APPENDIX A

The recommended policies designed to increase health inducing behavior for Thailand

Kanchit Suknark[1], Jirakom Sirisrisakulchai[2] and Songsak Sriboonchitta[3]

Abstract

Global health care costs are rising and become a big problem for developing countries. Health promotion and disease prevention are considered as the approaches to reduce health care costs. However, the factors affecting health behaviors have to be well specified before considering the optimal health promotion and prevention policy. Investing in lower cost-effectiveness policies may not give the economic benefit as expected. This paper proposes a new method to analyze the factors affecting health behaviors by accounting for the dependence between each health behavior. The proposed model will give more efficient parameter estimates in comparison with the traditional models. Moreover, understanding the dependencies between choices for each health behaviors gives useful information for designing more efficient health promotion and disease prevention programs. The data from Thai National Health Examination Survey were used to analyze the factor affecting the physical activity, tobacco consumption and alcohol consumption behaviors, simultaneously. The empirical results show the negative correlation between tobacco consumption and physical activity behaviors. We also found the positive correlation between alcohol consumption and physical activity behaviors. Finally, from the empirical results, the recommended policies designed to reduce health-risk behavior and increase health inducing behavior for Thailand are discussed in the paper.

Keywords

Labor, Health risk behavior, Policies, Health behavior, Health care costs, Thailand

1. Introduction

Global health care costs are rising and become a big problem for developing countries. Health promotion and disease prevention are considered as the approaches to reduce health care costs.

---

[1] Phd. Student in Economics, Faculty of Economics, Chiang Mai University

[2] Dr., Faculty of Economics, Chiang Mai University

[3] Prof.Dr., Faculty of Economics, Chiang Mai University

The explicit burden on society due to health-risk behaviors, particularly alcohol and tobacco consumption, includes health care costs, productivity loss, property damage costs, costs of criminal justice as well as law enforcement. To reduce health-risk behaviors, Thailand should aim to reduce alcohol consumption and prevent initiation of drinking. While Thailand already implements alcohol related policies, such as high alcohol taxation, restricted alcohol sale times, more effective measures at the societal level to control alcohol consumption and alcohol-related harms are still required. The national survey in 2011 reported that about 17.7 million people or 20.8% of the population aged 15 years and over are alcohol users. Men used alcohol at a higher rate than women (NSO, 2011). Equally, tobacco consumption control policies have been implemented to reduce tobacco consumption and prevent initiation of smoking, especially in younger people. Current policies include high rates of tobacco taxation, control of tobacco advertising, non-smoking areas and bans on smoking in public places. These policies have been shown to be successful in decreasing the proportion of smokers in the Thai population (aged 15 years and older) from 32 % in 1991 to 20% in 2013 (NSO, 2011).

On the other hand, the Thai Health Promotion Foundation has promoted physical activity in the Thai population by sponsoring and supporting several public campaigns nationally on the benefits of physical activity and advising people on the effective levels of frequency, duration and intensity required to achieve physical fitness since 2010. Such programs have also been supported at the local and regional level in many areas of the country. Most of the projects are mainly focused on increasing perceptions, attitudes, and practices related to physical activity generally (Katewongsa et al, 2014).

The previous studies on the factors affecting alcohol consumption, tobacco consumption, and physical activity were based on a single equations (Katewongsa et al, 2014) (Suwannashote P.,2009) (Praponsin S., 2007) and (Sirirassamee T., 2009).

In this paper, we simultaneously determined the factors affecting including alcohol-consumption, tobacco-consumption and physical activity to quantify the factors affecting health behaviors have to be well specified before considering the optimal health promotion and prevention policy. Investing in lower cost-effectiveness policies may not give the economic benefit as expected and attempted to quantify the dependence measures between these pairs using the copula approach. This paper proposes a new method to analyze the factors affecting health behaviors by accounting for the dependence between each health behavior. The proposed model will give more efficient parameter estimates in comparison with the traditional models. Moreover, understanding the dependencies between choices for each health behaviors gives useful information for designing more efficient health promotion and disease prevention programs.

2. Data

The data used in this study are from the Thai National Health Examination Survey, No.4 (NHES IV) from 2009. The data consists of a sample of 20,450 individuals. The ordered dependent variables are alcohol consumption (Y1), tobacco smoking (Y2), and physical activity in leisure time (Y3). The independent variables are sex, age, income, chronic diseases, marital status, education level, and occupation. The alcohol consumption variable (Y1) were: 0 for non-alcohol consumption; 1 for alcohol consumption. For the tobacco consumption variable (Y2) were: 0 for non-smoking; 1 for smoking. For the physical activity variable (Y3), the levels of physical activity or exercise in leisure time were: 0 for non-physical activity; 1 for physical activity. Table 1. presents a description of the variables and related statistics.

Table 1 Main statistics and description of variables

variables	Description	N	Mean	SD	Min.	Max.
Y1	1 if individual consume alcohol; 0 otherwise	20450	0.446	0.627	0	1
Y2	1 if individual consume tobacco; 0 otherwise	20450	0.052	0.339	0	1
Y3	1 if individual has physical activity in leisure time; 0 otherwise	20450	2.201	0.845	0	1
Sex	1 if individual is male; 0 otherwise	20450	0.524	0.499	0	1
Age	in Year	20450	52.917	18.236	14	98
Income	in 1,000 Baht	20450	3.310	5.698	0	32,480
bachelor	1 if individual graduated from Bachelor degree or higher; 0 otherwise	20450	0.061	0.24	0	1
Age	1 if individual works in agricultural sector; 0 otherwise	20450	0.176	0.381	0	1
Whi	1 if individual is white-collar worker	20450	0.035	0.184	0	1
police	1 if individual works as police or soldier; 0 otherwise	20450	0.012	0.108	0	1
Labor	1 if individual is in labor sector; 0 otherwise	20450	0.48	0.499	0	1
Married	Marital status where 1 indicates married; 0 otherwise	20450	0.636	0.481	0	1
Pe_bmi25	1 if individual has body mass index more than 25; 0 otherwise	20450	0.348	0.476	0	1
Pe_tc200	1 if individual has chloresterol level more than 200; 0 otherwise	20450	0.561	0.496	0	1
qlhealth	Self health quality assessment, where 5 is the highest level	20450	3.708	0.867	0	5
NCD	Number of chronic diseases	20450	0.632	0.959	0	10

3. Methods

3.1 Binary Choice Models

The health behaviors that we are going to analyze are binary choices of alcohol consumption, tobacco consumption, and physical activity. Let $Y_i = \{0,1\}$,for i = 1, 2, and 3, be the binary choices, where 1 , 2, and 3 indicate alcohol consumption, tobacco consumption, and physical activity choices, respectively. For the set $\{0,1\}$ of each choice outcome, 0 indicates individual has not chosen that choice and 1 indicates individual has chosen that choice. Let X_i ,for i = 1, 2, and 3,where 1, 2, and 3, are previously defined, be the vector of all the explanatory variables thought to explain health behaviors, and β_i are the vector of the parameters to be estimated corresponding to X_i. For each Y_i, we can think of this random variable as being generated from the following binomial distribution:

$$f(k; n, p) = \Pr(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ for } k = 0,1,2,\ldots,n, \qquad (3.1)$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, in which n is the number of trials and p is the probability of choosing this choice. Note that we drop the subscript i for the sake of simplicity.

In our case, we assume that individual make choice only once, hence $n=1$ (which is called Bernoulli distribution), and if p is assumed to be a standard normal distribution (Φ), we can derive the probability distribution as follows:

$$\Pr(Y=1|X,\beta) = \Phi(X\beta), \quad (3.2)$$

$$\Pr(Y=0|X,\beta) = 1-\Phi(X\beta) \quad (3.3).$$

The above marginal distribution model has exactly the same functional form as probit model. We introduce the idea of viewing binary choice model as derivation from binomial distribution in contrast with derivation from latent continuous variable to model the multivariate joint distribution using discrete copula in the next section.

3.2 Copula

Copula is increasing popularity in many fields of application especially in economics. The readers will refer to Joe (1997) and Nelsen (2006) for the introduction of copula theory and some applications. The examples of the application in health econometric models can be found in Sirisrisakulchai and Sriboonchitta (2014a, 2014b) and Kanchit et al. (2016a, 2016b).

The fundamental theorem of copula is Sklar's theorem (1959). The theorem briefly states that there exists a copula function C such that

$$F(y_1, y_2, ..., y_m) = C(F_1(y_1), F_2(y_2), ..., F_m(y_m)) \quad (3.4)$$

where $\mathbf{y} = (y_1, y_2, ..., y_m)$ is the realization of an m-dimensional random vector $\mathbf{Y} = (Y_1, Y_2, ..., Y_m)$, F is a joint distribution function, and F_j is a marginal distribution function corresponding to each marginal j. Sklar's theorem provides us the link between the dependence structure of multivariate distribution and their univariate margins.

For continuous variables, the joint density $f(y_1, y_2, ..., y_m)$ can be easily obtained by taking the derivative of both the sides of equation (3.4), which gives

$$f(y_1, y_2, ..., y_m) = c(F_1(y_1), F_2(y_2), ..., F_m(y_m)) f_1(y_1) f_2(y_2) ... f_m(y_m),$$

where f is a joint density function, f_j is a marginal density function corresponding to each marginal j, and c is a copula density function. The copula function is unique for the continuous random vector $\mathbf{Y}$. However, the copula function is unique only over the Cartesian product of the ranges of the marginal distribution function in a discrete random vector (Genest and Neslehova, 2007). For modeling issues, parametric modeling of discrete variables by copula acquires dependence properties in the same way as in the continuous case.

For discrete variables, the probability mass function can be evaluated by taking the difference of the copula function. The joint probability mass function (pmf) of $\mathbf{Y}$ can be obtained as follows:

$$\Pr(\mathbf{Y}=y)=\sum_{i_1=0,1}\cdots\sum_{i_m=0,1}(-1)^{i_1+\ldots+i_m}C(F_1(y_1-i_1),\ldots,F_m(y_m-i_m)) \tag{3.5}$$

Note that, to compute this pmf, we have to evaluate 2^m times of the copula functions. However, one can approximate the pmf of $\mathbf{Y}$ by building up from the number of bivariate copulas. This approach is called pair copula constructions (PCC). Pair copula constructions (PCC) were initiated by Joe (1996) and developed in more detail by Bedford and Cook (2001, 2002), and Kurowicka and Cooke (2006).

For continuous $\mathbf{Y}$, a PCC can be derived by factorizing the joint density function into the conditional density function and the marginal density function, as follows:

$$f(y_1,y_2,\ldots,y_m)=f_{1|2,\ldots,m}(y_1|y_2,\ldots,y_m)\,f_{2|3,\ldots,m}(y_2|y_3,\ldots,y_m)\ldots f_m(y_m) \tag{3.6}$$

Aas et al. (2009) have shown that the conditional density function on the right hand side of equation (3.6) can be decomposed into the product of a bivariate copula density and a univariate conditional density by using Sklar's theorem. This can be done recursively to each of the terms on the right hand side of equation (3.6) until $f(y_1,y_2,\ldots,y_m)$ is decomposed into the product of m(m-1)/2 bivariate copulas (Panagiotelis, 2012).

For discrete margins, we can decompose a pmf by using the method proposed by Panagiotelis, 2012 as follows:

$$\begin{aligned}\Pr(Y_1=y_1,\ldots,Y_m=y_m)&=\Pr(Y_1=y_1|Y_2=y_2,\ldots,Y_m=y_m)\times\\&\qquad\Pr(Y_2=y_2|Y_3=y_3,\ldots,Y_m=y_m)\times\ldots\times\Pr(Y_m=y_m)\end{aligned}. \tag{3.7}$$

We can perform the same decomposition as in a continuous case for each term on the right hand side of equation (3.7) to get the product of a bivariate copula.

For example in the case of m = 3, three-dimensional discrete margin PCC can be obtained as follows:

$$\begin{aligned}&\Pr(Y_1=y_1,Y_2=y_2,Y_3=y_3)\\&\quad=\Pr(Y_1=y_1|Y_2=y_2,Y_3=y_3)\times\Pr(Y_2=y_2|Y_3=y_3)\times\Pr(Y_3=y_3)\end{aligned}, \tag{3.8}$$

where

$$\begin{aligned}&\Pr(Y_1=y_1|Y_2=y_2,Y_3=y_3)\\&\quad=\frac{\left\{\sum_{i_1=0,1}\sum_{i_2=0,1}(-1)^{i_1+i_2}C_{12|3}(F_{1|3}(y_1-i_1|y_3),F_{2|3}(y_2-i_2|y_3))\right\}}{\Pr(Y_2=y_2|Y_3=y_3)},\end{aligned} \tag{3.9}$$

and the arguments in the copula function are

$$F_{1|3}(y_1-i_1|y_3)=\frac{C_{13}(F_1(y_1-i_1),F_3(y_2))-C_{13}(F_1(y_1-i_1),F_3(y_3-1))}{\Pr(Y_3=y_3)},$$

and

$$F_{2|3}(y_2-i_2|y_3)=\frac{C_{23}(F_2(y_2-i_2),F_3(y_3))-C_{23}(F_2(y_2-i_2),F_3(y_3-1))}{\Pr(Y_3=y_3)}.$$

Since the dominator of equation (3.9) cancels with the second term on the right hand side of equation (3.8), the full expression for the pmf of the three-dimensional discrete margin PCC is

$$
\begin{aligned}
&\Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) \\
&= \left\{ \sum_{i_1=0,1} \sum_{i_2=0,1} (-1)^{i_1+i_2} C_{12|3}\left(\frac{C_{13}(F_1(y_1 - i_1), F_3(y_3)) - C_{13}(F_1(y_1 - i_1), F_3(y_3 - 1))}{F_3(y_3) - F_3(y_3 - 1)}, \right.\right. \\
&\left.\left. \frac{C_{23}(F_2(y_2 - i_2), F_3(y_3)) - C_{23}(F_2(y_2 - i_2), F_3(y_3 - 1))}{F_3(y_3) - F_3(y_3 - 1)} \right) \right\} \left[F_3(y_3) - F_3(y_3 - 1) \right]
\end{aligned}
$$

The above model can be used to analyzed the dependence between each health behavior considered in this paper and can also determine the factors affecting those behaviors as the same time.

4. Results and Discussion

We select Frank copula for all bivariate copula that building up to approximate the multivariate joint distribution of multivariate binary probit models.

4.1 Factors Affecting Alcohol Consumption, Tobacco Consumption, and Physical Activity Behaviors

Table 2 presents the results of estimation of the multivariate binary probit models for alcohol consumption, tobacco consumption, and physical activity choices. The first dependent variable to be discussed is alcohol consumption level. The explanatory variables included in the model that significant are age, income, high cholesterol, gender, non-communication diseases, occupation, and married status. The coefficient interpretations are: 1) young individuals, individuals with higher income, individuals with lower cholesterol of 200 mg/dl or a lower number of chronic diseases, and married are more likely to alcohol consumption; 2) males are more likely to consume alcohol than females; 3) individuals who work in the agricultural sector and work in risky occupations such as police and soldiers are more likely to consume alcohol than white-collar workers and those from the labor sector.

The second dependent variable is the level of tobacco consumption, the explanatory variables included in the model that significant are age, quality of health assessment, gender, non-communication diseases, occupation only agriculture and labor, married, and Body Mass Index. The coefficient interpretations are: 1) Young individuals, individuals who lower health quality assessment or lower number of chronic diseases, individuals who education lower than bachelor degree, and individuals who non-obese (BMI < 25) are more likely to tobacco consumption; 2) male are more likely to alcohol consumption than female; 3) individuals who work in agricultural sector are more likely to tobacco consumption than labor sector.

The third dependent variable is physical activity level in leisure time. The explanatory variables included in the model that are significant are age, income, health quality assessment, gender, non-communicable diseases, occupation only agriculture and police, married status, education, and Body Mass Index. The coefficient interpretations are: 1) young individuals, individuals with higher income, number of chronic diseases, individuals who are non-married, individuals who education in bachelor degree, and individuals who are obese (BMI > 25) are more likely to undertake physical activities; 2) males are more likely to undertake physical activities than females; 3) individuals who work in the agricultural sector are more likely to undertake physical activities than those from the risky occupations such as police and soldiers.

Table 2 presents the results of estimation of the multivariate binary probit models for alcohol consumption, tobacco consumption, and physical activity choices.

Variables	Y1		Y2		Y3	
	Coeff.	Std.err	Coeff.	Std.err	Coeff.	Std.err
age	**-0.0122**	**7.149e-04**	**-9.249e-03**	**1.536e-03**	**-8.608e-03**	**7.470e-04**
income	**1.394e-05**	**1.892e-06**	5.206e-06	3.962e-06	**-5.957e-06**	**1.981e-06**
pe_tc200	**-0.0507**	**0.0201**	0.0507	0.0433	-0.0138	0.0209
qlhealth	3.277e-03	0.0111	**-0.0963**	**0.0249**	**0.1781**	**0.0125**
sex	**0.9327**	**0.0202**	**1.427**	**0.08859**	**0.3800**	**0.0209**
NCD	**-0.1042**	**0.01189**	**-0.0854**	**0.0296**	**0.0698**	**0.0115**
agr	**0.3053**	**0.0324**	**0.2449**	**0.0709**	**-0.1988**	**0.0348**
white	**0.1866**	**0.0619**	7.133e-03	0.1318	-0.0621	0.0636
police	**0.3109**	**0.0886**	0.1578	0.1536	**-0.3393**	**0.0991**
labor	**0.1623**	**0.02900**	**0.1419**	**0.0663**	-0.0142	0.0299
married	**0.0580**	**0.0218**	-0.0802	0.0504	**-0.0891**	**0.0221**
bachalor	-0.0411	0.0452	**-0.5405**	**0.1287**	**0.2943**	**0.0442**
pe_bmi25	-0.0362	0.0214	**-0.1751**	**0.0511**	**0.0698**	**0.02197**
θ1(Y1-Y3)	**0.9121**	**0.1974**				
θ2(Y2-Y3)	**-0.5775**	**0.2093**				
θ3((Y1-Y2)	-0.0270	0.1349				

4.2 Dependence Measures of Health Behaviors Pairs

The two for three dependence parameter estimated from the Frank copula multivariate ordered probit are significant. Firstly, the dependence parameter estimated for alcohol consumption and physical activity behavior is 0.912, that can be interpreted as the positive correlation between alcohol consumption and physical activity behaviors. The secondly, the dependence parameter estimated for tobacco consumption and physical activity behaviors is -0.577, that can be interpreted as the negative correlation between alcohol consumption and physical activity behaviors.

5. Concluding Remarks

From the empirical results previously discussed, the followings are the recommended policies designed to reduce health-risk behavior and increase health inducing behavior for Thai citizens:

a) The policies should be focus on children and teenage more than adult stage to reduce alcohol and tobacco consumption, because of the empirical results show that the young individuals are more likely to alcohol consumption, tobacco consumption, and more likely to physical activity in leisure time.

b) Campaigns aimed at reducing alcohol consumption should have a greater focus on workers in the agricultural sector and in risky occupations.

c) The empirical results show that there is a negative correlation between tobacco consumption behavior and physical activity behavior. Thus, anti-smoking policies would have a more positive impact when the policy makers promote physical activity campaign.

d) Thai citizens should be know about their health status, because of the empirical results show that the individuals who are low health status, for example, obese, high cholesterol, more non-communication diseases, have less health-risk behaviors and more health behaviors.

For further study, the copula-based ordered probit model should be generalized to a multivariate model that increase the level of ordinal outcomes. However, the empirical results of this paper cannot confirm that there is some dependence between alcohol and tobacco consumption as discussed in the Alcohol Alert, 2007. That study found that people who smoke are much more likely to drink, and people who drink are much more likely to smoke.

Referencess

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009), “Pair-Copula Constructions of Multiple Dependence,” Insurance, Mathematics and Economics, 44, 182–198.

Bedford, T., Cooke, R.M., (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. Annals of Mathematics and Artificial Intelligence 32, 245–268.

Bedford, T., Cooke, R.M., (2002). Vines — a new graphical model for dependent random variables. Annals of Statistics 30 (4), 1031–1068.

Genest, C., and Neslehová, J. (2007), “A Primer on Copulas for Count Data,” The Astin Bulletin, 37, 475–515.

Joe, H., (1996). Families of m-variate distributions with given margins and m(m − 1)/2 bivariate dependence parameters, in: Rüschendorf, L., Schweizer, B., Taylor, M. D., (Eds.), Distributions with fixed marginals and related topics, pp. 120–141.

Joe, H., (1997). Multivariate Models and Dependence Concepts. Chapman & Hall, London.

Katewongsa P., Y. Sawangdee, C. Yousomboon, P. Choolert. (2014). Physical activity in Thailand: The general situation at national level. Journal of Science and Medicine in Sport. Vol 18. p 100-101.

Kurowicka, D., and Cooke, R. (2006), Uncertainty Analysis With High Dimensional Dependence Modelling (Wiley Series in Probability and Statistics), Chichester: Wiley.

Nelsen, R.B., (2006). An Introduction to Copulas, second ed. Springer, New York.

Panagiotelis, A., C., Czado and Harry Joe (2012): Pair Copula Constructions for Multivariate Discrete Data, Journal of the American Statistical Association, 107:499, 1063-1072

Praponsin, S., (2007). Factor Affecting Alcohol Consumption of Thai People. The 3rd National Alcohol Conference "Alcohol : No Ordinary Commodity", 21st-22nd November 2007, Richmond Hotel, Nonthaburi, Page 159.

Sirirassamee, T., B. Sirirassamee, A. Jampaklay, R. Borland, G. T., Fong, (2009). Risk Factors of Tobacco use Among Thai Adolescents: Finding from International Tobacco Control Policy Survey Southeast Asia (ICT-SEA), Journal Med Assoc Thai Vol. 92 Suppl. 3

Sirisrisakulchai, J., and S. Sriboonchitta (2014a). Modeling dependence of accident related outcomes using pair copula constructions for discrete data, "Modeling Dependence in Econometrics", Advances in Intelligent Systems and Computing Volume 251, pp 215 – 228.

Sirisrisakulchai, J., and S. Sriboonchitta (2014b). Factors Affecting Hospital Stay Involving Drunk Driving and Non-Drunk Driving in Phuket, Thailand, "Modeling Dependence in Econometrics", Advances in Intelligent Systems and Computing Volume 251, pp 479 – 489.

Suknark, K., J. Sirisrisakulchai, and S. Sriboonchitta (2016a). Modeling Dependence of Health Behaviors Using Copula-Based Bivariate Ordered Probit, "Causal Inference in Econometrics", Studies in Computational Intelligence, Volume 622, pp 295-306.

Suknark, K., J. Sirisrisakulchai, and S. Sriboonchitta (2016b). Reinvestigating the Effect of Alcohol Consumption on Hypertension Disease, "Causal Inference in Econometrics", Studies in Computational Intelligence, Volume 622, pp 307-318.

Suwannashote, P. (2009). Factors affecting alcohol consumption behavior: case study of the air technical students of the Royal Thai Air Force. Bangkok: Mahidol University.

The National Statistical Office of Thailand . (2011). Executive summary: The 2011 Survey on Population Behavior in Playing Sport or Physical Exercising and Mental Health.

APPENDIX B

Modeling Dependence of Health Behaviors Using Copula-Based Bivariate Ordered Probit

Kanchit Suknark, Jirakom Sirisrisakulchai and Songsak Sriboonchitta

Abstract This study simultaneously determines the factors affecting each pairing of health behaviors such as alcohol-consumption and physical activity, tobacco-consumption and physical activity, and alcohol-consumption and tobacco-consumption. The measure of dependence between these pairs was quantified using the copula approach. The Copula-based Ordered Probit Model was used to control any common unobserved factors that might affect the random errors related to each pair of health behaviors. The results is more efficient parameter estimates, in terms of lower standard errors, in comparison with separate estimations. Moreover, understanding the dependencies between ordinal choices for each pair of health behaviors gives useful information for designing more efficient health care programs.

Keywords Copula · Bivariate ordered probit · Alcohol consumption · Tobacco consumption · Physical activity

1 Introduction

Thailand is a medium-high income country where morbidity and mortality are primarily related to chronic rather than infectious diseases. Cardiovascular disease is the main cause of death with cancer as the next highest [11]. The risk factors for raising the mortality rate were health behaviors. For example, alcohol consumption, smoking, poor eating habits and diet, urban air pollution, obesity, physical inactivity, and unsafe sex [4]. Health behaviors are particularly important factors for health policy planning.

The explicit burden on society due to health-risk behaviors, particularly alcohol and tobacco consumption, includes health care costs, productivity loss, property damage costs, costs of criminal justice as well as law enforcement. To reduce health-risk behaviors, Thailand should aim to reduce alcohol consumption and prevent initiation of drinking. While Thailand already implements alcohol related policies,

---

K. Suknark (✉) · J. Sirisrisakulchai · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail: kanchitsook@hotmail.com

 295
V.-N. Huynh et al. (eds.), *Causal Inference in Econometrics*,
Studies in Computational Intelligence 622, DOI 10.1007/978-3-319-27284-9_18

such as high alcohol taxation, restricted alcohol sale times, more effective measures at the societal level to control alcohol consumption and alcohol-related harms are still required. The national survey in 2011 reported that about 17.7 million people or 20.8 % of the population aged 15 years and over are alcohol users. Men used alcohol at a higher rate than women [10].

Equally, tobacco consumption control policies have been implemented to reduce tobacco consumption and prevent initiation of smoking, especially in younger people. Current policies include high rates of tobacco taxation, control of tobacco advertising, non-smoking areas and bans on smoking in public places, workplaces, public transport, schools and other areas and facilities, supporting quit-smoking programs and publicity campaigns. These policies have been shown to be successful in decreasing the proportion of smokers in the Thai population (aged 15 years and older) from 32 % in 1991 to 20 % in 2013 [10].

Since 2010, the Thai Health Promotion Foundation has promoted physical activity in the Thai population by sponsoring and supporting several public campaigns nationally on the benefits of physical activity and advising people on the effective levels of frequency, duration and intensity required to achieve physical fitness. Such programs have also been supported at the local and regional level in many areas of the country. Most of the projects are mainly focused on increasing perceptions, attitudes, and practices related to physical activity generally [3]. The national survey in 2011 reported that about 26.1 % of the population played some form of sport or physical exercised, but this is actually a decrease of about 3 % when compared with the 2007 levels [10].

The previous studies on the factors affecting alcohol consumption, tobacco consumption, and physical activity were based on a single equations [3, 6, 7, 9]. In this paper, we simultaneously determined the factors affecting each pair of some important health behaviors including alcohol-consumption and physical activity pair, tobacco-consumption and physical activity pair, and alcohol-consumption and tobacco-consumption pair, and attempted to quantify the dependence measures between these pairs using the copula approach. A bivariate ordered probit model was used to control for the common unobserved factors that might affect the random errors in each pair of health behaviors. If these random errors are ignored, and not correlated, inefficiency in parameter estimation is likely [1]. Moreover, understanding the dependencies between the ordinal choices for each pair of health behaviors will give information useful for designing more efficient health care programs.

2 Data

The data used in this study are from the Thai National Health Examination Survey, No. 4 (NHES IV) from 2009. The data consists of a sample of 20,450 individuals. The ordered dependent variables are alcohol consumption (Y_1), tobacco smoking (Y_2), and physical activity in leisure time (Y_3). The independent variables are sex, age, income, chronic diseases, marital status, education level, and occupation. The

Table 1 Main statistics and description of variables

Variables	Description	N	Mean	SD	Min.	Max.
Y1	Level of alcohol consumption	20450	0.446	0.697	0	3
Y2	Level of Tobacco consumption	20450	0.052	0.339	0	3
Y3	Level of Physical Activity	20450	2.201	0.845	0	3
Sex	1 if individual is male; 0 otherwise	20450	0.524	0.499	0	1
Age	In year	20450	52.917	18.236	14	98
Income	In 1,000 Baht	20450	3.310	5.698	0	32.480
Bachelor	1 if individual graduated from Bachelor degree or higher; 0 otherwise	20450	0.061	0.24	0	1
Agr	1 if individual works in agricultural sector; 0 otherwise	20450	0.176	0.381	0	1
Whi	1 if individual is white-collar worker	20450	0.035	0.184	0	1
Police	1 if individual works as police or soldier; 0 otherwise	20450	0.012	0.108	0	1
Labor	1 if individual is in labor sector; 0 otherwise	20450	0.48	0.499	0	1
Married	Marital status where 1 indicates married; 0 otherwise	20450	0.636	0.481	0	1
pe_bmi25	1 if individual has body mass index more than 25; 0 otherwise	20450	0.348	0.476	0	1
pe_tc200	1 if individual has chloresterol level more than 200; 0 otherwise	20450	0.561	0.496	0	1
qlhealth	Self health quality assessment, where 5 is the highest level	20450	3.708	0.867	0	5
NCD	Number of chronic diseases	20450	0.632	0.959	0	10

alcohol consumption variable (Y_1) was stated as an amount of ethanol consumption on average per day in a year, and was classified into four levels: 0 for non-alcohol consumption; 1 for less than or equal to 40 g of ethanol on average per day (considered to be a responsible level of consumption); 2 for 41–60 g of ethanol on average per day (a harmful level); and 3 for over 61 g of ethanol on average per day (hazardous level). For the tobacco consumption variable (Y_2), measured as an amount of cigarettes per day, it can be classified into four levels: 0 for non-smoking; 1 for up to 10 cigarettes per day; 2 for more than 10 and up to 20 cigarettes per day; 3 for more than 20 cigarettes per day. For the physical activity variable (Y_3), the levels of physical activity or exercise in leisure time were: 0 for non-physical activity; 1 for low level of activity; 2 for moderate level of activity; and 3 for high level of activity. These are obviously indicative levels rather than attempting to quantify physical activity by number of hours or some other more precise measure. Table 1 presents the description of variables and main statistics.

3 Copula-Based Bivariate Ordered Probit Models

A Bivariate Ordered Probit Model is a system of two equations that can be used to model a simultaneous relationship of two ordinal outcome variables. The traditional Bivariate Ordered Probit Model uses the bivariate normal distribution to model the dependence between two equations [1]. In this study, we used a copula distribution function to model the dependence between two ordinal outcome responses. This is more flexible than the bivariate normal distribution. The Copula Function is a joint distribution with uniform margins. Let $U_1, \ldots, U_q$ be the possibly dependent uniform random variables on [0, 1]-interval. Copula can be defined as

$$C_\theta(u_1, \ldots, u_q) = \Pr(U_1 \le u_1, \ldots, U_q \le u_q), \tag{1}$$

where $C_\theta(\cdot, \ldots, \cdot)$ is a Copula Function with the dependent parameter θ, and u_m, for $m = 1, \ldots, q$ is a realization of U_m. The Copula Function must be grounded and increasing on the unit hypercube on its domain $[0, 1]^q$ (see, [5] for more details). By Sklar's Theorem (1959), for q marginal distribution functions, $F_1(\cdot), \ldots, F_q(\cdot)$ and $(z_1, \ldots, z_q)$ are arbitrary, we can derive the joint distribution $H(\cdot, \ldots, \cdot)$ for the random variables, $Z_1, \ldots, Z_q$ as follows:

$$C_\theta(F_1(z_1), \ldots, F_q(z_q)) = \Pr(F^{-1}(U_1) \le z_1, \ldots, F^{-1}(U_q) \le z_q) \equiv H(z_1, \ldots, z_q), \tag{2}$$

where $Z_m = F_j^{-1}(U_m), m = 1, \ldots, q$. Thus, we can construct a joint distribution function from a set of margins by using the Copula Function to combine them.

Now we can start deriving our copula-based bivariate ordered probit model. Suppose that each individual i selects the level of two dependent ordinal responses based on the following system of two equations:

$$Y_{i1}^* = X_{i1}\beta_1 + \varepsilon_{i1}, \tag{3}$$
$$Y_{i2}^* = X_{i2}\beta_2 + \varepsilon_{i2}, \tag{4}$$

where i indexes individual $i = 1, \ldots, N$, Y_{i1}^* and Y_{i2}^* are latent variables, X_{i1} and X_{i2} are the $K \times N$ matrices of explanatory variables, β_1 and β_2 are conformable vectors of parameters to be estimated, and ε_{i1} and ε_{i1} are random errors.

We can model the observed level of two dependent ordinal responses, Y_{i1}, and Y_{i2} by the following threshold crossing conditions:

$$Y_{ij} = r_j \quad \text{if} \quad \tau_{r_j,j} \le Y_{ij}^* < \tau_{r_j+1,j}, \quad r_j = 1, \ldots, R_j, \quad j = 1, 2 \tag{5}$$

where R_j are the number of ordinal levels of $Y_i j$ and $\tau_{r_j,j}$ are threshold parameters to be estimated from the model, with $\tau_{1,j} = -\infty$ and $\tau_{R_j,j} = +\infty$. The joint distribution of the individual selected the level of two ordinal response outcomes can be expressed as follows:

$$\begin{aligned}
&\Pr(\tau_{r_1,1} \le Y_{i1}^* < \tau_{r_1+1,1}, \tau_{r_2,2} \le Y_{i2}^* < \tau_{r_2+1,2}) \\
&= \Pr(\tau_{r_1,1} \le X_{i1}\beta_1 + \varepsilon_{i1} < \tau_{r_1+1,1}, \tau_{r_2,2} \le X_{i2}\beta_2 + \varepsilon_{i2} < \tau_{r_2+1,2}) \\
&= \Pr(\tau_{r_1,1} - X_{i1}\beta_1 \le \varepsilon_{i1} < \tau_{r_1+1,1} - X_{i1}\beta_1, \tau_{r_2,2} - X_{i2}\beta_2 \le \varepsilon_{i2} < \tau_{r_2+1,2} - X_{i2}\beta_2) \\
&= C_\theta(F_1(\tau_{r_1+1,1} - X_{i1}\beta_1), F_2(\tau_{r_2+1,2} - X_{i2}\beta_2)) \\
&\quad - C_\theta(F_1(\tau_{r_1,1} - X_{i1}\beta_1), F_2(\tau_{r_2+1,2} - X_{i2}\beta_2)) \\
&\quad - C_\theta(F_1(\tau_{r_1+1,1} - X_{i1}\beta_1), F_2(\tau_{r_2,2} - X_{i2}\beta_2)) \\
&\quad + C_\theta(F_1(\tau_{r_1,1} - X_{i1}\beta_1), F_2(\tau_{r_2,2} - X_{i2}\beta_2)).
\end{aligned}$$

For the traditional Bivariate Ordered Probit Model, the marginal distribution $F_1(\cdot)$ and $F_2(\cdot)$ are specified as the standard normal distribution and the copula function is specified as a Gaussian copula. Therefore, the traditional Bivariate Ordered Probit Model is the special case of copula-based Bivariate Ordered Probit Model. To capture a wider range of dependencies and distributional shapes of random errors, we use different type of copula functions and a mixture of two normal components for random errors.

The most general form of normal mixtures can be expressed as

$$F_j(z) = \pi_j \Phi\left(\frac{z - \mu_{j1}}{\sigma_{j1}}\right) + (1 - \pi_j)\Phi\left(\frac{z - \mu_{j2}}{\sigma_{j2}}\right) \tag{6}$$

where Φ is the standard normal distribution, π_j is the mixing parameter, μ_{j1} and μ_{j2} are location parameters, and σ_{j1} and σ_{j2} are dispersion parameters. The location and dispersion parameters have to be constrained to satisfy the mean and variance normalizations as follows:

$$\pi_j\mu_{j1} + (1 - \pi_j)\mu_{j2} = 0, \quad \pi_j(\sigma_{j1}^2 + \mu_{j1}^2) + (1 - \pi_j)(\sigma_{j2}^2 + \mu_{j2}^2) = 1. \tag{7}$$

This normal mixtures distribution can capture the varieties of skewness or bimodality in the shape of random errors.

The log-likelihood of the copula-based Bivariate Ordered Probit Model is given by

$$
\begin{aligned}
LL = \sum \log(&C_\theta(F_1(\tau_{r_1+1,1} - X_{i1}\beta_1), F_2(\tau_{r_2+1,2} - X_{i2}\beta_2)) \\
&- C_\theta(F_1(\tau_{r_1,1} - X_{i1}\beta_1), F_2(\tau_{r_2+1,2} - X_{i2}\beta_2)) \\
&- C_\theta(F_1(\tau_{r_1+1,1} - X_{i1}\beta_1), F_2(\tau_{r_2,2} - X_{i2}\beta_2)) \\
&+ C_\theta(F_1(\tau_{r_1,1} - X_{i1}\beta_1), F_2(\tau_{r_2,2} - X_{i2}\beta_2))).
\end{aligned}
$$

The corresponding vector of parameters $\beta_1, \beta_2, \tau_{r_1,1}, \tau_{r_2,2}$, parameters of random errors $\pi_1, \mu_{11}, \mu_{12}, \sigma_{11}, \sigma_{12}, \pi_2, \mu_{21}, \mu_{22}, \sigma_{21}, \sigma_{22}$, and dependence parameter θ can be estimated simultaneously using the maximum likelihood estimation. This study uses STATA software [8] and user written command BICOP [2] to estimate all parameters in the models.

4 Results and Discussion

We consider both Frank copula and Gaussian copula that allow for both positive and negative dependence. For the marginal distribution of each residual, we consider three specifications including specifying each marginal as a standard normal distribution, and specifying one of the random errors as a normal-mixture distribution and another as a standard normal distribution, and specifying each marginal as a normal-mixture distribution. For all three pairs, the best fitted model (in terms of Akaike Information Criteria, AIC) is the Frank copula with standard normal distribution for both random errors. In comparison with the two separate univariate ordered probit model (independent copula), we found that the estimated standard errors of bivariate models are lower than those of univariate models (the results are not shown here). However, the differences are very small (five digits after the decimal point) corresponding with the low level of correlation between each random error.

4.1 Factors Affecting Alcohol Consumption and Physical Activity Behaviors

Table 2 presents the model estimation results for the level of alcohol consumption and physical activity behaviors pair. The first dependent variable to be discussed is alcohol consumption level. The explanatory variables included in the model that significant are age, income, high cholesterol, gender, non-communication diseases, occupation, education, and Body Mass Index. The coefficient interpretations are: (1) young individuals, individuals with higher income, individuals with lower cholesterol

Table 2 Parameter estimates for level of alcohol consumption and level of physical activity model

Variables	Y1		Y3	
	Coeff.	Std. err	Coeff.	Std. err
Sex	0.956	0.019	0.143	0.017
Age	−0.012	0.001	−0.005	0.001
Income	1.48E−05	1.71E−06	−2.40E−06	1.60E−06
Bachelor	−0.08	0.042	0.035	0.038
Agr	0.249	0.031	0.523	0.027
Whi	0.179	0.057	0.125	0.052
Police	0.251	0.079	0.194	0.077
Labor	0.168	0.027	0.339	0.023
Married	0.025	0.02	0.097	0.018
pe_bmi25	−0.034	0.02	0.059	0.018
pe_tc200	−0.062	0.019	−0.053	0.017
qlhealth	−0.011	0.011	0.09	0.009
NCD	−0.12	0.011	−0.041	0.009
$\tau_{1,j}$	−1.626	0.072	−2.244	0.068
$\tau_{2,j}$	−0.135	0.071	−0.576	0.064
$\tau_{3,j}$	0.107	0.071	0.205	0.064
θ	0.623	0.060		
LL	−35,863.919			

of 200 mg/dl or a lower number of chronic diseases, individuals who have education lower than bachelor degree, and individuals who are non-obese (BMI $<$ 25) are more likely to alcohol consumption; (2) males are more likely to consume alcohol than females; (3) individuals who work in the agricultural sector and work in risky occupations such as police and soldiers are more likely to consume alcohol than white-collar workers and those from the labor sector.

The second dependent variable is physical activity level. The explanatory variables included in the model that are significant are age, high cholesterol, health quality assessment, gender, non-communicable diseases, occupation, married status, and Body Mass Index. The coefficient interpretations are: (1) young individuals, individuals with higher health quality assessment, individuals with lower cholesterol of 200 mg/dl or lower, number of chronic diseases, individuals who are married, and individuals who are non-obese (BMI $<$ 25) are more likely to undertake physical activities; (2) males are more likely to undertake physical activities than females; (3) individuals who work in the agricultural sector are more likely to undertake physical activities than those from the other sectors.

4.2 Factors Affecting Tobacco Consumption and Physical Activity Behaviors

Table 3 presents the model estimation results of tobacco consumption and physical activity behaviors. For the first dependent variable, namely, the level of tobacco consumption, the explanatory variables included in the model that significant are age, quality of health assessment, gender, non-communication diseases, occupation only agriculture and labor, married, and Body Mass Index. The coefficient interpretations are: (1) Young individuals, individuals who lower health quality assessment or lower number of chronic diseases, individuals who education lower than bachelor degree, and individuals who non-obese (BMI < 25) are more likely to tobacco consumption; (2) male are more likely to alcohol consumption than female; (3) individuals who work in agricultural sector are more likely to tobacco consumption than labor sector.

For the second dependent variable, which is physical activity level, the explanatory variables included in the model that significant are age, high cholesterol, health quality assessment, gender, non-communication diseases, occupation, married status, and Body Mass Index. The coefficient interpretations are: (1) Young individuals, individuals who higher health quality assessment, individuals who lower cholesterol 200 mg/dl or lower number of chronic diseases, individuals who married, and

Table 3 Parameter estimates for level of tobacco consumption and level of physical activity model

Variables	Y2		Y3	
	Coeff.	Std. err	Coeff.	Std. err
Sex	1.426	0.088	0.143	0.017
Age	−0.009	0.002	−0.005	0.001
Income	5.70E−06	3.85E−06	−2.40E−06	1.64E−06
Bachelor	−0.533	0.131	0.035	0.038
Aagr	0.259	0.072	0.523	0.027
Whi	0.048	0.159	0.124	0.052
Police	0.135	0.156	0.194	0.077
Labor	0.161	0.069	0.339	0.023
Married	−0.06	0.049	0.097	0.018
pe_bmi25	−0.157	0.05	0.06	0.018
pe_tc200	0.06	0.043	−0.053	0.017
qlhealth	−0.098	0.024	0.09	0.009
NCD	−0.08	0.029	−0.041	0.009
$\tau_{1,j}$	−0.569	0.174	−2.244	0.068
$\tau_{2,j}$	−0.413	0.174	−0.576	0.064
$\tau_{3,j}$	0.051	0.176	0.205	0.064
θ	−0.528	0.170		
LL	−23,904.808			

individuals who non-obese (BMI < 25) are more likely to physical activities; (2) male are more likely to physical activities than female; (3) individuals who work in agricultural sector and labor are more likely to physical activities than the other sector.

4.3 Factors Affecting Alcohol Consumption and Tobacco Consumption Behaviors

Table 4 presents the model estimation results of alcohol consumption and tobacco consumption behaviors. The estimated parameters are similar to the previous subsections. More information from Table 4 is just the dependence parameter estimation, which will be discussed in the next subsection. The marginal effects of each dependent variable are shown in Tables 5, 6 and 7.

Table 4 Parameter estimates for level of alcohol consumption and level of tobacco consumption model

Variables	Y1		Y2	
	Coeff.	Std. err	Coeff.	Std. err
Sex	0.956	0.019	1.429	0.088
Age	−0.012	0.001	−0.009	0.002
Income	1.48E−05	1.71E−06	5.00E−06	3.87E−06
Bachelor	−0.081	0.042	−0.529	0.131
Aagr	0.249	0.031	0.263	0.072
Whi	0.182	0.057	0.046	0.158
Police	0.253	0.079	0.154	0.155
Labor	0.169	0.028	0.156	0.069
Married	0.025	0.02	−0.056	0.05
pe_bmi25	−0.034	0.02	−0.15	0.05
pe_tc200	−0.062	0.019	0.066	0.043
qlhealth	−0.01	0.011	−0.1	0.024
NCD	−0.12	0.011	−0.081	0.029
$\tau_{1,j}$	−1.622	0.072	−0.581	0.175
$\tau_{2,j}$	−0.131	0.071	−0.425	0.175
$\tau_{3,j}$	0.111	0.071	0.039	0.176
θ	0.979	0.178		
LL	−17,171.767			

Table 5 Marginal effects for level of alcohol consumption

Variables	Level of alcohol consumption			
	Level 0	Level 1	Level 2	Level 3
Age	0.0037	−0.0026	−0.0003	−0.0008
Income	−4.65e−06	3.28e−06.	4.13e−07	0.000000957
pe_tc200	0.0195	−0.0138	−0.0017	−0.004
qlhealth	0.0033	−0.0023	−0.0003	−0.0007
Sex	−0.3007	0.2119	0.0267	0.0621
NCD	0.0377	−0.0266	−0.0033	−0.0078
Agr	−0.0786	0.0554	0.007	0.0162
Whi	−0.0573	0.0404	0.0051	0.0118
Police	−0.0792	0.0558	0.007	0.0164
Labor	−0.0533	0.0375	0.0047	0.0111
Married	−0.0077	0.0054	0.0007	0.0016
Bachelor	0.0255	−0.018	−0.0022	−0.0053
Pe_bmi25	0.0109	−0.0077	−0.001	−0.0022

Table 6 Marginal effects for level of tobacco consumption

Variables	Level of tobacco consumption			
	Level 0	Level 1	Level 2	Level 3
Age	0.0005	−0.0001	−0.0002	−0.0002
Income	−3.14E−07	7.19E−08	1.44E−07	9.81E−08
pe_tc200	−0.0034	0.0008	0.0015	0.0011
qlhealth	0.0052	−0.0012	−0.0024	−0.0016
Sex	−0.0753	0.0172	0.0344	0.0237
NCD	0.0042	−0.0009	−0.0019	−0.0014
Agr	−0.0136	0.0031	0.0062	0.0043
Whi	−0.0017	0.0004	0.0008	0.0005
Police	−0.0068	0.0016	0.0031	0.0021
Labor	−0.0082	0.0019	0.0037	0.0026
Married	0.0031	−0.0007	−0.0014	−0.001
Bachelor	0.0281	−0.0064	−0.0128	−0.0089
Pe_bmi25	0.0081	−0.0019	−0.0037	−0.0025

4.4 Dependence Measures of Health Behaviors Pairs

The dependence parameters for three different pairs are significant, indicating the need to model these behaviors simultaneously. The dependence parameter estimated from the Frank copula bivariate ordered probit for alcohol consumption and physical activity behaviors is 0.623. This dependence parameter can be interpreted as

Table 7 Marginal effects for level of physical activity

Variables	Level of physical activity			
	Level 0	Level 1	Level 2	Level 3
Age	0.0002	0.0014	0.0004	−0.002
Income	7.00E−08	6.53E−07	1.85E−07	−9.08e−07
pe_tc200	0.0016	0.0147	0.0041	−0.0204
qlhealth	−0.0026	−0.0246	−0.0069	0.0341
Sex	−0.0042	−0.0392	−0.0111	0.0545
NCD	0.0012	0.0113	0.0032	−0.0157
Agr	−0.0154	−0.1433	−0.0405	0.1992
Whi	−0.0036	−0.0339	−0.0096	0.0471
Police	−0.0057	−0.0534	−0.0151	0.0742
Labor	−0.0099	−0.0929	−0.0262	0.129
Married	−0.0028	−0.0265	−0.0075	0.0368
Bachelor	−0.001	−0.0097	−0.0027	0.0134
Pe_bmi25	−0.0017	−0.0162	−0.0046	0.0225

a concordance measure (Kendall's Tau) equal to 0.07. The dependence parameter estimated from the Frank copula bivariate ordered probit for tobacco consumption and physical activity behaviors is −0.528. This dependence parameter can be interpreted as a concordance measure (Kendall's Tau) equal to −0.06. For the parameter estimated from the Frank copula bivariate ordered probit for alcohol consumption and tobacco consumption behaviors, the dependence parameter is 0.979, corresponding with 0.108 as the concordance measure. The concordance measure for all three models are quite small but statistically significant. Thus, we can not ignore these dependencies in model estimation.

5 Concluding Remarks

From the empirical results previously discussed, the followings are the recommended policies designed to reduce health-risk behavior and increase health inducing behavior for Thai citizens:

(a) Campaigns aimed at reducing alcohol consumption should have a greater focus on workers in the agricultural sector and in risky occupations.

(b) The empirical results show that there is a negative correlation between tobacco consumption behavior and physical activity behavior. Thus, anti-smoking policies would have a more positive impact when the policy makers promote physical activity campaign.

(c) Finally, the empirical results confirm that there is some dependence between alcohol and tobacco consumption as discussed in the Alcohol Alert, 2007. This study found that people who smoke are much more likely to drink, and people who drink are much more likely to smoke. Thus, the alcohol consumption reduction policies and anti-smoking policies would have more positive impact when they are more closely associated.

For further study, the copula-based ordered probit model should be generalized to a multivariate model. However, the main concern on this issue is the curse of dimensionality. When the level of ordinal outcomes and the number of outcomes itself increase, it will give more computational burden on model estimation. Practitioners have to consider about the trade-off between computational cost and efficiency gain.

References

1. Greene, W.H., Hensher, D.A.: Modeling Ordered Choices: A Primer. Cambridge University Press, Cambridge (2010)
2. Hernandez-Alava, M., Pudney, S.: BICOP: a Stata command for fitting bivariate ordinal regressions with residual dependence characterised by a copula function and normal mixture marginals, Working Paper Series No. 2015-02 (2015). https://www.iser.essex.ac.uk/research/publications/working-papers/understanding-society/2015-02
3. Katewongsa, P., Sawangdee, Y., Yousomboon, C., Choolert, P.: Physical activity in Thailand: the general situation at national level. J. Sci. Med. Sport **18**, 100–101 (2014)
4. Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., Murray, C.J.L.: Global Burden of Disease and Risk Factors. A copublication of Oxford University Press and The World Bank (2006)
5. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer, New York (2006)
6. Praponsin, S.: Factor affecting alcohol consumption of Thai people. In: The 3rd National Alcohol Conference "Alcohol : No Ordinary Commodity", Richmond Hotel. Nonthaburi, pp. 159, 21–22 November 2007
7. Sirirassamee, T., Sirirassamee, B., Jampaklay, A., Borland, R., Fong, G.T.: Risk factors of tobacco use among thai adolescents: finding from international tobacco control policy survey Southeast Asia (ICT-SEA). J. Med. Assoc. Thai **92**, Suppl 3 (2009)
8. StataCorp.: Stata statistical software: Release 13. StataCorp LP, College Station, TX (2013)
9. Suwannashote, P.: Factors affecting alcohol consumption behavior: case study of the air technical students of the Royal Thai Air Force. Mahidol University, Bangkok (2009)
10. The National Statistical Office of Thailand (2011). Executive summary: The 2011 survey on population behavior in playing sport or physical exercising and mental health
11. World Health Organization: Noncommunicable Diseases (NCD) Country profiles. WHO Documentation Production Service, Geneva (2014)

APPENDIX C

Reinvestigating the Effect of Alcohol Consumption on Hypertension Disease

Kanchit Suknark, Jirakom Sirisrisakulchai and Songsak Sriboonchitta

Abstract The researchers reinvestigate the effect of alcohol consumption on hypertension from observational data, taken from the Thai National Health Examination Survey. In the observed samples, the treatment assignment is not ignorable, thus using treatment as a dummy variable in the statistical model will lead to the bias estimation of treatment effects. Factors affecting self-selection (drink/not drink) may cause the dummy variable of treatment to be correlated with random errors in the outcome model, which leads to the biased parameters estimation. We propose to use copula-based endogenous switching regression for ordinal outcomes as the more appropriate model for treatment effect estimation. The new results should give us more a accurate and reliable treatment effect for causal inference.

Keywords Alcohol consumption · Hypertension disease · Copula · Endogenous switching regression

1 Introduction

Hypertension is a major risk factor for cardiovascular disease. Tobacco consumption and alcohol consumption are the most important avoidable causes of cardiovascular diseases worldwide [14]. In Thailand, cardiovascular disease and hypertension are major health problems. The cardiovascular morbidity rate and Hypertension have been reported as an important cause of morbidity for past several years (1) The cardiovascular disease was ranked fourth largest cause of death in 2012 (32.9 people per 100,000) (2) Hypertension was the third largest cause of death (37.4 people per 100,000). Both causes are increasing in frequency [22]. Table 1 presents the causes of death in Thailand.

Epidemiological and experimental investigations have established a close association between alcohol consumption and hypertension [6]. A number of population studies have almost unanimously shown an empirical link between high levels of

K. Suknark (✉) · J. Sirisrisakulchai · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail: kanchitsook@hotmail.com

 307

V.-N. Huynh et al. (eds.), *Causal Inference in Econometrics*,
Studies in Computational Intelligence 622, DOI 10.1007/978-3-319-27284-9_19

Table 1 Causes of death in Thailand (per 100,000 inhabitants)

Cause groups	2008	2009	2010	2011	2012
1. Malignant neoplasm, all forms	87.6	86.3	91.2	95.2	98.5
2. Accident and poisonings	55.1	55.6	51.6	52.8	51.6
3. Hypertension and cerebrovascular disease	24.7	24.7	31.4	–	37.4
4. Disease of the heart	29.8	29.0	28.9	31.4	32.9
5. Pneumonia and other diseases of lung	23.0	22.9	25.7	26.3	26.1

Source Bureau of Policy and Strategy, Office of the Permanent Secretary, Ministry of Public Health

alcohol consumption and hypertension. In many former studies, risk factors were assessed using multiple regression models; i.e. age, gender, family history, weight and height for hypertension [2, 9, 12, 15, 21, 24, 25].

In 1985, [15] studied the direct effects of alcohol consumption. He received responses from 46 male drinkers (22–55 years) to a questionaire. Similar research has been conducted by Yadav et al. [25] and Taraman et al. [21]. They studied both tobacco and alcohol consumption among people in Southern India [21]. The results suggest that hypertension becomes more frequent among women of advancing age who are alcohol users, especially if they smoke [25]. Yadav studied the prevalence of Hypertension in northern India among subjects of approximately 30 years of age. The results indicate that increasing age, body mass index, obesity and impaired glucose tolerance were significantly associated with Hypertension [25]. They all used regression analysis in their research. Oyunbileg used regressive methods to assess the correlation between alcohol consumption and arterial hypertension among the population of the Gobi region of China [12].

In Thailand, multiple regression modeling research shows an empirical link between high levels of alcohol consumption and hypertension. Leelarassme et al. studied a sample in Bangkok [7]. Pati and Siviroj studied a sample in Northern Thailand [13, 19]. Howteerakul studied people living in the rural areas in Thailand [5]. In these studies the researchers used data collected in a national survey [13, 19].

In this paper, the researchers reinvestigate the effect of alcohol consumption on the levels of hypertension. Moreover, the researchers investigate whether alcohol consumption may be a cause of hypertension by introducing the switching regression model for observational data analysis. This investigation of causality has been done using the Neyman-Rubin counterfactual framework [17]. In this framework, groups were compared to extract the differences that are due to changes in the level of alcohol consumption and the affect on levels of hypertension.

Ideally, we just want to compare two groups that only differ with regard to whether they consumed alcohol or not. Two groups should be identical in all relevant characteristics, so that we can deduce changes due to the effect of alcohol consumption on hypertension level. This idea can be done in practice by randomly assigning individuals to each group by using the same probability of being assigned to either the control or treatment groups. The characteristics of the individuals in each group may be viewed as interchangeable and are the same, at least on average, if we have a large enough sample.

The potential outcomes framework [10] uses a what-if scenario as the baseline for making causal inferences. This baseline allows each individual the same chance for potential outcomes. However, we can only observe the individual in one treatment condition. Under the interchangeable assumption and given that each group only differs with regard to the treatment condition, the causal effect can then be statistically estimated as the difference in the means of an outcome of interest between the control and treatment groups [1].

To consistently estimate the average treatment effect, we have to satisfy the ignorable treatment assignment assumption [16]. This condition ensures that the outcome of interest is independent of the treatment assignment mechanism [10]. This assumption can be satisfied when we assign each individual randomly to the treatment and control groups. In other words, we can only consistently estimate the average treatment effect if there is no selection bias in the treatment assignment.

In our study, the survey participants self-selected themselves into treatment condition (whether to drink alcohol or not), thus it is hard to believe that the independence assumption holds in this situation. In the observed data, the treatment assignment is not ignorable, thus using treatment as a dummy variable in the statistical model will lead to the bias estimation of the treatment effect. Factors affecting the self-selection might cause the dummy variable of treatment condition to be correlated with the random errors in the outcome model, which leads to the biased parameters estimation. By introducing the endogenous switching regression model for ordered outcomes, the new treatment effect estimation should give us more accurate and reliable results for causal inference.

2 Data

This study uses the Thai National Health Examination Survey, No. 4 (NHES IV) data of 2009. The ordered dependent variables are the blood pressure levels of participants. Blood pressure levels may be classified into three groups of individuals as follows (Thailand context). First, those with an average diastolic blood pressure (DBP) and systolic blood pressure (SBP) of <80/120 mmHg as normal blood pressure level; second, those with DBP and SBP of 80–90/120–140 mmHg as pre-hypertension level; and third those with DBP and SBP of greater than 90/140 mmHg as regarded as having hypertension. The independent variables are gender, age, income, chronic diseases (if any), marital status, level of education, and occupation. Table 2 presents a description of the variables and related statistics.

Table 2 Description of variables and statistics

Variables	Description	N	Mean	SD	Min.	Max.
BP	Level of blood pressure	20,450	0.446	0.697	0	2
Gender	1 if individual is male; 0 otherwise	20,450	0.524	0.499	0	1
Age	In Year	20,450	52.917	18.236	14	98
Income	In Baht	20,450	3.31	5.698	0	32.48
Bachelor	1if individual graduated in Bachelor degree or higher; 0 otherwise	20,450	0.061	0.24	0	1
Agr	1 if individual works in agricultural sector; 0 otherwise	20,450	0.176	0.381	0	1
Whi	1 if individual is white collar	20,450	0.035	0.184	0	1
Police	1 if individual works as police or soldier; 0 otherwise	20,450	0.012	0.108	0	1
Labor	1 if individual is classified as labor; 0 otherwise	20,450	0.48	0.499	0	1
Married	Marital status where 1 indicates married; 0 otherwise	20,450	0.636	0.481	0	1
pe_bmi25	1 if individual has body mass index more than 25; 0 otherwise	20,450	0.348	0.476	0	1
pe_tc200	1 if individual has cholesterol level more than 200; 0 otherwise	20,450	0.561	0.496	0	1
qlhealth	Health quality assesment where 5 is the highest level	20,450	3.708	0.867	0	5
NCD	Number of chronic diseases	20,450	0.632	0.959	0	10

3 Switching Regression Model for Level of Hypertension

In the sample used to estimate the effect of alcohol consumption on the level of hypertension, the participants are not randomly drawn from the population from which we wanted to draw inferences, but participants who self-selected themselves into treatment. The approach to self-selection used, is that proposed by Heckman [4]. The assumption is that the self-select mechanism may be modeled by a binary choices model. The switching regression model, was supplemented with copula by using copula to model the correlation between the random errors from a decision model and outcome models [4].

Consider two decisions, $S = 0, 1$, where 1 is 'drink alcohol' and 0 is not. Let $S^* = Z\gamma + \nu$ be the latent variable for the decision mechanism. The decision rule is the following condition

$$S = \begin{cases} 1 & \text{if } s^* > 0 \\ 0 & \text{if } s^* \leq 0 \end{cases}$$

where Z is the matrix of the explanatory variables explaining the self-select mechanism, and γ is the corresponding vector of parameters to be estimated. The individuals are observed either in decision $S = 0$, or in decision $S = 1$, but never in both.

Consider the outcome of interest, the level of hypertension $Y_s = 0, \ldots, J$, can be modeled using the latent variable framework and can be determined by the following condition:

$$Y_s = j \quad \text{iff} \quad \kappa_{s,j-1} < Y_s^* \leq \kappa_{s,j}, \quad s = 0, 1, \quad j = 0, \ldots, J \tag{1}$$

where $\kappa_{s,j}$ are the threshold values, which form a partition of the real line, i.e., $\kappa_{s,0} = -\infty$, $\kappa_{s,J} = \infty$, and $\kappa_{s,j} > \kappa_{s,j-1}$ for all j.

Let $Y_0^* = X\beta_0 + \varepsilon_0$ be the latent variable for the individual decision not to drink $S = 0$, and $Y_1^* = X\beta_1 + \varepsilon_1$ be the latent variable for the individual to consume alcohol $S = 1$, where, X is the vector of all the explanatory variables, β_0 and β_1 are the vector of the parameters to be estimated.

As previously discussed, there might be some unobserved factors affecting both the self-selected mechanism and the response outcome, therefore the probability of observing $Y_s = j$ depends on the self-selected variable S. Given that S and Y_s are not necessarily independent. We have

$$\begin{aligned} Pr(Y_0 = j, S = 0|X, Z) &= Pr(\kappa_{0,j-1} - X\beta_0 \leq \varepsilon_0 < \kappa_{0,j} - X\beta_0, \nu \leq -Z\gamma) \\ &= Pr(\varepsilon_0 < \kappa_{0,j} - X\beta_0, \nu \leq -Z\gamma) \\ &\quad - Pr(\varepsilon_0 \leq \kappa_{0,j-1} - X\beta_0, \nu \leq -Z\gamma) \end{aligned}$$

$$\begin{aligned} Pr(Y_1 = j, S = 1|X, Z) &= Pr(\kappa_{1,j-1} - X\beta_1 \leq \varepsilon_1 < \kappa_{1,j} - X\beta_1, \nu \leq -Z\gamma) \\ &= Pr(\varepsilon_1 < \kappa_{1,j} - X\beta_1) - Pr(\varepsilon_1 < \kappa_{1,j-1} - X\beta_1) \\ &\quad - Pr(\varepsilon_1 \leq \kappa_{1,j} - X\beta_1, \nu \leq -Z\gamma) \\ &\quad + Pr(\varepsilon_1 \leq \kappa_{1,j-1} - X\beta_1, \nu \leq -Z\gamma) \end{aligned}$$

To model the above probability, we have to specify the appropriate joint distribution functions. In this paper, we suggest combining the marginal distributions (ε_s and ν) by using copula.

Copula was introduced in 1959 by Abe Sklar. Sklar's Theorem succinctly states that there is a copula function that connects multivariate distributions to their univariate marginal distributions being uniformly distributed on the unit interval [0,1]. Therefore, copula is the joint distribution of a multivariate uniform random vector. Introduction and standard reference on copula theory can be found in [11].

For a bivariate joint distribution H with marginal distributions F_1 and F_2, the copula $C : [0, 1]^2 \rightarrow [0, 1]$, which combines these two marginal distributions, can be expressed as follows:

$$H(x_1, x_2) = C\left\{F_1(x_1), F_2(x_2)\right\}, (x_1, x_2) \in \mathbf{R^2} \tag{2}$$

The copula function is uniquely determined for the continuous random vector (F_1, F_2). For a discrete random vector, the copula function is unique only over the Cartesian product of the range of the marginal distribution function [3]. Thus, in discrete cases the mapping from two marginal distributions and copula to a bivariate joint distribution is not one-to-one. However, the region outside the Cartesian product of the range of the marginal distribution function is not of interest [11]. Moreover, [3] demonstrated that parametric modeling of discrete random vector by copula acquires dependence properties in a way that is similar to the continuous case.

For any copula, the marginal distribution implied by bivariate copula are $C(u, v) \leq C(u, 1) = u$ and $C(u, v) \leq C(1, v) = v$, for all $0 \leq u, v < 1$, and so $W(u, v) = max(u + v - 1, 0) \leq C(u, v) \leq min(u, v) = M(u, v)$. The copula $M(u, v)$ and $W(u, v)$ are called the Frechet upper bound and Frechet lower bound, respectively. We can interpret the Frechet lower bound as the copula with the maximum negative dependence and Frechet upper bound as the copula with the maximum positive dependence. In modeling switching regression, it is essential that the copula should allow for both positive and negative dependence, since the direction of the selection bias can be in both directions. We should not restrict the direction of selection bias a priori. The selection pattern should be explained by the data itself.

Copula has had limited use in the endogenous switching regression models. Some, but not all examples, are [18, 23] for modeling endogenous switching regression in count outcomes, [20] for modeling endogenous switching regression of continuous variables and [8] for modeling endogenous switching regression in ordered outcomes.

In this paper, we consider six copula functions, namely, the Normal copula, the FGM (Farlie-Gumbel-Morgenstern) copula, AMH (Ali-Mikhail-Haq) copula, the t copula, and the Frank copula. However, only the Normal copula, t copula, and Frank copula can reach the Frechet lower bound and upper bound, and thus can span the full range of dependence.

For any given copula, the two required joint distribution, $Pr(Y_0 = j, S = 0|X, Z)$ and $Pr(Y_1 = j, S = 1|X, Z)$ are fully determined. Therefore,

$$
\begin{aligned}
Pr(Y_0 = j, S = 0|X, Z) = {} & C_0(F_1(\kappa_{0,j-1} - X\beta_0), F_2(-Z\gamma); \theta_0) \\
& - C_0(F_1(\kappa_{0,j-1} - X\beta_0), F_2(-Z\gamma); \theta_0)
\end{aligned}
$$

$$
\begin{aligned}
Pr(Y_1 = j, S = 1|X, Z) = {} & C_1(F_1(\kappa_{1,j-1} - X\beta_1), 1; \theta_1) - C_1(F(\kappa_{1,j} - X\beta_1), 1; \theta_1) \\
& - C_1(F_1(\kappa_{1,j} - X\beta_1), F_2(-Z\gamma); \theta_1) \\
& + C_1(F_1(\kappa_{1,j-1} - X\beta_1), F_2(-Z\gamma); \theta_1)
\end{aligned}
$$

where $C_0(u, v)$ and $C_1(u, v)$ are copula functions and F_1 and F_2 are marginal functions which can be either normal or logistic distribution which correspond to the Probit and Logit models, respectively.

4 Results and Discussion

A total of three models were estimated using the Independence copula, the Normal copula, and the Frank copula. We selected the best fitted model based on Akaike Information Criteria (AIC), which is the Frank copula model. Table 4 shows the log-likelihood values for the Independence copula and the Frank copula models. A likelihood ratio test rejects the Independence copula model in the Frank copula model. Therefore, the results provided here are only from the Independence copula and the Frank copula models. In the following subsection, we will discuss the results from the Frank copula model for the policy implications.

4.1 Binary Choice Equation for Alcohol Consumption

Table 3 gives the results of the selection equation. The results of the binary outcome equation of self-selected alcohol consumption provide the effects of the variable on the propensity toward alcohol consumption relative to non-alcohol consumption. All the parameter estimates were statistically significant at the standard level. The coefficient interpretations are: (1) young individuals, individuals income, individuals who are married are more likely to consume alcohol; (2) males are more likely to use alcohol than females; (3) individuals who work in the agricultural sector and work in high risk occupations such as the police and military are more likely to use alcohol than white-collar workers and those in the labor sector.

Table 3 Estimation results of selection equation for alcohol consumption

Variables	Independent		Frank	
	Coeff.	Std.err	Coeff.	Std.err
Selection equation				
Intercept	−0.347	0.050	−0.343	0.050
Age	−0.014	0.001	−0.014	0.001
Income	1.34E−05	1.88E−06	1.36E−05	1.88E−06
Sex	0.946	0.02	0.946	0.020
Agr	0.345	0.032	0.347	0.032
Whi	0.200	0.061	0.203	0.061
Police	0.315	0.088	0.317	0.088
Labor	0.181	0.029	0.184	0.029
Bachelor	−0.044	0.044	−0.042	0.045
Married	0.048	0.022	0.048	0.022

4.2 Factors Affecting Hypertension Level for Non-alcohol Users

Table 4 presents the model estimation results of hypertension levels for non-alcohol users. The significant explanatory variables included in the model are age, income, high cholesterol, gender, non-communicable diseases, occupation, education, and Body Mass Index. The coefficient interpretations are: (1) older individuals, individuals with lower income, individuals who have higher cholesterol 200 mg/dl, or the number of chronic diseases, or obese (BMI $<$ 25) individuals, or individuals with education lower than bachelor degree level are more likely to develop hypertension; (2) females are more likely to have hypertension than males; (3) individuals who work in the agricultural sector are more likely to have hypertension.

4.3 Factors Affecting Hypertension Level for Alcohol Users

Table 4 also presents the model estimation results of hypertension levels for alcohol users. The explanatory variables included in the model that are significant are age, income, high cholesterol, gender, non-communicable diseases, occupation, education, and Body Mass Index (BMI). The coefficient interpretations are: (1) older individuals, individuals who have a higher health quality assessment, individuals who have higher cholesterol 200 mg/dl, or higher numbers of chronic diseases, or individuals who are obese (BMI $<$ 25), individuals whose educational levels are lower than bachelor degree, and individuals who are married are more likely to have

Table 4 Estimation results of blood pressure level equations

Regime	Independent				Frank			
	Alcohol user		Non-alcohol user		Alcohol user		Non-alcohol user	
Variables	Coeff.	Std.err	Coeff.	Std.err	Coeff.	Std.err	Coeff.	Std.err
Intercept	−1.696	0.075	−1.718	0.100	−1.838	0.073	−1.701	0.126
Age	0.026	0.001	0.022	0.001	0.027	0.001	0.022	0.001
Income	−1.68E−06	2.38E−06	4.26E−06	2.44E−06	−4.91E−06	2.28E−06	4.12E−06	2.51E−06
Sex	0.198	0.023	0.538	0.032	−0.138	0.046	0.528	0.051
Agr	−0.010	0.035	−0.040	0.047	−0.083	0.036	−0.044	0.051
Whi	−0.066	0.075	0.033	0.086	−0.102	0.072	0.031	0.087
Police	−0.038	0.126	0.041	0.108	−0.112	0.115	0.038	0.109
Labor	0.080	0.029	0.124	0.044	0.039	0.029	0.122	0.045
Bachelor	−0.176	0.053	−0.195	0.061	−0.156	0.051	−0.195	0.062
Married	−0.032	0.022	−0.085	0.032	−0.042	0.022	−0.085	0.032
qlhealth	0.006	0.012	−0.040	0.017	0.006	0.011	0.040	0.017
NCD	0.149	0.011	0.160	0.019	0.143	0.011	0.160	0.019
pe_bmi25	0.481	0.022	0.511	0.031	0.460	0.022	0.512	0.031
pe_tc200	0.252	0.022	0.250	0.028	0.236	0.021	0.245	0.028
$\kappa_{s,1}$	1.028	0.014	1.045	0.018	0.976	0.022	1.045	0.018
θ					−2.583	0.588	−0.097	0.412
LL	−30334.15				−30324.16			

Table 5 Predicted probabilities of blood pressure level within sample for the Frank copula model

Outcome	Mean of predicted probabilities						
	Sober			Drink			Difference of mean
	Min	Mean	Max	Min	Mean	Max	
Normal	0.035	0.52	0.942	0.009	0.404	0.907	0.116
Moderate	0.053	0.289	0.375	0.084	0.34	0.399	−0.051
High	0.005	0.191	0.797	0.009	0.256	0.904	−0.065

hypertension; (2) males are more likely to hypertension than females; (3) individuals who work in the agricultural sector are more likely to have hypertension.

4.4 Effect of Alcohol Consumption on Blood Pressure Level

From Table 4, the dependence parameter θ tells us about the direction of self-selection biases. The t-test is used for hypothesis testing. The null hypothesis that $\theta = 0$ implies that there is no self-selection bias. If the null is rejected, the quantification of the selection effects can be computed by comparing the outcome distribution of $\Pr(Y_0 = j|S = 1)$ with the counterfactual predicted distribution $\Pr(Y_0 = j|S = 0)$ of an individual who chooses to consume alcohol but is hypothetically allocated to non-alcohol user regime [8].

The parameter θ for alcohol user regime is negative and significant. This indicates that the two random errors (ε_0 and ν) tend to move in the opposite direction. The negative correlation means that the alcohol user counterfactual blood level of those who actually chose not to drink are below than that of an average. For the non-alcohol user regime, the dependence parameter is not significant, indicating that the blood pressure level distribution of those who are non-alcohol users do not differ from the distribution of an arbitrary individual with the same characteristics.

To quantify the effect of alcohol consumption on the probability of each level of blood pressure, we compute the average of predicted probabilities of each level of blood pressure for alcohol user and non-alcohol user regimes. Table 5 shows the results of the mean of predicted probabilities of outcome within the sample. The results indicate that alcohol consumption is more likely to lead to higher level of blood pressure and hypertension disease.

5 Concluding Remarks

This paper applied a copula-based endogenous switching regression for ordinal outcomes to examine the effect of alcohol consumption on levels of hypertension, using the data from National Health Examination Survey in 2009. We present the Frank copula and Independence copula models in this paper. We found statistical evidence for positive self-selection on alcohol users

From the empirical results previously discussed, the following are the recommended policy designs to reduce the levels of hypertension in alcohol users. In non alcohol users, other policies may help reduce levels of hypertension:

(a) The protection and prevention program for hypertension for non-alcohol users, should focus more on the needs of women, individuals from lower income groups and those with lower educational levels.
(b) For alcohol users, the protection and prevention program for hypertension should be focus more on male alcohol users, undergraduates, and manual workers.
(c) The high risk group includes individuals who are older,have higher numbers of chronic diseases, high cholesterol, and who may be obese. These people should receive regular follow-up medical examinations and take appropriate measures, including life-style changes to prevent hypertension.

References

1. Antonakis, J., Lalive, R.: Counterfactuals and causal inference: methods and principles for social research. In: Morgan, S.L., Winship, C. (eds.) Structural Equation Modeling, pp. 152–159. Cambridge University Press, Cambridge (2011)
2. Fuchs, F.D., Chambless, L.E., Whelton, P.K., Nieto, F.J., Heiss, G.: Alcohol consumption and the incidence of hypertension. Hypertension **37**, 1242–1250 (2001)
3. Genest, C., Neslehova, J.: A primer on copulas for count data. Astin Bull. **37**, 475–515 (2007)
4. Heckman, J.: Sample selection bias as a specification error. Econometrica **47**, 153–161 (1979)
5. Howteerakul, N., Suwannapong, N., Sittilerd, R., Rawdaree, P.: Health risk behaviours, awareness, treatment and control of hypertension among rural community people in Thailand. Asia Pac. J. Public Health **18**, 3 (2006)
6. Klatsky, A.L.: Alcohol and hypertension. Health effects of moderate alcohol consumption: a paradigmatic risk factor. Clin. Chim. Acta. **246**(12), 91–105 (1996)
7. Leelarasamee, A., Aswapokee, N., Muangmanee, L., Charoenlarp, K., Jaroonvesama, N.: Prevalence of hypertension in Thai population in Bangkok. J. Med. Assoc. Thail. **61**(5), 301–308 (1968)
8. Luechinger, S., Stutzer, A., Winkelmann, R.: Self-selection models for public and private sector job satisfaction. Res. Labor Econ. **30**, 233–251 (2010)
9. Mlunde, L.: Knowledge, attitude and practices towards risk factors for hypertension in Kinondoni Municipality, Dar es Salaam. DMSJ **14**(2), 59 (2007)
10. Morgan, S.L., Winship, C.: Counterfactuals and Causal Inference: Methods and Principles for Social Research. Cambridge University Press, New York (2007)
11. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer, New York (2006)
12. Oyunbileg, D., Bolorm, I., Tsolmon, U., Chimedsuren, O.: Environmental and occupational risk factors effect to arterial hypertension: correlation between arterial hypertension and lead. J. Geosci. Env. Prot. **3**, 60–65 (2015)
13. Pati, K.: Hypertension in Nan Province. Faculty of Graduate Studies. Mahidol University (2004)
14. Pechacek, T.F., Asma, S., Blair, N., Eriksen, M.P.: Tobacco, global and community solutions. In: Yusuf, S., Cairns, J.A., Camm, A.J., Fallen, E.L., Gersh, B.J. (eds.) Evidence-Based Cardiology, vol. 10313, 2nd edn. BMJ Books, London (2003)

15. Puddey, I., Lawrence, B., Beilin, J., Vandongen, R., Rouse, I.L., Rogers, P.: Evidence for a direct effect of alcohol consumption on blood pressure in normotensive men. A randomized controlled trial. Hypertension **7**(5), 707 (1985)
16. Rosenbaum, P.R., Rubin, D.R.: The central role of the propensity score in observational studies for causal effects. Biometrika **70**(1), 41–55 (1983)
17. Shenyang, G., Fraser, M.W.: Propensity Score Analysis. Statistical Methods and Applications. SAGE Publications, Los Angeles (2010)
18. Sirisrisakulchai, J., Sriboonchitta, S.: Factors affecting hospital stay involving drunk driving and non-drunk driving in Phuket, Thailand, modeling dependence in econometrics. Adv. Intell. Syst. Comput. **251**, 479–489 (2014)
19. Siviroj, P., Pengpid, S., Peltzer, K., Morarit, S.: Alcohol use and associated factors among older adults in northern Thailand. J. Hum. Ecol. **42**(2), 187–193 (2013). (Kamla-Raj)
20. Smith, M.: Using copulas to model switching regimes with an application to child labour. Econ. Rec. **81**, 547–557 (2005)
21. Venkataraman, R., Kumar, S., Kumaraswamy, M., Singh, R., Pandey, M., Tripathi, P., George, V.J., Dahal, P., Vaibhav, P.: Smoking, alcohol, and hypertension. Int. J. Pharm. Pharm. Sci. **5**(4), 28–32 (2013)
22. The Permanent Secretary Ministry of Thailand. Statistical Thailand. Bureau of Policy and Strategy (2013)
23. van Ophem, H.A.: Modeling selectivity in count-data models. J. Bus. Econ. Stat. **18**(4), 503–511 (2000)
24. Xin, X., He, J., Frontini, M.G., Ogden, L.G., Motsamai, O.I., Whelton, P.K.: Effects of alcohol reduction on blood pressure. A meta-analysis of randomized controlled trials. Departments of Epidemiology and Biostatistics, School of Public Health and Tropical Medicine, and Department of Medicine, School of Medicine, Tulane University, New Orleans, La; and Louisiana State Office of Public Health (M.G.F.). New Orleans. Hypertension **38**, 1112–1117 (2001)
25. Yadav, S., Boddula, R., Genitta, G., Bhatia, V., Bansal, B., Kongara, S., Julka, S., Kumar, A., Singh, H.K., Ramesh, V., Bhatia, E.: Prevalence and risk factors of pre-hypertension and hypertension in an affluent north Indian population. Departments of Endocrinology and Pathology. Indian J. Med. Res. **128**, 712–720 (2008)

Curriculum Vitae

Author's name	Mr. Kanchit Suknark
Date of Birth	2 April 1973
Place of Birth	Bangkok, Thailand
Education	2006 Master of Science (Public Policy and Human Development), University Maastricht, The Netherlands
	2005 Master of Public Health, Naresuan University, Thailand
	2002 Master of Economics, Thammasat University, Thailand
	1995 Bachelor of Science (Agricultural Economics), Kasetsart University, Thailand
Scholarship	2005-2006 International Labour Organization (ILO)
	2009-2012 Naresuan University
Experience	Lecturer, Faculty of Business Economics and Communication, Naresuan University, Phitsanulok, since 2002