A count data analysis of the determinants of cigarette consumption evidence from the city of Sulaimanyah in 2015

Hawnaz Omer Muhammad^{1*} Shaho Muhammad Wstabdulla²

- 1. Economic Department, School of Administrations and Economics, Sulaimani University, P.O.Box: 334 Sulaimani-IRAQ
- 2. Statistics Department, School of Administrations and Economics, Sulaimani University, P.O.Box: 334 Sulaimani-IRAQ

Abstract

Problem statement: The literature of smoking cigarettes well documented the harmful consequences of smoking on health statues. Exposure to smoking cigarette is associated with precocious death, economic losses to society, and a substantial burden on the health-care system. **Significance of the study:** The importance of the study comes from determining the influential factor of smoking decisions, thus, such studies might help antismoking policies to be more efficient in reducing the prevalence of smoking. **Objective:** this paper attempts to identify factors determining the participation decisions to smoke at the individual level in the City of Sulamnyah. **Approach:** Random samples of 650 individuals and employees have been selected to participate in survey questionnaires to study their participations of smoking. **Hypothesis:** It is assumed that age, education level, marital status, income, living with others or being alone, the influence of friends, having children, and health status are determine the number of smoking cigarette per day. **Methodology:** This paper will use Poisson regression model and negative binomial (NB) regression model to acquire the Hypothesis. **Results:** It is evident from the results that the elder, educated and married persons smoke less than younger, single, and less educated individuals. Moreover, friends have no influences in the participation to smoke. The results also identify that, income is not relevant in smoking determinations.

Keywords: Determinants of smoking cigarettes, Poisson regression model, Negative binomial regression model (NB), Sulamnyah city in Kurdistan region- Iraq

1. Introduction

The role of count data techniques has become increasingly popular in many applied economic researches. Bulk bodies of econometric studies utilize count data to determine the influential factors affecting individual level decision of participation and incidence rate. The realization of nonnegative integer values in count data leads its application to be more desirable. In general, univariate statistical models of incidence counts reveal the probability distribution of the number of events happening identified by some parameters. The probability distribution of count data is attempting to estimate those unknown parameters. Moreover, the count model contains nonetheless any other variables and its literature has usually assumed that number of events is independently and identically distributed (iid) (Cameron and Trivedi, 1998). Consequently, it is applicable to choose count data to model the determination factors that affect individual to smoke a cigarette (see, for instance, Mullahy, 1997; Ground and Koch, 2007; and Muhammad and Ahmad, 2010)

It is well known that tobacco use remains the leading preventable cause of death and disease, and the World Health Organization (WHO) has warned tobacco use "the single most preventable cause of death in the world today" (WHO, 2008). A strong link has been found between tobacco use and cancer, and according to American Cancer Society in (2014) smoking is responsible for almost one in five deaths in the United State. The literature of cigarette smoking generally accepts that smoking cigarettes have deleterious effects on human health (Wasserman et al., 1991; Garcia and Labeaga, 1996; Keeler et al., 2001; and Yen, 2005a). Indeed, consequences of smoking are serious this may cause the determinations of smoking cigarette has been a focus of interest for decades.

In many country around the world studies on the determinants of tobacco consumption have been implemented based on issues such as the influence of the smoking behavior of parents and peers or education on smoking incidence (see, for instance, Gruber and Zinman, 2000; and Sander, 1995), or the impact of gender differences in smoking behavior (see, for example, Bauer et al., 2007; and A Moghimbeigi et al., 2009). However, surprisingly, fewer studied have been implemented about determinants of smoking decisions in the Arab world and more specifically in Iraq and Kurdistan Region.

To the best of authors' knowledge, there is no piece of research employing data from Iraq in general and Kurdistan region particularly to identify social and economical determinations of smoking decisions based on count data modeling. The only mention about determinants of smoking cigarette in Iraq have been implemented by (Alghabban, 2009) who provides the empirical result for prevalence of smoking habit among college students at Kerbala university bases on the significance factor's *P*-values.

The literature of tobacco consumption provides sufficient evidence that smoking have significant adverse health effects on both individual smokers and passive smokers too (see, for example, Bennett et al, 1999; and Farrell et al, 2003). Smoking not only threats individual health condition but also causes enormous cost arising from the productivity lost and from the medical expenses associated smoking-induced diseases. This may explain the primary concern of professionals and policy makers on smoking around the world. Thus, the objective of this paper is to identify factors determining the participation decisions to smoke at the individual level in the City of Sulamnyah in 2015.

The paper focuses on the determination of cigarette smoking implying Poisson regression model and negative binomial (NB) regression model. The micro data have been used that is provided by questionnaires in Sylaimanyah in 2015. In the next section a brief survey of the empirical literature of the subject have been presented, and followed by the data and econometric techniques used. In section 4 will present the study results and finally section 5 provides the conclusion of the paper.

2. Literature Review

For several decades, numerous studies have been implemented by social scientists to understand individual cigarette smoking behavior. These considerations come from two aspects of the behavior. Firstly, smoking cigarettes is deleterious for smokers health and societies have to bear enormous costs arising from the lost of productivity and from the medical expense associated with smoking induced diseases. Therefore, explaining the factors that influence cigarette smoking has important implication for intervention policies. Secondly, the smoking behavior itself is interesting. It is puzzling why individuals maintain to participate in a seemingly devastating habit although most of them are conscious of the harmful consequences of their addictive nature.

It is quite often in economics that interest lies in modeling the factors that affect number of cigarette smoking per day. There seems to consensus in the literature regarding those factors. Many studies have suggested almost the same determinates of tobacco consumption, however their outcome is deferent based on the conducted econometric technique. For instance, in the literature of smoking, It is well known that more males than females smoke. According to Jha et al. (2002), nearly 47% of all men are smoker, but only 11% of all woman smoke. (see, for instance, Waldron, 1991; and Moghimbeigi et al., 2009). The psychological researchers, for example Waldron, 1991, conclude that gender differences in smoking cigarettes are mainly due to different behavior, its roots come from traditional sex roles. Whereas, Bauer et al., (2006) explains gender differences in Germany by firstly the differences in socioeconomic characteristics between males and females and secondly it is due to differences in coefficients indicating substantial differences in the smoking behavior between men and women rather than differences in characteristics.

An empirical research about marital status and smoking in Korea by Cho et al. (2008) finds that the prevalence of smoking is lower among the married individuals comparing to the unmarried. Moreover, in the US, Hersch (2000) detects that married individuals have fewer propensity to smoke.

According to Bauer et al. (2007) employed individuals have higher probability to smoke. It can be explained by the fact that employed individuals have kind of job stress and also they are more financially independent as compared to unemployed individuals. For instance, Ayyagari and Sindelar (2010) declare that job stress is positively related to continuing to smoke and to the number of cigarettes smoked for current smokers. However, Manrique and Jensen (2004) uses the households data in Spain to investigate the influencing factors of both smoking and drinking behavior, and indicate that household heads who are currently being employed are less probable to smoke than the unemployed. Moreover, Moghimbeigi et al (2009) employing data from the National Health Survey of Iran, by applying a Zero Inflated Poisson model concluded that the unemployed adults are more at risk of smoking.

Previous studies have consistently found that education has a significant impact on smoking. Many studies have concluded that education is negatively associated with the probability to smoke (see, for instance, Aristei &

Pieroni, 2008; Bilgic et al., 2010; Lin, 2010). In other words, individuals with higher educational background are less likely to smoke as a result of knowledge of the adverse effects of smoking.

3. Methodology

The purpose of this study is to determine the most important factors that affect the number of smoking cigarette in the City of Sulamanyah in 2015. The dependent variable of this study is the number of smoking cigarette per day, for the majority of result, with explanatory variables being age, education level, marital status, income, living with, the influence of friends, having children, and health status.

3.1 Sample

Sample was composed of 650 individuals and employees. 561 surveys questionnaires were returned, and the total response rate for this paper was 86.3%. Of the respondents, 84.0% were male, 16.0 % were female; 0.0 % was less than 15 years old, 4.1 % were 15-25 years old, 47.4 % were 25-35 years old, 30.5 % were 35-45 years old, 10.2 % were 45-55 years old, 7.8 % was beyond 55 years old. The education level of the respondents varied: 4.1 % was illiterate, 10.5 % had primary level, 12.7 % had secondary level, 16.4 % had high education, 19.6 % had a diploma's degree, 32.4 % had a bachelor's degree, and 4.3 % had a high level's degree. The respondents "marital statue" as follow: 46.7 % were single person and 53.3 % were married. The below table provides summery statistics about all variables.

Variables	Definition	Mean <u></u> SD	Obs.
Age	Age of participants in year	34.3 ± 6.5	561
Age2	Age squared		
Education	Education level of participants	3.5 ± 1.59	561
Maritalsta	Marital status	1.53 <u>+</u> 0.49	561
income	Monthly income	1.8±0.79	561
Livewith	Participants living with family, spouse, or alone	2.4 ± 0.61	561
Friends	Friend influences in participation decisions to	1.59±0.41	561
	smoke		
Children	Having children at participants' home	1.52 ± 0.52	561
Health	Health status of participants	1.56±0.76	561
Smoking	Number of smoking cigarettes per day	15±10.4	561

Table (1): Summary Statistics

3.2 Data collection

The data have been collected from both smokers and nonsmokers through web-based survey emailed to all possible respondents in the city of Sulamanyah and also a random section of completed questionnaires have been selected in order to be analyzed. The final result will be seen by all respondents as part of this report upon permission from following publication. The primary research will be used to obtain all results and statistical software such as STATA- program; SPSS and Excel will be used to derive the final conclusion.

4. Data Analysis & Results

The paper presents the application of both Poisson and negative binominal (NB) regression modules for determinants of smoking cigarette.

From the Stata outcome for both models in tables (2) and (3), it can be seen that, all regressors in Poisson and NB are jointly significant at 5% level because their prob>chi2 statistic tests are less than 0.05. Furthermore, all coefficients in both models are individually significant at 5% level because their P values are less than 0.05, except income and living with variables in both models is not statistically significant at 0.05 levels, even though, the coefficient of having children in NB model is significant at 10% level. Additionally, choosing between negative binomial model and Poisson regression model relies on the nature of the distribution of the response variable; negative binomial regression has been commonly selected by researchers because its assumption are observed with social data. Though, Poisson regression model are far from non-existent with several researches

even seeing presence of both NB regression and Poisson regression model within the same study (Braga & Bond, 2008).

Table 2 and table 3 displays the results of Poisson and negative binomial regression models and the impact of the explanatory variables on the response variable can be practically determined by the regression coefficient. Furthermore, the techniques of count regression model are the log of incident count; the coefficient in both methods can be interpreted as follows: for a one unit change in the explanatory variable, the log of response variable is predicted to change by the value of the coefficient.

Iteration 0: log likelihood		-1111.6028		Number of obs		561	
Iteration 1: log likelihood		-1111.6015		LR chi2 (9)		242.58	
Iteration 2: log likelihood		-1111.6015		Prob >chi2		0.000	
Log likelihood		-1111.6015		Pseudo R^2		0.6925	
No.Smoking	Coef	Std.Err	Z	P>/Z/ [95% Cof.		Interval	
Age	.2913	.1575	6.55	0.000	.7231	1.3407	
Age^2	0064	.0237	-5.08	0.000	16687	0739	
Education	03564	.01777	-2.01	0.044	-07034	00094	
Marital status	44186	.06901	-6.40	0.000	57713	3066	
Income	04705	.03656	-1.29	0.198	1187	.02460	
Live with	.0716	.0474	1.51	0.131	0214	.1647	
Friends	4466	.05004	-8.92	0.000	5447	3485	
Children	1041	.0530	-1.96	0.049	20813	00025	
Health	.0369	.01439	2.57	0.010	.00087	.06516	
Cons	.73598	.33004	2.23	0.026	.089119	1.3828	

Table (2): Poisson Regression Model

Iteration 0: log likelihood		-1131.2646		Number of obs		561	
Iteration 1: log likelihood		-1114.9543		LR chi2 (9)		190.68	
Iteration 2: log likelihood		-1111.8636		Prob >chi2		0.000	
Iteration 3: log likelihood		-1111.3387		Pseudo R^2			
Iteration 4: log likelihood		-1111.2876					
Iteration 5: log likelihood		-1111.2861				0.6925	
Iteration 6: log likelihood		-1111.2801					
Log likelihood		-1111.2861					
No.Smoking	Coef	Std.Err	Z	P>/Z/ [95% Cof. In		nterval	
Age	.28910	.16789	6.32	0.000	.73207	1.3900	
Age^2	00616	.02513	-4.94	0.000	17357	07502	
Education	0365	.01844	-1.98	0.048	072667	00036	
Marital status	45137	.07296	-6.19	0.000	59439	30835	
Income	-04782	.03791	-1.26	0.207	12213	.02648	
Live with	.07252	.049624	1.46	0.144	024739	.16978	
Friends	45151	.05221	-8.65	0.000	55384	34917	
Children	10303	.055039	-1.87	0.061	-2.1091	.004835	
Health	.038400	.015120	2.54	0.014	.00875	.068051	
Cons	.70857	.34423	2.06	0.040	.03300	1.3832	
Inalpha	2.5321	.033122			2.3541	2.7100	
alpha	12.6427	.02988			11.9900	13.298	
Likelihood-ratio test for alpha=0: chibar2 (01): 1.e+0.5 prob>chibar2=0.000							

Table (3): Negative Binomial (NB) Regression Model

The result for both methods starts with the coefficient of age and age square. The coefficients of age in both methods are approximately 0.289. This means that the number of smoking cigarette will increase by 28.9% per day if age increases by one year, but at a decreasing rate of 0.0064% in Poisson and by 0.0061 in NB, holding all other variables constant. This might be because old people may be more careful about their health. Moreover, income and living with coefficients in both models are not statistically significant because their p-values are greater than the common alpha level 0.05. The result of income might be logically acceptable because the price of cigarette is cheap and most people can afford it. However, the result of living with coefficient is unpredictable because it is expected that lonely individuals smokes more than people who live with families and children.

It is obvious that the remaining regressors from both tables (2) and (3) are Dummies, which require using exp (Bi)-1 for interpreting their coefficients. For example, the coefficient of education level in Poisson regression model shows that if education level promotes by one unit then the smoking cigarette will decrease by 3.431% and by 3.585% in NB regression, ceteris paribus. Meaning that, people with degrees smoke less than people without degrees because the educated people may have a good knowledge about the drawbacks of smoking and it has an impact on educated people to reduce smoke or quit it.

The influences of friend variable have a negative impact on smoking in both regression models but it has a different coefficient. In negative binomial regression, friend influences have affected in decreasing the number of smoking by 36.33% and by 36.02% in Poisson regression after holding all other variables constant. However, the variables of health status in Poisson regression model and negative binomial regression are unexpected; the

result declares that it has a positive effect on smoking. Means that having illnesses in NB regression increase the number of smoking cigarette by 3.91% and by 3.76% in Poisson regression model and ceteris paribus.

The marital status and having children have a negative impact on smoking in both models. In more details, the number of smoking cigarette will fall by10.97 % in Poisson and by 9.78% in negative binomial regression model after holding all other regressors constant and if the number of children increase by one unit then the number of smoking cigarette will fall by 36.32% in NB regression and by 35.68% in Poisson regression model after holding all other independent variables constant. According to (Freundet al, 1992) is that the marital status has a negative effect on reducing the number of smoking cigarette and quit it.

The tables (4) and (5) demonstrate the average of marginal effects in Poisson and negative binominal regression models respectively. The marginal effect usually measures the effect of changing the explanatory Variable upon response variable.

No.Smoking	Dy/dx	Delta-Method Std. Erre	Z	P>/Z/	[95% Cof. Interval	
Age	.51125	0.804	7.41	0.000	0.55288	0.9096
Age^2	00870	8.04	-11.49	0.000	.55288	.90961
Education	05719	.05190	-1.10	0.270	15892	.0445
Marital status	-1.2104	.20510	-5.90	0.000	-1.6024	80847
Income	08053	.10703	75	0.452	29033	.12925
Live with	.18406	.139812	1.32	0.188	089960	.45809
Friends	-1.3642	.151306	-9.02	0.000	-1.6608	-1.0677
Children	33775	.155658	-2.17	0.030	64283	03266
Health	.114736	.042538	2.70	0.007	.03136	.19810

Table (4): Margins in Poisson Regression

Table (5): Margins in Non-Binomial Regression

No.Smoking	Dy/dx	Delta-Method Std. Erre	Z	P>/Z/	[95% Cof. Interval	
Age	.48430	.50448	6.21	0.000	2.14552	4.1230
Age^2	00850	.07507	-4.89	0.000	51429	22002
Education	10786	.05457	-1.98	0.048	21483	00088
Marital status	-1.3333	.21904	-6.09	0.000	-1.7626	90399
Income	14127	.11206	-1.26	0.207	360909	.078363
Live with	.21422	.146706	1.46	0.144	073315	.5017
Friends	-1.3337	.158772	-8.40	0.000	-1.64490	-1.0225
Children	30436	.16277	-1.87	0.061	-0.6233	.01466
Health	.11343	.044823	2.53	0.011	.025578	.201283

It is preferred to interpret one qualitative variable and one quantitative variable in both models as an example to show the marginal effects. To beginning with the coefficient of age, the cigarette smoking rises by 0.5112 per day but at decreasing rate of 0.0087 if age increase by one year in Poisson regression and it rises by 0.4843 at a decreasing rate of 0.0085 in NB regression model after holding all other variables constant. In addition, education level variable was positively affected on decreasing the number of smoking cigarette by 0.057 in Poisson regression and by 0.1078 in negative binomial regression and ceteris paribus.

The result of p-value demonstrates statistical significance for each variable. The most of variables were statistically significance in Poisson regression because p-value for those variables are less than the common alpha level 0.05 but several variables were statistically significance at 0.05 level in NB regression. Thus, the negative binomial regression model is preferred.

5. Conclusion

This study has attempted to explain the determinants of cigarette smoking in the city of Sulaimanyah using two count data econometric methods these ;the Poisson and the negative binomial regression models. Fitting both models, it was found that the Negative Binomial provided the best fit for the data.

Using 561 surveys questionnaires, the findings in this study show that age, education, marital status, friend influences, having children, and health status are very important determinants of cigarette smoking; aged persons have less prevalent smoking compared to the young people, the probability of smoking decrease with increased education, married individuals have lees chance of smoking, friends have negative impact on the number of smoking cigarettes, having children decreases the likelihoods of smoking. Unexpectedly, health status has positive influences in smoking; meaning that people with diseases smokes more than healthier individuals. It is also noticeable that, the coefficients in living with and income variables are not statistically significant in both Poisson and the NB regression models. For further research that would be important to participate the other cities to provide a detailed analysis of the determinants of cigarette consumption in Kurdistan region.

Moreover, we find that income has no impact on the participations of smoking because its price is quit cheap, indicating that anti-smoking policies can be more effective if they take this into account by imposing more taxes and quotas on cigarettes.

It is suggested from the study that smoking prevention programs should increase the younger's knowledge of high risks of health problems associated with exposure to tobacco smoke through media programs as well as work on awareness prevalence among adults especially in schools and universities. It is also significance to pay special attention to develop and implement effective tobacco control strategies such as price increase and raising taxes on cigarette importation in Iraq.

Reference

[1] American Cancer Society (2014) " Cancer Facts and Figures". *American Cancer Society*, 500814. Available from: www.cancer.org/downloads/STT

[2] Alghabban, S., I. (2009) " Prevalence of Current Smoking among Students in University of Kerbala". *Karbala journal of Medicine*, 2(8), PP. 645-662.

[3] Aristei, D. and Pieroni, L. (2008) " A double-hurdle approach to modeling tobacco consumption in Italy". Applied Economics, *40*, PP. 2463-2476.

[4] Ayyagari, P. and Sindelar, Jody L. (2010) " The Impact of Job Stress on Smoking and Quitting: Evidence from the HRS" *The B.E. Journal of Economic Analysis & Policy*, *10*(1), PP. 1-32.

[5] Bauer, T., Gohlmann, S., and Sinning, M. (2007) "Gender differences in smoking behavior". *Health Economics*, *16*, PP. 895-909.

[6] Bennett, WP., Alavanja, MC., Blomeke, B., Vahakangas, KH., Castren, K., Welsh, JA., and et al. " Environmental tobacco smoke, genetic susceptibility, and risk of lung cancer in neversmoking women". *Journal of the National Cancer Institute*, *91*, PP. 2009-2014.

[7] Bilgic, A., Florkowski, WJ., and Akbay, C. (2010) " Demand for cigarettes in Turkey: An application of count data models ". *Empirical Economics*, *39*, PP. 733-765.

[8] Braga, A., and Bond, B. (2008) "Policing crime and disorder hot spots: A randomized controlled trail ". *Criminology*, *46*(*3*), PP 577-607.

[9] Cameron, C., and Trivedi, P. (1998) *Regression Analysis of Count Data*. Cambridge University Press, New York.

[10] Cho, HJ., Khang, YH., Jun, HJ., and Kawachi, I. (2008) "Marital Status and smoking in Korea: The influence of gender and age" *Social Science and Medicine*, *66*, PP.609-619.

[11] Farrell, L., Tim R. L. F., and Harris, M. N. (2003) "A Pack A Day For Twenty Years: Smoking And Cigarette Pack Sizes," Department of Economics - Working Papers Series 887, The University of Melbourne.

[12] Freund, K. M. at el. (1992) " Predictors of Smoking Cessation: The Framingham Study ". American Journal of Epidemiology, 135, PP. 957-964.

[13] Garcia, J. and Labeaga, JM. (1996) "Alternative approaches to modelling zero expenditure: an application to Spanish demand for tobacco". *Oxford Bulletin of Economics and Statistics* 58 (3), PP.489-506

[14] Gruber, J. and Zinman, J. (2000) "Youth Smoking in the U.S.: Evidence and Implications", NBER Work Paper Series No. 7780.

[15] Hersch, J. (2000) "Gender, income levels, and the demand for cigarettes". *Journal of Risk Uncertainty*, 21, PP. 263-282.

[16] Jha, P., Ranson, M. K., Nguyen, S. N., and Yach, D. (2002) "Estimates of Global and Regional Smoking Prevalence in 1995, by Age and Sex". *American Public Health Association*, *92*(6), PP. 1002-1006.

[17] Keeler, TE., Hu, TW., Barnett, PG., and Manning, WG. (1993) "Taxation, regulation, and addiction: a demand function for cigarettes based on time-series evidence". *Health Economics*, *12*, PP. 1–18.

[18] Khwaja, A., Sloan, F., and Chung, S. (2006) "Learning about individual risk and the decision to smoke". *International Journal of Industrial Organization* 24(4), PP. 683-699.

[19] Lin, SJ. (2010) "Estimating the determinants of smoking behavior in Taiwan". *Substance Use & Misuse*, 45, PP. 482-95.

[20] Manrique, J. and Jensen, HH. (2004) "Consumption of tobacco and alcoholic beverages among Spanish consumers". *Southwestern Economic Review*, 31, PP. 41-56.

[21] Moghimbeigi, A., Eshraghian, MR., Mohammad, K., Nourijelyani, K., Nourijelyani, M. (2009) "Determinants Number of Cigarette Smoked with Iranian Adolescents: A Multilevel Zero Inflated Poisson Regression Model". *Iranian Journal of Public Health*, *38*, PP. 91-96.

[22] Muhammad, A. and Ahmad, I. (2010) "Panel Count Data Analysis of Determinants of Cigarette Smoking: Evidence from British Household Panel Survey, 2001-2009". International Journal of Business and Economics, 1(1), PP. 10-23.

[23] Mullahy, J. (1997) "Instrumental-Variable Estimation Of Count Data Models: Applications To Models Of Cigarette Smoking Behaviour" *The Review of Economics and Statistics, 794, PP. 586-593.*

[24] Taylor, D. H., Hasselblad, V., Henley, S.J., Thun, M.J., and Sloan, F.A. (2002) "Benefits of Smoking Cessation for Longevity". *American of Public Health association*, *92(6)*, PP. 990-996.

[25] Waldron, I. (1991) "Patterns and causes of gender differences in smoking". *Social Science and Medicine*, *32*(9), PP. 989-1005.

[26] Wasserman, J., Manning, W., Newhouse, J., and Winkler, J. D. (1991) "The effects of excise taxes and regulations on cigarette smoking". *Health Economics*, *10*(*1*), PP. 43–64.

[27] World Health Organization (2008) "WHO Report on the Global Tobacco Epidemic". *The MPOWER Package.Geneva: World Health Organization*.

[28] Yen, S. T. (2005a) "Zero observations and gender differences in cigarette consumption". *Applied Economics*, *37*, PP.1839–1849.