ADDICTION

The effect of a major cigarette price change on smoking behavior in California: a zero-inflated negative binomial model

Mei-ling Sheu[a],*, Teh-wei Hu[b], Theodore E. Keeler[b], Michael Ong[c] and Hai-Yen Sung[c]

[a] *Taipei Medical University, Taiwan*
[b] *University of California, Berkeley, USA*
[c] *University of California, San Francisco, USA*

Summary

The objective of this paper is to determine the price sensitivity of smokers in their consumption of cigarettes, using evidence from a major increase in California cigarette prices due to Proposition 10 and the Tobacco Settlement.

  The study sample consists of individual survey data from Behavioral Risk Factor Survey (BRFS) and price data from the Bureau of Labor Statistics between 1996 and 1999. A zero-inflated negative binomial (ZINB) regression model was applied for the statistical analysis.

  The statistical model showed that price did not have an effect on reducing the estimated prevalence of smoking. However, it indicated that among smokers the price elasticity was at the level of −0.46 and statistically significant.

  Since smoking prevalence is significantly lower than it was a decade ago, price increases are becoming less effective as an inducement for hard-core smokers to quit, although they may respond by decreasing consumption. For those who only smoke occasionally (many of them being young adults) price increases alone may not be an effective inducement to quit smoking. Additional underlying behavioral factors need to be identified so that more effective anti-smoking strategies can be developed. Copyright © 2004 John Wiley & Sons, Ltd.

Keywords   price change; cigarette consumption; tobacco tax; smoking prevalence

Introduction

California has increased its state tobacco tax three times since 1989 to discourage cigarette consumption and increase revenues. Proposition 99, the California Tobacco Tax and Promotion Act, increased the cigarette tax from 10 to 35 cents per pack in January 1989. To fund breast cancer research and the services for early detection, two more cents were added to the tax in January 1994. The most recent tax increase occurred on January 1, 1999 when Proposition 10, the California Children and Families First Act, increased the state tax by an additional 50 cents per pack.

  In addition, the retail price of cigarettes has also increased because of the nationwide Master Settlement Agreement (MSA) in 1998. The settlement included an agreement that the tobacco industry would have to pay $206 billion to 46 states over 25 years [1]. Consequently, the tobacco companies have raised the price of cigarettes per pack by approximately 40 cents at the end of 1998 and another 20 cents in 1999. Overall, the average price of cigarettes (including the tobacco tax) in

---

*Correspondence to: School of Health Care Administration, Taipei Medical University, 250 Wu-hsing St, Taipei 110, Taiwan.
E-mail: mlsheu@tmu.edu.tw

real terms was 63% higher in 1999 than in 1997 in California.

With respect to the estimates of price elasticity of cigarette demand, Sung *et al.* used panel data for 11 western states over the period between 1969 and 1990, and found that price elasticity was about –0.40 in the short run and −0.48 in the long run [2]. Keeler *et al.* studied the cigarette demand based on time-series data from 1980 to 1990 in California and concluded that the short-run price elasticities were between −0.30 and −0.50, and long-run price elasticities between –0.50 and –0.60 [3]. Hu *et al.* analyzed the impact of Proposition 99 and the anti-smoking media campaign [4]. The estimated elasticity of cigarette sales with respect to the tobacco tax was −0.30. In order to determine the influence of Proposition 10 and the tobacco settlement on tobacco consumption, we have analyzed monthly cigarette sales data in California from 1991 to 1999 [5]. The results show an estimated elasticity of approximately −0.34, and cigarette sales have decreased between 0.8 and 0.9 packs per capita per month because of the drastic price increase.

Although the estimates based on aggregate sales data demonstrate that price increases reduced cigarette consumption, a clear understanding of the way individuals make decisions on cigarette consumption requires behavioral information at the individual level. Previous studies based on individual survey data showed that the price elasticities ranged from −0.23 to −0.42 [6–8]. Considering individual behavioral risk factors and cigarette consumption in California, Hu *et al.* analyzed the data of Behavioral Risk Factor Survey (BRFS) from 1985 to 1991 [9]. With a two-part model, they found an overall price elasticity of –0.46, and the reduction in cigarette consumption to be equally attributable to the decrease in smoking prevalence and the reduction in consumption among smokers.

Due to the efforts of anti-smoking campaigns and the increase in cigarette tax, the smoking prevalence has been steadily decreasing in California. According to the BRFS data, the smoking prevalence was 25.8% among the adult population (age 18 and over) in California in 1984. It hit a low rate, 16.4%, in 1995, but rose to 18.6% in 1996 and 18.8% in 1999. It is difficult to determine how low the prevalence rate could go. Nevertheless, smoking prevalence is unlikely to continue its decrease, given that the tobacco companies continue to advertise to attract young participants [10–12]. The average prevalence among individuals aged 18–24 was 18% between 1990 and 1994 and 22% between 1995 and 1999. Given that the current prevalence of cigarette consumption is significantly lower than that a decade ago, it is worth revisiting the issue as to whether further price increases will continue to reduce both smoking prevalence and cigarette consumption among smokers. The present paper aims to estimate the impact of this most recent wave of major price increase on cigarette consumption in California. We also pay special attention to the smoking patterns among individuals with different demographic characteristics.

Methods

Data

With the exception of data on cigarette prices, the main data source of this research is the BRFS, which is conducted by the California Department of Health Services in collaboration with the Centers for Disease Control and Prevention. We include the data from 1996 to 1999 for the analysis in this paper for two main reasons. First, the survey questions underwent a significant change in 1996. This issue will be discussed in further detail in the next section. Second, this time span covers the pre- and post-Proposition 10 periods and the number of observations is sufficiently large for the statistical analysis. The original BRFS data during this time frame include 16 260 individual observations. When observations with missing or unknown values for the key variables are dropped, the total sample reduces to 16 147. Since cigarette price data are only available for the three metropolitan areas, Los Angeles, San Diego and San Francisco, only individuals (12 189 observations) residing in these three areas are selected.[a] Moreover, we decided to focus on the smoking behavior of the three major ethnic groups, Whites, Blacks and Hispanics.[b] After individuals of other ethnicities are excluded, the final sample size of this research is 11 180.

Cigarette price information was obtained from the Bureau of Labor Statistics. This information consists of monthly price indices of tobacco products for the Los Angeles, San Diego and
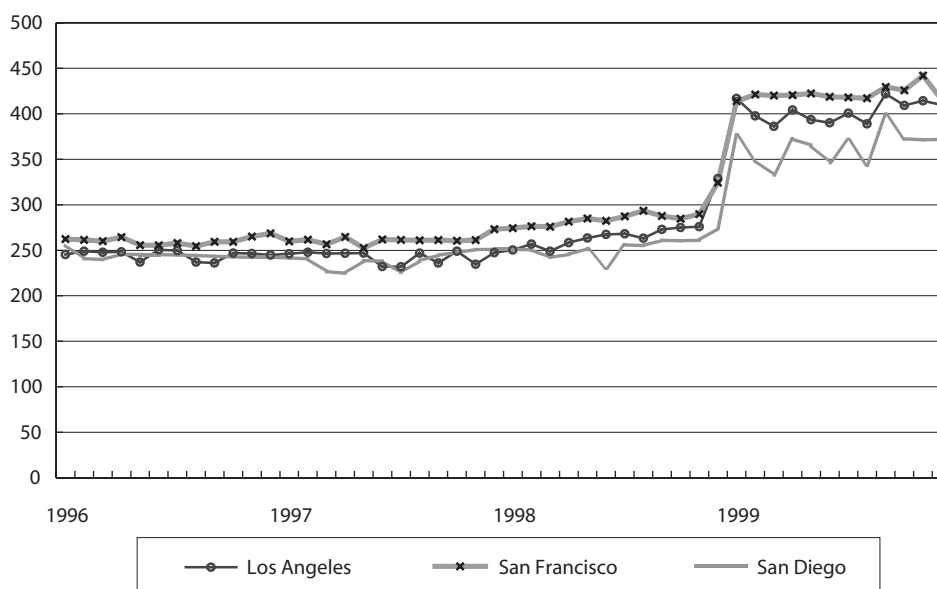
Chart 1. Monthly deflated cigarette prices in three California metropolitan areas

San Francisco metropolitan areas; these indices are deflated to 1982–1984 dollars using the consumer price index for all urban consumers. Each individual respondent in the BRFS dataset is then pegged to the corresponding deflated price index based on the individual's county of residence and the date of the interview (month and year). Chart 1 shows the trend of the deflated price indices for all three metropolitan areas over the 4 year study period.

Dependent and explanatory variables

The dependent variable is the number of cigarettes smoked per day. It is constructed based on the BRFS data. One important point about the variable is that the way in which smokers and the amount of cigarettes consumption are defined in the BRFS data has changed after 1996. Prior to 1996, two key questions defined smokers. First, the individuals were asked if they had smoked 100 cigarettes in their entire life. Second, for those who answered yes to the first question, they were asked if they currently smoke. If the answer was yes, they were classified as current smokers and were asked the number of cigarettes they smoked per day. Starting from 1996, these questions were modified. The first question remained the same, but the second question for those who responded as

having smoked 100 cigarettes was changed to 'Do you now smoke cigarettes everyday, some days or not at all?' Those who smoked everyday (the regular smokers) were asked the average number of cigarettes they smoked per day. Those who smoked some days or not at all were asked whether they had smoked cigarettes in the past 30 days. If the answer was yes, they were asked the number of days they had smoked during the past 30 days. Those who responded to the second question as having smoked some days (the irregular smokers) were also asked the number of cigarettes they smoked per day on the days that they did smoke.

Through the new design in the BRFS survey questions, a fair amount of people revealed that they were irregular smokers. Table 1 shows that 17.9% of the individuals in the sample were current smokers with 12.5% regular smokers and 5.5% irregular smokers. To make the level of cigarette consumption comparable between these two types of current smokers, we need to construct the number of cigarettes smoked per day for those irregular smokers. The number of smoking days was multiplied by the number of cigarettes smoked per day and was then divided by 30 days. Table 1 indicates that the average number of cigarettes consumed per day between regular and irregular smokers is quite different, 16.1 and 2.0, respectively.

Table 1. Sample size, prevalence rate of smoking, and average number of cigarettes smoked per day by socio-demographic characteristics (1996–1999)

| Characteristics | Total sample | Percent smokers | | | Average number of cigarettes smoked per day | |
		Total	Smoke everyday	Smoke some days	Smoke everyday	Smoke some days[a]
Total	11 180	17.9%	12.5%	5.5%	16.1	2.0
Ethnicity						
White	7448	18.1%	13.5%	4.6%	17.9	2.5
Black	758	24.9%	17.5%	7.4%	12.2	2.1
Hispanic	2974	15.8%	8.6%	7.2%	11.2	1.3
Age (years)						
18–24	1080	22.3%	12.9%	9.4%	12.0	1.5
25–34	2402	18.1%	11.1%	7.0%	13.2	1.5
35–44	2673	19.1%	13.3%	5.8%	16.1	1.7
45–54	2008	19.8%	15.1%	4.6%	17.6	2.7
55–64	1221	19.2%	15.5%	3.8%	19.7	3.2
over 65	1796	10.3%	7.7%	2.6%	17.6	3.7
Gender						
Male	4851	20.3%	13.9%	6.4%	17.0	2.0
Female	6329	16.1%	11.3%	4.8%	15.3	2.0
Marital status						
Couple	6005	14.1%	10.0%	4.1%	15.4	1.8
Single	2353	22.5%	13.8%	8.7%	14.7	1.8
Divorced	1939	25.5%	19.0%	6.5%	18.3	2.5
Widowed	883	15.2%	11.3%	3.9%	16.5	3.2
Education						
Some high/tech school	1687	19.6%	12.9%	6.7%	13.7	1.6
High school graduate	2822	24.1%	17.6%	6.4%	16.6	2.2
Some college	3012	19.4%	13.3%	6.0%	16.8	2.3
College graduate	3659	11.2%	7.5%	3.7%	16.1	1.7
Employment						
Employed	10 554	17.3%	12.1%	5.3%	16.0	2.0
Unemployed	626	27.6%	19.0%	8.6%	16.7	1.9
Income						
Less than $10 000	1156	23.3%	15.9%	7.4%	15.4	1.5
10 000–20 000	1717	22.4%	15.1%	7.3%	15.0	2.1
20 000–34 000	2088	20.7%	14.8%	5.9%	15.8	2.1
35 000 and over	5545	14.9%	10.4%	4.5%	17.2	2.0
No response	674	13.6%	9.5%	4.2%	14.1	2.9
Health status						
Excellent	3127	12.4%	7.9%	4.5%	15.9	1.6
Good	6486	19.5%	13.7%	5.8%	15.9	2.1
Fair	1179	20.9%	14.8%	6.0%	15.9	2.0
Poor	388	27.3%	20.4%	7.0%	19.1	2.6
Year						
1996	2761	18.0%	12.9%	5.1%	16.1	2.7
1997	2786	16.7%	11.7%	5.0%	16.5	1.5
1998	2789	19.0%	13.5%	5.6%	16.7	2.0
1999	2844	18.0%	11.8%	6.2%	15.0	1.8

[a] Forty-four individuals reported that they smoked some days but did not smoke any cigarette during the past 30 days. The number of cigarettes smoked is coded as zero and included in this calculation.

The explanatory variables include cigarette price and socio-demographic variables such as ethnicity, age, gender, marital status, education, employment status, income, and health status. The price variable is transformed into logarithms. The socio-demographic variables are categorical, and the reference groups in the regression analysis are those who are white, age 18–24, female, couples, with some high school (or technical school) education, employed, with income less than $10 000, and with excellent health condition. To control for any consistent pattern in cigarette consumption over time, a year trend variable is also included.

Table 1 shows the sample size, smoking prevalence, and cigarette consumption per day by socio-demographic characteristics. Blacks had the highest smoking prevalence, 24.9%, while Whites consumed the highest number of cigarettes, 17.9 per day, among those regular smokers. With regard to age, the youngest group, age 18–24, had the highest smoking prevalence rate but they consumed fewer cigarettes per day than the other age groups. Males had higher smoking prevalence and smoked more than females. Couples had the lowest smoking prevalence; however, singles consumed fewer cigarettes per day because most of them were younger people. High school graduates had the highest smoking prevalence while college graduates had the lowest. The unemployed had a higher prevalence rate than the employed. Individuals who had income $35 000 or above and who did not reveal their income had the lowest smoking prevalence rate. However, among the regular smokers, those with incomes $35 000 or over had the highest cigarette consumption per day. Individuals with a poor health status had the highest smoking prevalence and consumed the most cigarettes per day.

Statistical methods

The distribution of individual cigarette consumption data is generally skewed to the right and contains a large proportion of zeros (i.e. excess zeros). To deal with these distributional characteristics, several estimation techniques have been reported in previous cigarette consumption studies using micro-level survey. These include the pseudo-Poisson model [8], the two-part model [4,6,8,13], and the hurdle model [14–17]. In this study, we have applied a zero-inflated

negative binomial (ZINB) regression model [18].

A ZINB model is a modified Poisson regression model that is designed to deal with two common issues that occur with the application of the Poisson model to count data. These include overdispersion and excess zeros [19,20]. In a basic Poisson regression model, the number of events y (such as the number of cigarettes smoked) for individual i has a Poisson distribution with a conditional mean λ depending on the characteristics, x, of the individual:

$$\lambda_i = E(y_i|x_i) = e^{x_i\beta} \tag{1}$$

And the probability of y given x is:

$$\Pr(y_i|x_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \tag{2}$$

A special characteristic of the Poisson distribution is that its variance is equal to its mean, λ. However, count data very often demonstrates 'overdispersion' meaning that the variance is larger than the mean. When overdispersion is an issue, the estimates based on Poisson regression will be inefficient [21]. Overdispersion is often caused by unobservable individual heterogeneity and/or excess zeros of the data.

Unobservable individual heterogeneity is likely to be an issue of the current study, given that the study sample consists of individuals with a wide variation of smoking status, nonsmokers, new smokers, former smokers, current occasional smokers, regular smokers, and heavy smokers. The identified individual socio-demographic characteristics might not be able to capture all of the heterogeneous cigarette consumption behaviors. One way to deal with this issue is to use the negative binomial (NB) regression models in which unobserved heterogeneity is considered by adding an error term, ε, to the conditional mean of the Poisson distribution.

$$\tilde{\lambda}_i = E(y_i|x_i) = e^{(x_i\beta+\varepsilon_i)} \tag{3}$$

Normally, $\exp(\varepsilon_i)$ is assumed to have a gamma distribution with mean 1 and variance α so that the conditional mean of y_i is still λ_i but the conditional variance of y_i becomes $\lambda_i(1 + \alpha\lambda_i)$:[c,d]

$$\frac{\text{Var}(y_i)}{E(y_i)} = 1 + \alpha E(y_i) \tag{4}$$

If α approaches zero, y becomes a Poisson distribution. As α becomes larger, the distribution will be more dispersed.

The phenomenon of excess zeros is definitely a concern in this study because 82.1% of the study sample were not current smokers. The issue of excess zeros can be dealt with through the application of zero-inflated Poisson (ZIP) regression models [19–22]. Lambert introduced the ZIP model [22]:

$y_i \sim 0$ with probability q_i

$y_i \sim \text{Poisson}(\lambda_i)$ with probability $1 - q_i$

$$(y_i = 0, 1, 2, 3, \ldots)$$

where $q_i = \dfrac{e^{z_i\gamma}}{1 + e^{z_i\gamma}}$ (5)

This model puts extra weight on the probability of observing a zero through a mixing specification. Conceptually, it divides individuals into nonusers, with probability q_i, and potential users, with probability $1-q_i$ [23]. The unobservable probability q_i is generated as a logistic function of the observable covariates to ensure nonnegativity. An observed zero for y_i is generated from either the logistic process or the Poisson process.

The mean and variance of y_i are

$$E(y_i) = q_i 0 + (1 - q_i)\lambda_i = (1 - q_i)\lambda_i \qquad (6)$$

and $\quad \text{Var}(y_i) = \lambda_i(1 - q_i)(1 + \lambda_i q_i)$ (7)

Then

$$\frac{\text{Var}(y_i)}{E(y_i)} = 1 + \lambda_i q_i = 1 + \left[\frac{q_i}{1 - q_i}\right] E(y_i) \qquad (8)$$

Therefore, if q_i approaches zero, a Poisson distribution emerges. Similar to α in the NB model, $q_i/(1-q_i)$ reflects the degree of overdispersion.

Greene has expanded the ZIP model by adopting different specifications to the two statistical processes; that is, the observed count variable y_i is generated as a product of the two latent variables z_i and $y_i{}^*$ [19]:

$$y_i = z_i y_i{}^* \qquad (9)$$

where z_i is a binary variable with values 0 or 1, and y_i^* has a Poisson or NB distribution. Then,

$\Pr(y_i = 0) = \Pr(z_i = 0) + \Pr(z_i = 1, y_i{}^* = 0)$

$\qquad\qquad = q_i + (1 - q_i)f(0)$

$\Pr(y_i = k) = (1 - q_i)f(k), \quad k = 1, 2, \ldots$

where $f(\cdot)$ is the Poisson or negative binomial probability distribution for y_i^*. The binary process z_i can be modeled using logit or probit or other

models. Therefore, there can be different combinations of zero-inflated models from the two processes. When the second process has a NB distribution, Equation (9) is defined as the ZINB model. We applied the ZINB model in which the binary process is estimated by the logit model. The ZINB model has the variance

$$\text{Var}(y_i) = \lambda_i(1 - q_i)[1 + \lambda_i(q_i + \alpha)] \qquad (10)$$

and $\quad \dfrac{\text{Var}(y_i)}{E(y_i)} = 1 + \left[\dfrac{q_i + \alpha}{1 - q_i}\right] E(y_i)$ (11)

Note that the overdispersion terms in Equations (4), (8), and (11) are from different sources, either from unobserved heterogeneity or excess zeros or both. Since the Poisson model and ZIP models are not nested (likewise for the NB and ZINB models), the Vuong non-nested test can be used to decide which model has a better fit [19,24]:

$$V = \frac{\sqrt{N}\bar{m}}{s_m} \qquad (12)$$

where $m_i = \ln[\hat{P}_1(y_i|x_i)/\hat{P}_2(y_i|x_i)]$ and $\hat{P}_1(y_i|x_i)$ and $\hat{P}_2(y_i|x_i)$ are the predicted probabilities of the two competing models. $\bar{m}$ is the mean and s_m is the standard deviation of m_i. V has a asymptotically normal distribution. If $|V|$ is less than the critical value, such as 1.96, neither model is preferred. If V is larger than the critical value 1.96, model 1 is favored. If V is smaller than the critical value -1.96, model 2 fits better. As noted by Greene and Grootendorst, we can choose the best model among the ZINB, ZIP, NB, and Poisson models by the following steps [19,25]. If the Vuong test shows that the ZINB model is rejected in favor of the NB model, the splitting mechanism is rejected. In this case, we will estimate the NB model and test if the heterogeneity parameter α is significant by using the t-test; a significant α suggests that unobservable heterogeneity accounts for dispersion. On the other hand, if the Vuong test shows that the NB is rejected in favor of the ZINB model, we will test if the parameter α in the ZINB model is significant. If the estimate of α is also significant, both the splitting mechanism and individual heterogeneity account for dispersion.

The Vuong test of ZINB and NB model for our study sample shows that $V = 26.87$ (Table 2). Therefore, we choose the ZINB rather than NB model. We apply a ZINB model rather than a ZIP model because the estimated results show

Table 2. Zero-inflated negative binomial regression model

Parameter	Logit		Negative binomial	
	Estimate	z-Value	Estimate	z-Value
Ln(price)	0.135	0.57	−0.459	−2.07**
Black	0.062	0.59	−0.461	−4.87***
Hispanic	−0.648	−8.16***	−0.731	−10.52***
Age 25–34	−0.052	−0.48	0.135	1.33
Age 35–44	−0.019	−0.17	0.439	4.26***
Age 45–54	−0.039	−0.33	0.576	5.18***
Age 55–64	−0.225	−1.75*	0.720	5.89***
Age over 65	−1.197	−8.52***	0.613	4.54***
Male	0.352	6.25***	0.153	2.80***
Single	0.452	5.86***	0.004	0.06
Divorced	0.648	8.84***	0.172	2.47**
Widowed	0.412	3.25***	−0.070	−0.56
High school graduate	0.088	0.94	0.223	2.65***
Some college	−0.268	−2.70***	0.124	1.41
College graduate	−0.877	−8.14***	−0.016	−0.17
Unemployed	0.351	3.22***	0.101	1.04
$10 000–$19 999	0.147	1.40	−0.047	−0.49
$20 000–$34 999	−0.007	−0.07	0.039	0.39
$35 000 and over	−0.253	−2.44**	0.062	0.65
Income unknown	−0.313	−2.10**	0.087	0.59
Good health	0.510	7.39***	0.168	2.41**
Fair health	0.611	5.79***	0.217	2.15**
Poor health	0.945	6.37***	0.345	2.52**
Year trend	−0.007	−0.17	0.041	1.00
Constant	−2.141	−2.02**	3.908	3.88***
Ln α	0.261	3.79***		
α	1.299			
Vuong test of ZINB versus NB		26.87***		
N	11 180			
(cigarette > 0)	18%			
Mean cigarette consumption	2.11			
LR χ²(24)	254***			

Logit model: Pr(Number of cigarette > 0).
*** = $p < 0.01$;
** = $p < 0.05$;
* = $p < 0.10$.

overdispersion (i.e. the estimated Ln α = 0.26 and the t-ratio = 3.79) after the excess zero issue is addressed.

Besides ZIP or ZINB models, two-part or hurdle models are commonly applied in count data with excess zeros. The basic idea is that the participation decision and the positive counts are generated by separate processes [23]. For cigarette consumption, the decision model for smoking is generally specified by a logit or probit model. The second part is a truncated at zero count model that only focuses on individuals with positive cigarette consumption.

It may not be easy to say which model, the zero-inflated model or the two-part/hurdle model, is better [26–29]. It depends on the reasoning of the study question. Cheung suggested that in clinical research the zero-inflated model should be more useful when there is a strong basis to expect a necessary condition for a subsequent Poisson outcome. On the other hand, if all individuals

are at risk of a certain event (e.g. recurrence of cancer for all cancer patients), the realization of that event represents the pass of a 'hurdle' and the two-part model should be more appropriate [26].

Empirically, using prescription drug utilization data, Grootendorst compared the two-part model with the zero-inflated model [25]. The results showed that the two-part model performed better than the alternatives. With the data of congressional responses to Supreme Court decisions, Zorn found that both zero-inflated and hurdle poisson specifications generated similar results [30].

The main rationale for our application of the zero-inflated model is at the conceptual level. First of all, the clear distinction (as suggested by the hurdle model) between nonsmokers and smokers based on zero cigarette consumption or not is oversimplified. The smoking behavior is much more complicated. Most people are not smokers. Many of them are not at risk and would not smoke regardless of any price level. Among smokers, some just started to try to smoke. Their cigarette consumption is not regular. For some other people, they were smokers but tried to quit smoking. They may not quit smoking at once but rather decrease smoking gradually [31]. Their cigarette consumption became irregular. Still, some smokers may just respond to price changes and adjust their cigarette consumption even to the level of zero. In our study sample, there were individuals revealed to be smokers who did not smoke regularly or even did not smoke during the past 30 days. They could be people of these three kinds. Since those with zero cigarette consumption includes individuals at risk as well as not at risk, the application of the zero-inflated model is appropriate [25].

In summary, given the decision-making just described, we deem it much more reasonable to assume that the decisions as to whether to smoke, and, if so, how much, are conceptually integrated for a given consumer, rather than being totally disjoint. This, in turn, suggests that the ZINB model is to be preferred to the two-part model on *a priori* grounds.

For comparison purposes and to check the robustness of the results from the ZINB model, we also applied the two-part model. The STATA program was used for the empirical analyses.

Results

Table 2 presents the results of the ZINB regression model, which include a logit and a NB regression. The original setup of the logit model is to predict the probability of being in the nonsmoking group. However, for the convenience of comparing the results to those of the NB regression model that estimates the cigarette consumption among the potential smokers, the signs of the coefficients have been changed so that the logit model reflects the probability of being in the potential smoking group. In the logit model, the price coefficient was not statistically significant and the sign was positive, indicating that price did not have an effect on reducing the likelihood of being a smoker. On the other hand, the NB model showed that among smokers the price elasticity was at the level of –0.46 and statistically significant. It indicates that for a 10% increase in price, the cigarette consumption per day would decrease approximately by 4.6%.

While Blacks were not statistically different from Whites in smoking prevalence, Hispanics were less likely than Whites to be smokers. The odds ratio of the smoking participation of Hispanics *versus* Whites was $0.52(= \exp(-0.648))$. The NB regression results showed that both Blacks and Hispanics consumed fewer cigarettes than Whites among smokers. For Blacks, the expected cigarette consumption decreased by a factor of $0.63(= \exp(-0.461))$. Similarly, Hispanics had a lower level of cigarette consumption by a factor of 0.48.

Consistent with the finding from the descriptive analysis, the smoking prevalence and cigarette consumption demonstrated different patterns among different age groups. Compared to the reference group defined as individuals between 18 and 24 years of age, older individuals at the age of and over 55 had lower smoking participation, although only those whose age was above 65 were statistically significant at *p*-value < 0.05. As to the amount of cigarettes consumed among smokers, those of age 35 and over smoked statistically more cigarettes than did the 18–24 cohort. Those between the age of 55 and 64 smoked more cigarettes than the base group at a factor of 205%. In the categories of marital status, the smoking prevalence for couples, the base group, was statistically significantly lower than all other groups. Among smokers, only those who were divorced smoked statistically significantly more than couples.

Education also had different influences on smoking participation and cigarette consumption. Individuals with some college education, especially college graduates, had a lower probability of being smokers. However, college educated smokers did not consume fewer cigarettes than the comparison group who only had some high school education. On the other hand, smokers who were high school graduates had higher cigarette consumption than the comparison group. Unemployed individuals had higher smoking participation, but their cigarette consumption was similar to those who had jobs. Individuals whose income was $35 000 or higher and whose income was unknown had lower smoking prevalence. However, no difference was found in cigarette consumption among smokers of different income categories. In terms of self-perceived health condition, individuals with excellent health (the reference group) had the lowest smoking prevalence. The odds of being a smoker for those with poor health were 2.57 times larger than those with excellent health. Among smokers, those with excellent health smoked fewer cigarettes than those in other health status groups. The year trend variable was not statistically significant.

Overall, we found that the factors distinguishing potential smokers from nonsmokers are somewhat different from those that affect the level of cigarette consumption among smokers. In the current study sample, the increase in cigarette price had no effect on smoking participation but was associated with a decrease in cigarette consumption among smokers. In terms of socio-demographic characteristics, those who were more likely to smoke were Whites and Blacks, younger, male, single (including those divorced or widowed), less educated, unemployed, of less income and with poorer health. Among potential smokers, the cigarette consumption was higher for those who were Whites, older, male, divorced, and with poorer health.

For comparison, we did two-part models and found that the results were consistent with ZINB models. Table 3 lists the price elasticities based on the ZINB model as well as the two-part model. For the two-part model, two types of models were estimated for the second part, a NB model and an OLS model. The price elasticity from the second part of the two-part model was estimated to be –0.42 for the NB model and –0.47 for the OLS model, both of these were very close to the estimate, –0.46, in the NB regression of the ZINB model, although the price elasticity from the OLS model was barely significant at P-value $= 0.05$.

Discussion

One major finding of this research that differs from the previous literature is that the price increase did not have an effect on the level of overall smoking prevalence, but did have a negative impact on the level of cigarette consumption among potential smokers during the study time period. Two important factors might have contributed to this result. The first factor is that smoking prevalence among young adults has been increasing over time. According to California Department of Health Services, the 18–24 age group exhibited the greatest increase in smoking prevalence among the four age groups (18–24, 25–44, 45–64 and 65+) after 1995. Furthermore, it is the only group continuing a rising trend after 1998 [32]. The second factor is related to the identification of the smoking pattern of irregular smokers. The average level of cigarette consumption for the irregular smokers was much lower than that of the regular smokers. Therefore, cigarette consumption showed larger variations than the situation where

Table 3. Price elasticities of different statistical models

| | ZINB | | | Two-part model | |
	Estimate	z-Value		Estimate	z-Value
			First part		
Logit	0.135	0.57	Logit	0.056	0.25
			Second part		
NB	−0.459	−2.07	(1) NB	−0.417	−2.43
			(2) OLS[a]	−0.474	−1.82

[a]For the OLS model, the dependent variable, the number of cigarette consumption, is logarithm transformed.

the irregular smokers had not been recognized. In addition, while the cigarette consumption for both types of smokers was decreasing after cigarette price increased, the smoker mix was also gradually changing, with relatively fewer regular smokers but more irregular smokers (Table 1). Consequently, the statistical model of this study showed a higher estimate in price elasticity among smokers than what was found in the previous studies.

These observations have important implications for anti-smoking policies. First, in order to further reduce smoking prevalence, smoking participation among young people needs to decrease. Two characteristics of smoking patterns can be found for the young adults: (1) smoking fewer cigarettes; and (2) having a higher proportion of people who smoke irregularly. Therefore, in theory, it should be easier for young people to quit smoking than for those who are older and have been addicted to smoking for a long time. Price increases alone may not be effective enough to encourage these people to quit smoking, especially for those who do not smoke much. Other measures such as education and anti-smoking campaigns are important to decrease smoking participation among young people, especially teenagers.

Second, the identification of regular and irregular smokers is important because it reveals that the usual dichotomy between smokers and non-smokers is not sufficient for the design of anti-smoking strategies. The irregular smokers could be either young adults who have not yet become addicted or individuals who are trying to quit smoking. Anti-smoking campaigns could be more effective if different strategies were used to target different smoking sub-groups.

The use of a cigarette tax has been justified because it has been proven to be effective in reducing cigarette smoking. On the other hand, because cigarette consumers are also relatively unresponsive to price changes, governments have used the cigarette tax to raise revenues. However, when smoking prevalence has declined to a certain level, this kind of policy intervention is less likely to be as effective in achieving both goals as before. It is more likely that as governments continue to increase the tobacco tax, which will eventually approach a limit, the marginal benefit would decrease in terms of reducing cigarette prevalence and consumption. Additional underlying behavioral factors need to be identified so that more effective anti-smoking strategies can be developed.

Finally, it is worth noting that there are likely to be some health benefits from these price increases, to the extent that they generated significant reductions in smoking by some individuals. Such benefits are obviously not so high as those stemming from complete smoking cessation, but they should also not be ignored.

Notes

a. Specifically, individuals are classified as Los Angeles metropolitan residents if they resided in Los Angeles, Orange, Riverside, San Bernardino, or Ventura counties; as San Diego metropolitan residents if they resided in San Diego county; and as San Francisco metropolitan residents if they resided in Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Santa Cruz, Solano, or Sonoma counties.

b. We adopt one of the ethnic groupings in the BRFS dataset where four groups are classified: White, Black, Hispanic and Other. We exclude the Other group from the analysis. Regression results are similar if this Other group is included.

c. Given that $\exp(\varepsilon_i) = \delta_i$, the gamma distribution is

$$g(\delta_i) = \frac{\theta^\theta}{\Gamma(\theta)} \delta_i^{\theta-1} e^{(-\delta_i\theta)} \quad \text{for } \delta_i > 0, \theta = \frac{1}{\alpha}$$

where the gamma function is defined as $\Gamma(\theta) = \int_0^\infty t^{\theta-1} e^{-t} \, dt$

d. A negative binomial regression is given by:

$$P(y_i|x_i) = \frac{\Gamma(y_i + \theta)}{\Gamma(\theta)y_i!} u_i^\theta (1 - u_i)^{y_i}$$

where $\theta > 0$, $y_i = 0, 1, 2, \ldots$, $u_i = \theta/(\theta + \lambda_i)$, with $E \times (y_i) = \lambda_i$ and $\text{Var}(y_i) = \lambda_i(1 + (1/\theta)\lambda_i) = \lambda_i(1 + \alpha\lambda_i)$.

References

1. Gruber J. Tobacco at the crossroads: the past and future of smoking regulation in the United States. *J Econ Perspect* 2001; **15**(2): 193–212.

2. Sung HY, Hu TW, Keeler TE. Cigarette taxation and demand: an empirical model. *Contemp Econ Policy* 1994; **12**(3): 91–100.

3. Keeler TE, Hu TW, Barnett PG, *et al*. Taxation, regulation, and addiction: a demand function for cigarettes based on time-series evidence. *J Health Econ* 1993; **12**(1): 1–18.

4. Hu TW, Sung HY, Keeler TE. Reducing cigarette consumption in California: tobacco taxes vs an anti-smoking media campaign. *Am J Public Health 1995*; **85**(9): 1218–1222.

5. Hu TW, Ong M, Keeler TE, *et al*. *California cigarette consumption: the effects of the Proposition 10 tax and national tobacco industry settlement.* Report submitted to Tobacco Related Disease Research Program, University of California, 2001.

6. Lewit ML, Coate D. The potential for using excise taxes to reduce smoking. *J Health Econ* 1982; **1**(2): 121–145.

7. Chaloupka F. Rational addictive behavior and cigarette smoking. *J Polit Econ* 1991; **99**(4): 722–742.

8. Wasserman J, Manning WG, Newhouse JP, *et al*. The effects of excise taxes and regulations on cigarette smoking. *J Health Econ* 1991; **10**(1): 43–64.

9. Hu TW, Ren QF, Keeler TE, *et al*. The demand for cigarettes in California and behavioural risk factors. *Health Econ* 1995; **4**(1): 7–14.

10. Pierce JP, Choi WS, Gilpin EA, *et al*. Tobacco industry promotion of cigarettes and adolescent smoking. *JAMA* 1998; **279**(7): 511–515.

11. Evans N, Farkas A, Gilpin E, *et al*. Influence of tobacco marketing and exposure to smokers on adolescent susceptibility to smoking. *J Natl Cancer Inst* 1995; **87**(20): 1538–1545.

12. Pierce JP, Distefan JM, Jackson C, *et al*. Does tobacco marketing undermine the influence of recommended parenting in discouraging adolescents from smoking? *Am J Prev Med* 2002; **23**(2): 73–81.

13. Lewit EM, Coate D, Grossman M. The effects of government regulation on teenage smoking. *J Law Econ* 1981; **24**: 545–569.

14. Jones AM. A double-hurdle model of cigarette consumption. *J Appl Econometrics* 1989; **4**: 23–39.

15. Jones AM. The UK demand for cigarettes 1954–1986, a double-hurdle approach. *J Health Econ* 1989; **8**: 133–141.

16. Blaylock JR, Blisard WN. US cigarette consumption: the case of low-income women. *Am J Agric Econ* 1992; **74**: 699–705.

17. Blaylock JR, Blisard WN. Self-evaluated health status and smoking behaviour. *Appl Econ* 1992; **24**: 429–435.

18. Jones AM. *Applied Econometrics for Health Economists—a Practical Guide*. Office of Health Economics: London, 2002.

19. Greene WH. *Accounting for excess zeros and sample selection in poisson and negative binomial regression models*. Department of Economics Working Paper EC-94-10, Stern School of Business, New York University, 1994.

20. Long JS. *Regression models for categorical and limited dependent variables*. Sage Publications: Thousand Oaks, 1997.

21. Cameron AC, Trivedi PK. *Regression Analysis of Count Data*. Cambridge University Press: Cambridge, 1998; 60.

22. Lambert D. Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics* 1992; **34**: 1–14.

23. Jones AM. Health econometrics. In *Handbook of Health Economic*, Culyer AJ, Newhouse JP (eds). Elsevier: Amsterdam, 2000; Chapter 13.

24. Vuong Q. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 1989; **57**: 307–334.

25. Grootendorst Paul V. A comparison of alternative models of prescription drug utilization. *Health Econ* 1995; **4**: 183–198.

26. Cheung, Yi-Bin. Zero-inflated models for regression analysis of count data: a study of growth and development. *Stat Med* 2002; **21**: 1461–1469.

27. Campbell MJ, Machin D, D'Arcangues C. Coping with extra Poisson variability in the analysis of factors influencing vaginal ring expulsions. *Stat Med* 1991; **10**(2): 241–254.

28. Demetrio CGB, Ridout MS. Coping with extra Poisson variability in the analysis of factors influencing vaginal ring expulsions (letter to the editor). *Stat Med* 1994; **13**(8): 873–874.

29. Campbell MJ, Machin D. Author's reply. *Stat Med* 1994; **13**(8): 875–876.

30. Zorn CJW. *Evaluating zero-inflated, hurdle poisson specifications*. Midwest Political Science Association Political Methodology Working Papers, 1996; April 18–20.

31. Gilpin EA, Pierce JP. Demographic differences in patterns in the incidence of smoking cessation: United States 1950–1990. *Ann Epidemiol* 2002; **12**: 141–150.

32. The California Department of Health Services website. http://www.dhs.cahwnet.gov/tobacco/documents/AdultSmoking.pdf [10 February 2003].