, we projected health expenditures by type of healthcare service in 2007-2016 for California adults as follows. First, we used the Centers for Medicare and Medicaid Services (CMS) National Health Expenditures projections by type of healthcare services for the U.S. for 2007 through 2016.[1] The second step involved estimating California health expenditures through 2016. CMS also estimates the state-level health expenditures as well as the ratio of each state’s health expenditure to national health expenditure for each type of healthcare service for 1991-2004.[2] According to these estimates, we found that the ratios of California to national health expenditures by type of healthcare services have been stable for a number of years. We applied the 2004 ratio to the national projections through 2016 to obtain estimates of California health expenditures for total health expenditures and for each component of healthcare services. The estimated expenditures for each component of healthcare services were then calibrated so that the sum of all components equals total health expenditures. These estimates are for all ages. Third, we determined California healthcare expenditures for all health services and for each type of healthcare services for adults aged 18 and older. This was done by applying the ratio of adult expenditure to all-age healthcare expenditures by type of care estimated from the 2007 MEPS data, and then multiplying each ratio by its corresponding all-age healthcare expenditures. The estimated adult expenditures for each type of healthcare services were again calibrated so that the sum of all types equals total adult health expenditures.

Estimation of Smoking-Attributable Healthcare Expenditures for Each Year and Scenario

The estimated SAFs were multiplied by California healthcare expenditures for adults to obtain smoking-attributable expenditures for each year from 2007 to 2016 under each of the four tobacco control funding scenarios. Then, the nominal values of smoking-attributable expenditures were expressed in 2009 constant dollars using the annual deflator from Consumer Price Index for Medical Care [3]. Because the price index data were not available for the years 2011-2016, we used a deflator estimated by averaging the annual growth rates in Consumer Price Index for Medical Care over the latest five years from 2005-2010.[3]

References for Appendix 3

1 Centers for Medicare & Medicaid Services (CMS). *National Health Expenditure Projections 2009-2019, Forecast summary and selected tables.* Washington, DC: Centers for Medicare & Medicaid Services, 2009. http://www.cms.gov/NationalHealthExpendData/downloads/proj2009.pdf. Accessed 2/24/11.

2 Centers for Medicare and Medicaid Services (CMS). Health expenditures by state of residence, state-specific tables, 1991–2004. Washington, DC: Centers for Medicare & Medicaid Services, 2007. https://www.cms.gov/NationalHealthExpendData/downloads/res-states.pdf. Accessed February 15, 2011.

3 U.S. Department of Labor, Bureau of Labor Statistics. *Consumer price index detailed report, Table 3A—consumer price index for all urban consumers (CPI-U): US city average, detailed expenditure categories. Various years*. http://www.bls.gov/cpi/#tables. Accessed March 12, 2011.