Supplementary materials

# Supplementary material A:

# Development and Estimation of the zero-inflated Negative Binomial regression model

In the zero-inflated Negative Binomial setup, each observation can fall into one of two cases. Case 1 occurs when the value of *N\_smoke* (representing the count of cigarettes smoked per day) is inherently 0, in which case the observation enters the zero-inflated part of the model. Case 2 arises when the values of *N\_smoke* (including zeros) are generated according to the Negative Binomial part of the zero-inflated Negative Binomial model. Assuming that Case 1 and Case 2 occur with probabilities and 1-, respectively, the probability distribution of the outcome variable can be written as:

(A.1)

Where is the logistic link function defined below; is the distribution function of the Negative Binomial distribution defined by:

(A.2)

Where Γ is the gamma function with α being the over-dispersion coefficient:

, (A.3)

, (A.4)

. (A.5)

The expression of in (A.2) is related to the following quantities involved in equation (1) in the main text:

= exp(, (A.6)

The regression coefficients , , …, are parameters to be estimated in the Negative Binomial part.

The logistic link function in (A.1) is given by:

(A.7)

Where

. (A.8)

The coefficients , , …, are parameters to be estimated in the zero-inflated part. The parameters ’s and ’s can be jointly estimated by Maximum Likelihood Estimation (MLE) techniques [29, 30].

# Supplementary material B:

Supplementary Table 1. Pearson Correlations between explanatory variables.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | N\_  smoke | Self\_  NCD | Fam\_  NCD | Fam\_  smoke | Age\_  yrs | Minority | Rural | Educ\_  yrs | HH\_  income | Working | Health\_  insurance | Drinking | Self\_  health | N\_smoke  \_lag1 |
| N\_smoke | 1.00 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Self\_NCD | −0.12 | 1.00 |  |  |  |  |  |  |  |  |  |  |  |  |
| Fam\_NCD | −0.06 | 0.15 | 1.00 |  |  |  |  |  |  |  |  |  |  |  |
| Fam\_smoke | 0.06 | 0.002 | 0.19 | 1.00 |  |  |  |  |  |  |  |  |  |  |
| Age\_yrs | −0.21 | 0.25 | 0.08 | 0.049 | 1.00 |  |  |  |  |  |  |  |  |  |
| Minority | 0.02 | −0.05 | −0.05 | 0.04 | 0.0003 | 1.00 |  |  |  |  |  |  |  |  |
| Rural | 0.11 | −0.13 | −0.06 | 0.07 | −0.08 | 0.11 | 1.00 |  |  |  |  |  |  |  |
| Educ\_yrs | −0.03 | 0.02 | −0.01 | −0.10 | −0.3 | −0.07 | −0.32 | 1.00 |  |  |  |  |  |  |
| HH\_income | −0.03 | 0.10 | −0.04 | −0.06 | −0.04 | −0.09 | −0.19 | 0.34 | 1.00 |  |  |  |  |  |
| Working | 0.18 | −0.22 | −0.16 | −0.03 | −0.51 | 0.07 | 0.22 | 0.07 | 0.03 | 1.00 |  |  |  |  |
| Health\_insurance | −0.03 | 0.13 | 0.08 | −0.07 | 0.08 | −0.09 | −0.16 | 0.20 | 0.35 | −0.1 | 1.00 |  |  |  |
| Drinking | 0.20 | −0.08 | −0.04 | 0.03 | −0.21 | 0.01 | 0.07 | 0.06 | 0.03 | 0.17 | −0.01 | 1.00 |  |  |
| Self\_health | 0.04 | −0.14 | −0.04 | −0.01 | −0.16 | −0.03 | −0.04 | 0.13 | 0.10 | 0.12 | −0.01 | 0.11 | 1.00 |  |
| N\_smoke\_lag1 | 0.69 | −0.11 | −0.05 | 0.05 | −0.20 | 0.02 | 0.11 | −0.03 | −0.04 | 0.14 | −0.04 | 0.15 | 0.03 | 1.00 |

Notes: Source: Author’s computation using data from the CHNS (2000, 2004, 2006, 2009, 2011, 2015).

Supplementary Table 2. Zero-inflated negative binomial estimates of associations between one’s own different specific chronic diseases, home environment, and smoking behavior of Chinese adult males (aged 45–85).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model (1) | | Model (2) | | Model (3) | | Model (4) | |
| Models | If smoking  (Logit Part) | Number smoked  per day  (Negative  Binomial) | If smoking  (Logit Part) | Number smoked  per day  (Negative  Binomial) | If smoking  (Logit Part) | Number smoked  per day  (Negative  Binomial) | If smoking  (Logit Part) | Number smoked  per day  (Negative  Binomial) |
| Sample | Never smokers dropped | | Never smokers dropped | | Never smokers dropped | | Never smokers dropped | |
| Hypertension | −0.170\*  (0.097) | −0.002  (0.025) |  |  |  |  |  |  |
| Diabetes |  |  | −0.268\*  (0.149) | −0.072  (0.047) |  |  |  |  |
| Stroke |  |  |  |  | −0.410\*  (0.216) | −0.057  (0.065) |  |  |
| Myocardial infarction |  |  |  |  |  |  | −0.725\*\*\*  (0.270) | 0.068  (0.101) |
| Fam\_chron | −0.108  (0.075) | 0.018  (0.017) | −0.051  (0.068) | 0.015  (0.015) | −0.073  (0.069) | 0.009  (0.016) | −0.078  (0.067) | 0.017  (0.015) |
| Fam\_smoke | 0.332\*\*\*  (0.080) | 0.025  (0.018) | 0.278\*\*\*  (0.075) | 0.032\*\*  (0.016) | 0.282\*\*\*  (0.075) | 0.027  (0.017) | 0.275\*\*\*  (0.074) | 0.030\*  (0.016) |
| Hypertension × Fam\_chron | 0.154  (0.137) | −0.016  (0.038) |  |  |  |  |  |  |
| Diabetes × Fam\_chron |  |  | −0.090  (0.209) | 0.011  (0.082) |  |  |  |  |
| Stroke × Fam\_chron |  |  |  |  | 0.229  (0.277) | 0.137\*  (0.076) |  |  |
| Myocardial infarction × Fam\_chron |  |  |  |  |  |  | 0.491  (0.384) | −0.246\*  (0.130) |
| Hypertension × Fam\_smoke | −0.218  (0.145) | 0.028  (0.041) |  |  |  |  |  |  |
| Diabetes × Fam\_smoke |  |  | −0.036  (0.235) | −0.023  (0.097) |  |  |  |  |
| Stroke × Fam\_smoke |  |  |  |  | −0.170  (0.306) | 0.139\*  (0.083) |  |  |
| Myocardial\_infarction× Fam\_smoke |  |  |  |  |  |  | 0.169  (0.328) | 0.069  (0.128) |
| Age\_yrs | −0.099\*\*\*  (0.035) | 0.023\*\*\*  (0.009) | −0.101\*\*\*  (0.035) | 0.023\*\*\*  (0.009) | −0.105\*\*\*  (0.035) | 0.024\*\*\*  (0.009) | −0.105\*\*\*  (0.035) | 0.022\*\*\*  (0.009) |
| Age\_sq | 0.074\*\*\*  (0.028) | −0.026\*\*\*  (0.007) | 0.076\*\*\*  (0.028) | −0.026\*\*\*  (0.007) | 0.079\*\*\*  (0.028) | −0.027\*\*\*  (0.007) | 0.078\*\*\*  (0.028) | −0.025\*\*\*  (0.007) |
| Minority | 0.016  (0.108) | −0.008  (0.029) | 0.025  (0.109) | −0.008  (0.029) | 0.018  (0.108) | −0.009  (0.029) | 0.010  (0.108) | −0.008  (0.029) |
| Rural | −0.201\*\*  (0.084) | 0.054\*\*\*  (0.020) | −0.202\*\*  (0.084) | 0.054\*\*\*  (0.020) | −0.192\*\*  (0.084) | 0.055\*\*\*  (0.020) | −0.203\*\*  (0.084) | 0.055\*\*\*  (0.020) |
| Educ\_yrs | 0.005  (0.010) | −0.007\*\*\*  (0.002) | 0.006  (0.011) | −0.006\*\*\*  (0.002) | 0.004  (0.011) | −0.007\*\*\*  (0.002) | 0.004  (0.011) | −0.007\*\*\*  (0.002) |
| HH\_Income | −0.043  (0.031) | 0.009  (0.006) | −0.045  (0.031) | 0.009  (0.006) | −0.045  (0.031) | 0.009  (0.006) | −0.045  (0.031) | 0.009  (0.006) |
| Working | 0.240\*\*\*  (0.072) | 0.049\*\*\*  (0.017) | 0.248\*\*\*  (0.073) | 0.048\*\*\*  (0.017) | 0.241\*\*\*  (0.072) | 0.048\*\*\*  (0.017) | 0.239\*\*\*  (0.073) | 0.049\*\*\*  (0.018) |
| Health\_Insurance | 0.061  (0.084) | 0.045\*\*  (0.019) | 0.038  (0.085) | 0.045\*\*  (0.019) | 0.051  (0.084) | 0.045\*\*  (0.019) | 0.059  (0.084) | 0.045\*\*  (0.019) |
| Drinking | 0.843\*\*\*  (0.065) | 0.014  (0.014) | 0.842\*\*\*  (0.065) | 0.014  (0.014) | 0.837\*\*\*  (0.065) | 0.015  (0.014) | 0.836\*\*\*  (0.065) | 0.014  (0.014) |
| Self\_health | 0.112\*\*  (0.055) | 0.001  (0.012) | 0.112\*\*  (0.054) | −0.001  (0.012) | 0.109\*\*  (0.054) | 0.002  (0.012) | 0.114\*\*  (0.054) | 0.001  (0.012) |
| N\_smoke\_lag1 | 0.090\*\*\*  (0.004) | 0.024\*\*\*  (0.001) | 0.090\*\*\*  (0.004) | 0.024\*\*\*  (0.001) | 0.090\*\*\*  (0.004) | 0.024\*\*\*  (0.001) | 0.090\*\*\*  (0.004) | 0.024\*\*\*  (0.001) |
| Province dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Survey-wave dummies | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | 3.425\*\*\*  (1.082) | 1.746\*\*\*  (0.272) | 3.519\*\*\*  (1.077) | 1.734\*\*\*  (0.273) | 3.634\*\*\*  (1.083) | 1.712\*\*\*  (0.271) | 3.654\*\*\*  (1.084) | 1.763\*\*\*  (0.272) |
| *N* | 9737 | | 9725 | | 9731 | | 9734 | |
| Log pseudolikelihood | −29,461.859 | | −29,431.678 | | −29,442.654 | | −29,451.007 | |

Notes: Source: Author’s computation using data from the CHNS (2000, 2004, 2006, 2009, 2011, 2015).

Estimated coefficients are reported in the table. Standard errors are reported in parentheses.

\* *p* < 0.1, \*\* *p* < 0.05, \*\*\* *p* < 0.01.