

Impact of Lifestyles and Social Attitudes on Happiness Based on CGSS Phase II Modular Data

Bojuan Zhao

Bo.juan_zhao@t.jufe.edu.cn

Tianjin University of Finance and Economics <https://orcid.org/0000-0001-5880-080X>

Research Article

Keywords: Bayesian, BRPN, FBRPN, modular data, stackable

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11264404/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Impact of Lifestyles and Social Attitudes on Happiness Based on CGSS Phase II Modular Data

Bojuan Barbara Zhao^{a*}

^aDepartment of Statistics, Tianjin University of Finance and Economics, Tianjin, China

***Corresponding authors:**

Bojuan Zhao, Ph.D. Professor of Statistics

N9301, 25 Zhujiang Road

Hexi District, Tianjin 300222, P.R. China

E-mail address: bojuan_zhao@tjufe.edu.cn

Orcid number: 0000-0001-5880-080X

Running title: Impact of lifestyles and social attitudes on happiness

Acknowledgments: The author would like to thank the CGSS team for collecting the data and allowing it to be used for public research.

Funding Declaration: no funding is received for this project.

Declarations of interest: none.

Impact of Lifestyles and Social Attitudes on Happiness Based on CGSS Phase II Modular Data

Abstract

The Bayesian BRPN and FBRPN variable screening procedures were applied within survey-year-adjusted logistic regression models to identify important and meaningful protective and risk factors for happiness within the lifestyle, social attitude, and socio-demographic modules of the CGSS data, analyzed both individually and collectively. For the modules collectively, in descending order of importance, the protective predictors include perceived fairness of society, trustworthiness of most people, participating in physical exercises, ideal number of children, relaxation, socializing, gathering with relatives, TV usage, total individual income, female gender, browsing the Internet, watching TV or DVDs, equal sharing of household chores, lifelong learning, application for Party membership, political affiliation, overnight stays away from home, education level, listening to music at home, shopping, magazine usage, and tolerance of premarital sex; the risk factors include acceptance of extramarital sex, being taken advantage of if not vigilant, prioritizing layoffs of females, prioritizing marriage over career, and assuming men surpass women in capability. These ranked factors, together with survey year, age groups, and elderly care responsibilities, constitute the representative variables of the three modules that significantly and meaningfully affect happiness. Finally, the origins of the Bayesian – Frequentist debate are discussed to motivate the use of the approaches.

Keywords: Bayesian; BRPN; FBRPN; modular data; stackable

Subject classification codes:

1. Introduction

The Chinese General Social Survey (CGSS) was launched in 2003 and adopted a modular questionnaire starting from 2010. The modules include socio-demographic, housing, health, migration, lifestyle, social attitude, class identification, political participation (behaviour and attitude), individual cognitive ability, labour market, social

welfare, and family. Eight sets of cross-sectional data were collected in 2010, 2011, 2012, 2013, 2015, 2017, 2018 and 2021. Due to missing items in the 2011 questionnaire, only the questions valid for everyone to answer from the seven valid waves were used in the analyses. In the social attitude module, question A36 asks, “Overall, how happy do you think your life is?” with five answer options: 1 – very unhappy, 2 – pretty unhappy, 3 – hard to say, 4 – pretty happy, and 5 – very happy. Based on these responses, a binary variable *hap* was defined as 1 if $a36 \geq 4$ and 0 otherwise.

Since the modules represent distinct aspects, we examine their impacts on happiness both individually and collectively. Due to word limits, this paper focuses on three modules: lifestyle, social attitude, and socio-demographic. We apply the Bayesian BRPN and FBRPN procedures (Zhao, 2026) within survey-year-adjusted logistic regression models to identify the most efficient models with important and meaningful protective and risk factors for each module, and check whether the representative variables of the modules are nearly stackable. Before the analysis, we conducted a literature review to verify the expected direction and significance of the Bayesian priors, i.e., the protective or risk factors for happiness across the three modules.

According to the World Happiness Report (2017-2023), based on the three-year rolling averages of global life evaluations, China ranked 79th(out of 155), 86th(156), 93rd(156), 94th(153), 84th(149), 72nd(146), and 64th(137) among the surveyed countries (or regions), indicating an improving trend in happiness. Using CGSS data from 2012, 2013, and 2015, Xu et al. (2023) found that social trust significantly enhances rural residents' happiness via increased social capital. Drawing on CGSS 2013 data, Zheng et al. (2018) identified a significant positive correlation between number of children and happiness among rural residents. Based on CGSS 2015 data, Shi et al. (2023) concluded

that physical activity significantly and positively affects both happiness and perceived social trust. Su et al. (2021) found that TV viewing and Internet use both positively influence happiness. Li et al. (2022) reported a U-shaped age-happiness curve for the urban population and an S-shaped curve for the rural population, with the latter showing a decline in happiness for the oldest-old (over 85). Furthermore, based on CGSS 2017 data, Yan et al. (2023) found that Internet use is significantly and positively associated with happiness, and that this impact varies across age groups: middle-aged adults benefit from more frequent use and larger social networks, while younger and older adults gain more from group-based communications.

Using the China Labour-force Dynamics Survey (CLDS) data from 2012 and 2014, Xu and Chen (2017) reported that both income and fairness have a significant positive effect on happiness. Based on the Survey of the Chinese Economic Life (SCEL) in 2011, Wei et al. (2015) found that happiness is significantly and positively influenced by income, leisure time, and passive leisure activities (e.g., watching TV and browsing the Internet), whereas it is not significantly associated with active leisure activities (e.g., exercising, socializing, and shopping) or age. Additionally, females are happier than males, and education level exhibits a non-linear association with happiness.

Based on a survey conducted in 29 Asian countries, Tokuda et al. (2010) concluded that female gender, age 20-29 or 60-69, being married, high income, high education, religious belief, good health, and high social trust are significantly and positively associated with happiness. Furthermore, compared to people living in low-social-trust areas, those in high-social-trust areas report higher levels of happiness. Using data from a national leisure survey in South Korea, Kang et al. (2018) found that leisure satisfaction, vacation frequency, club participation, volunteering, gender, marital

status, household income, and education level are significantly associated with happiness.

In the following sections, we apply the Bayesian BRPN and FBRPN procedures to survey-year-adjusted logistic regression models using CGSS modular data, and propose model-building strategies for analyzing the modules individually and collectively. The final section presents the discussion and conclusions.

2. The Bayesian variable screening procedures and model-building strategies for modular data

For the seven waves of cross-sectional data, consider the survey-year-adjusted logistic regression model

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \beta_0 + \sum_{i=1}^m \beta_i X_i + \alpha \cdot year, \quad (1)$$

where $\hat{\pi}$ is the probability of being happy ($hap=1$); $\beta_0, \beta_1, \dots, \beta_m$ and α are unknown parameters; $X_1, \dots, X_m$ represent predictors from the lifestyle, social attitude and socio-demographic modules; and $year$ is the survey year. For simplicity, model (1) is referred to as a simple or multiple year-adjusted logistic regression model for $m=1$ or $m \geq 2$, respectively. For the data, we obtain the fitted parameter $\hat{\alpha} = 0.047$ ($p < 0.001$) for $m = 0$. Note that though the relationship between $year$ and the proportion of ($hap=1$) appears as a check-mark shape, with the lowest value in 2013, if $year^2$ is added, the term is non-significant.

Zhao (2026) proposed two Bayesian variable screening procedures, FBRPN (locally significant Forward and Backward procedure with Restrictions on the Positive or Negative effects) and BRPN (Backward procedure with Restrictions on the Positive or Negative effects) to identify important and meaningful protective and risk factors.

The steps of the FBRPN procedure in model (1) are: (a) Include all predictors that are significant ($p < 0.05$) in fitted simple year-adjusted logistic regression models into a multiple year-adjusted logistic regression model, and fit the model to the data. (b) Compare the estimates from the fitted simple and multiple year-adjusted logistic regression models; remove the most significant predictor among those whose effects change direction. If no predictor has opposite effects, remove the least significant predictor ($p > 0.05$) among those with the same sign of effects. Only one predictor is removed per iteration. (c) Refit the model with the remaining predictors. (d) Repeat steps (b) and (c) until no further predictors can be removed. For the BRPN procedure, the steps are identical to those of the FBRPN procedure, except for step (a), where all predictors including non-significant ones from the fitted simple year-adjusted logistic regression models are entered into a multiple year-adjusted logistic regression model and fitted to the data. Note that both procedures are essentially backwards, and the predictors retained in the final fitted models are the steadiest ones that have defeated all other variables fairly and sequentially, while preserving their meanings as protective or risk factors.

Starting with the common predictors in the FBRPN and BRPN models, we add each remaining predictor. A new model is selected only if the added predictor yields a significant and meaningful effect and the most efficient reduction in deviance, evaluated using a chi-squared test for the change in deviance. This forward selection procedure is iterated until no remaining predictor can be added; at that point, we obtain the most efficient model. Note that this procedure can be stepwise, as missing observations can cause a previously included predictor to become non-significant.

The variables in the most efficient model are referred to as the representative variables (RVs) of the original predictors. The order of $|Z|$ -values (or $|T|$ -values) of their

estimates shows the order of importance of the RVs (excluding multi-category predictors). For a multi-category predictor, a low or negative Pearson (or Spearman) correlation between the two vectors of estimates from the fitted simple and multiple year-adjusted logistic regression models may indicate inconsistent effect directions. Consequently, each predictor requires individual evaluation. To facilitate computation and importance ranking, we minimize the number of multi-category predictors.

For the modular data, suppose the q modules ($A_1, A_2, \dots, A_q$) cover different substantive domains. For the combined $A_1 + A_2 + \dots + A_q$, step (a) of the BRPN procedure includes all predictors across the q modules, while step (a) of the FBRPN procedure includes all variables selected by the FBRPN procedures applied to each of the q modules. Starting with the common predictors from the BRPN and FBRPN models, we sequentially add important and meaningful predictors to derive the most efficient model, with model comparisons evaluated using chi-squared tests.

Unlike the Bayesian variable selection methods that assume “prior and posterior distributions” (e.g., Mitchell and Beauchamp, 1988; Chen et al., 1999), the FBRPN and BRPN procedures assume “positive and negative effects” based on prior knowledge of whether variables are protective or risk factors, respectively. Compared with the backward or stepwise methods in the textbooks, the new procedures are Bayesian, in that not only the significance but also the directions (signs) of effects are considered. Compared with biased methods such as ridge regression, Lasso, and LAR, the unbiased BRPN procedure performs best among the compared methods in cross-validation comparisons using the Efron diabetes dataset. This dataset (Efron et al., 2004) includes 64 dummy variables, providing a fair basis for comparing different methods.

Before applying the procedures to survey-year-adjusted logistic regression models, it is essential to verify the direction (positive or negative) and significance of predictor

effects obtained from fitted simple year-adjusted logistic regression models. If inconsistencies with Bayesian priors on protective or risk factors arise, they can be temporarily ignored pending later model revision, given the relative independence of the selected representative variables. Note that these contradictions may indicate the presence of exogenous or endogenous variables, as discussed in the final section.

3. Descriptive statistics

Across the merged seven waves of data, we identified 76605 valid responses to question A36. Within this subset, we calculated the minimum, maximum, and sample size for every variable across the three modules. For each wave, we then calculated the mean of each variable, along with the minimum and maximum values of these means across the seven waves, as presented in Table 1.

Table 1 Descriptive statistics of variables in lifestyle, social attitude and socio-demographic modules, including the minimum and maximum values of the means of the seven sets of data

Variable	(min, max)	(min, max) of the means	sample size (<i>n</i>)
Dependent: happiness (a36)	(1,5)	(3.755,3.983)	76605
happiness (<i>hap</i>)	(0,1)	(0.723,0.814)	76605
lifestyle module:			
newspaper usage (a281)	(1,5)	(1.596,2.258)	76545
magazine usage (a282)	(1,5)	(1.505,1.890)	76522
radio usage (a283)	(1,5)	(1.584, 1.872)	76483
TV usage (a284)	(1,5)	(3.413, 4.133)	76565
Internet usage (a285)	(1,5)	(1.927, 3.322)	76480
mobile news (a286)	(1,5)	(1.521, 1.995)	76443
TV as the main source of information (a29)	(0,1)	(0.362, 0.778)	74847
watching TV or DVDs (a301)	(1,5)	(3.728, 4.474)	76481
going to the cinema (a302)	(1,5)	(1.288, 1.574)	76404
shopping (a303)	(1,5)	(2.528, 2.721)	76443
reading books/newspapers/magazines(a304)	(1,5)	(2.047, 2.473)	76352
attending cultural events (a305)	(1,5)	(1.402,1.487)	76335
gathering with relatives (a306)	(1,5)	(2.070,2.252)	76435
gathering with friends (a307)	(1,5)	(2.267,2.413)	76450
listening to music at home (a308)	(1,5)	(2.224,2.842)	76377

participating in physical exercises (a309)	(1,5)	(2.067,2.828)	76386
watching live sports (a3010)	(1,5)	(1.214,1.614)	76321
making handicrafts (a3011)	(1,5)	(1.383,1.455)	76360
browsing the Internet (a3012)	(1,5)	(2.032,3.543)	76444
socializing (a311)	(1,5)	(2.634,2.853)	76507
relaxation (a312)	(1,5)	(3.321,3.578)	76498
lifelong learning (a313)	(1,5)	(1.851,2.122)	76432
overnight stays away from home (a32)	(1,6)	(1.656,1.931)	76242
social attitude module:			
trustworthiness of most people(a33)	(1,5)	(3.283,3.662)	76410
being taken advantage of if not vigilant (a34)	(1,5)	(2.916,3.066)	76253
perceived fairness of society (a35)	(1,5)	(2.986,3.464)	76348
ideal number of children log(1+a371)	(0,4.143)	(1.061,1.089)	74216
ideal number of sons log(1+a372)	(0,4.533)	(0.672,0.709)	62337
ideal number of daughters log(1+a373)	(0,2.485)	(0.657,0.669)	62293
tolerance of premarital sex (a38)	(1,5)	(1.832,2.182)	75320
acceptance of extramarital sex (a39)	(1,5)	(1.248,1.338)	75524
acceptance of homosexual (a40)	(1,5)	(1.310,1.658)	73992
elderly care responsibilities (a41)	(1,4)	(2.578,2.663)	76148
men working and women keeping house (a421)	(1,5)	(3.134,3.594)	76359
assuming men surpass women in capability (a422)	(1,5)	(2.797,3.012)	76207
prioritizing marriage over career(a423)	(1,5)	(2.949,3.123)	76030
prioritizing layoffs of females (a424)	(1,5)	(1.958,2.223)	75180
equal sharing of household chores (a425)	(1,5)	(3.721,4.038)	76291
Socio-demographic module:			
female gender (a2)	(0, 1)	(0.489,0.544)	76605
age groups (<i>agr</i>)	(1, 3)	(1.818, 2.004)	76600
ethnic group (a4)	(0, 1)	(0.073, 0.093)	76545
religious belief (a5)	(0, 1)	(0.076, 0.147)	76594
education level (a7a)	(0, 1)	(0.155,0.206)	76524
total individual income log(1+a8a)	(0,16.118)	(8.094,8.623)	70158
individual's occupational income log(1+a8b)	(0,16.118)	(5.987,6.607)	70395
application for party membership (a9)	(0, 1)	(0.140, 0.198)	76380
political affiliation (a10)	(0, 1)	(0.150, 0.197)	76415
survey year (<i>year</i>)	(2010, 2021)		76605

In Table 1, the lifestyle module includes 23 variables, defined as follows: a281-a286 (1-never to 5-very frequently); a29 (0-newspaper, magazines, radio, Internet and mobile news, 1-TV); a301-a3012 (1-never, 2-yearly or less, 3-monthly, 4-weekly, and 5-daily); a311-a313 (1-never to 5-very frequently); a32 (1-never, 2-1 to 5 nights, ..., 6-

over 30 nights). The social attitude module includes 15 variables, defined as follows: a33-a34 (1-strongly disagree to 5-strongly agree); a35 (1-completely unfair to 5-completely fair); a371-a373 (in $\log(1+x)$ scale, as they fit better in the analyses); a38-a40 (1-completely wrong to 5-completely right); a41 (1-the government, 2-children, 3-the elderly, 4-shared by the government, children, and the elderly, with proportions 9.86%, 51.57%, 5.86%, and 32.71%, respectively); a421-a425 (1-strongly disagree to 5-strongly agree). The socio-demographic module includes 10 variables, defined as follows: a2 (0-male, 1-female); *age* (survey year - year of birth); *agr* (1- $16 \leq \text{age} < 40$, 2- $40 \leq \text{age} < 65$, 3- $\text{age} \geq 65$, with proportions 29.22%, 49.53%, and 21.25%, respectively); a4 (0-Han, 1-non-Han); a5 (0-no, 1-yes); a7a (0-below college (< 9), 1-college and above (≥ 9)); a8a-a8b (heavily right-tailed, in $\log(1+x)$ scale); a9 (0-not, 1-yes); a10 (0-the masses, 1-members of the Communist Youth League, the Communist Party or democratic parties).

The sample sizes for the seven waves of data are 11767, 11724, 11380, 10953, 12561, 12772, and 5448, for the years 2010, 2012, 2013, 2015, 2017, 2018, and 2021, respectively. Due to varying missing observations across variables, the sample sizes for different fitted models also vary.

4. Individual impacts of the three modules on happiness

4.1 Impact of the lifestyle module on happiness

For each predictor in the lifestyle module, the estimates and significance levels of parameters in simple year-adjusted logistic regression models are presented in the 2nd and 5th columns of Table 2. The results show that a301 (watching TV or DVDs), a302 (going to the cinema), a303 (shopping), a304 (reading books/newspapers/magazines), a305 (attending cultural events), a306 (gathering with relatives), a307 (gathering with

friends), a308 (listening to music at home), a309 (participating in physical exercises), a3010 (watching live sports), a3011 (making handicrafts), a3012 (browsing the Internet), a311-313 (socializing, relaxation, and lifelong learning, respectively), a32 (overnight stays away from home), and a281-a286 (usage of newspapers, magazines, radio, TV, Internet, and mobile news, respectively) are all significantly and positively associated with happiness. Predictor a29 (TV as the main source of information) is significantly and negatively associated with happiness. These conclusions are consistent with common perceptions and established Bayesian priors regarding protective and risk factors.

Table 2. Impact of the lifestyle module on happiness, estimates and significance (or Z-values) using the Bayesian procedures

Variable	Simple	BRPN (FBRPN) ($ Z $)	Variable	Simple	BRPN(FBRPN) ($ Z $)
constant		-106.1***	a305	0.191***	
<i>Year</i>		0.052 (17.95)	a306	0.288***	0.114 (8.65)
a281	0.184***		a307	0.202***	
a282	0.250***	0.051 (4.39)	a308	0.151***	0.027 (3.81)
a283	0.130***		a309	0.194***	0.091 (13.21)
a284	0.183***	0.090 (8.18)	a3010	0.111***	
a285	0.109***		a3011	0.065***	
a286	0.115***		a3012	0.099***	
a29 (1 vs.0)	-0.227***		a311	0.191***	0.089 (10.01)
a301	0.160***	0.073 (6.82)	a312	0.227***	0.133 (14.11)
a302	0.191***		a313	0.263***	0.130 (12.12)
a303	0.166***	0.032 (3.73)	a32	0.114***	0.030 (4.09)
a304	0.183***				

Notes: Symbols '***', '**' and '*' represent significant at levels of 0.1%, 1% and 5%, respectively

Since all the estimates in the 2nd and 5th columns are significant, the FBRPN and BRPN procedures yield identical results. Specifically, a3010, a305, a281, a302, a286, a3011, a285, a307, a3012, a29, a283, and a304 are excluded from the final model due to

opposite or non-significant effects. The final fitted model is presented in the 3rd and 6th columns of Table 2. In descending order of importance, relaxation (a312), participating in physical exercises (a309), lifelong learning (a313), socializing (a311), gathering with relatives (a306), TV usage (a284), watching TV or DVDs (a301), magazine usage (a282), overnight stays away from home (a32), listening to music at home (a308), and shopping (a303), together with survey year (*year*), constitute the representative variables (RVs) of the module that significantly and meaningfully affect happiness.

4.2 Impact of the social attitude module on happiness

For each predictor in the social attitude module, the estimates and significance levels of parameters in simple year-adjusted logistic regression models are presented in the 2nd and 5th columns of Table 3. The results show that a33 (trustworthiness of most people), a35 (perceived fairness of society), a38 (tolerance of premarital sex), a425 (equal sharing of household chores), a371-a373 (ideal numbers of children, sons, and daughters) are all significantly and positively associated with happiness; predictors a34 (being taken advantage of if not vigilant), a39 (acceptance of extramarital sex), a421 (men working and women keeping house), a422 (assuming men surpass women in capability), a423 (prioritizing marriage over career), and a424 (prioritizing layoffs of females) are all significantly and negatively associated with happiness. Predictor a40 (acceptance of homosexual) is negatively but non-significantly associated with happiness. Predictor a41 (elderly care responsibilities) is significantly associated with happiness. Specifically, compared with those choosing the government category, individuals in each of the other three categories (children, the elderly, and shared responsibility among the government, children, and the elderly) are significantly happier. These conclusions are consistent with common perceptions and Bayesian priors

on protective or risk factors.

Using the BRPN procedure, all variables in the 2nd and 5th columns were included in step (a), and a372 and a40 were removed due to opposite or non-significant effects. Using the FBRPN procedure, significant variables in the 2nd and 5th columns were included in step (a), and a372 was removed due to its opposite effect. The two final fitted models are identical, as shown in the 3rd and 6th columns of Table 3. In descending order of importance, perceived fairness of society (a35), trustworthiness of most people (a33), prioritizing layoffs of females (a424), tolerance of premarital sex (a38), acceptance of extramarital sex (a39), equal sharing of household chores (a425), assuming men surpass women in capability (a422), being taken advantage of if not vigilant (a34), prioritizing marriage over career (a423), ideal number of daughters (a373), men working and women keeping house (a421), and ideal number of children (a371), together with survey year (*year*) and elderly care responsibilities (a41), constitute the RVs of the module that significantly and meaningfully affect happiness.

Table 3. Impact of the social attitude module on happiness, estimates and significance (or $|Z|$ -values) using the Bayesian procedures

Variable	Simple	BRPN(FBRPN) ($ Z $)	Variable	Simple	BRPN(FBRPN) ($ Z $)
Constant		-53.521***	a40	-0.003	
<i>Year</i>		0.026 (7.86)	a41(2 vs.1)	0.503***	0.336 (9.86)
a33	0.367***	0.209 (20.48)	(3 vs.1)	0.567***	0.468 (8.46)
a34	-0.196***	-0.060 (6.00)	(4 vs.1)	0.383***	0.219 (6.15)
a35	0.592***	0.525 (50.92)	a421	-0.065***	-0.028 (2.71)
log(1+a371)	0.286***	0.125 (2.57)	a422	-0.102***	-0.063 (6.10)
log(1+a372)	0.129***		a423	-0.128***	-0.052 (5.31)
log(1+a373)	0.331***	0.145 (3.08)	a424	-0.157***	-0.123 (11.14)
a38	0.028***	0.098 (9.59)	a425	0.113***	0.089 (8.96)
a39	-0.146***	-0.151 (9.23)			

Notes: Symbols '***', '**' and '*' represent significant at levels of 0.1%, 1%, and 5%, respectively

4.3 Impact of the socio-demographic module on happiness

For each predictor in the socio-demographic module, the estimates and significance levels of parameters in simple year-adjusted logistic regression models are presented in the 2nd and 6th columns of Table 4. The results show that a2 (female gender), a7a (education level), a8a (total individual income), a8b (individual's occupational income), a9 (application for party membership), and a10 (political affiliation) are all significantly and positively associated with happiness. Age groups (*agr*) is significantly associated with happiness; compared with youth, middle-aged people are significantly less happy, and although the elderly are happier, the difference is non-significant. Ethnic group (a4) and religious belief (a5) are not significantly associated with happiness. Non-Han people are slightly happier than Han people, and religious believers are slightly less happy than non-religious believers. These conclusions are consistent with common perceptions and literature reviews, except for religious belief (a5). In subsequent analyses, we first identify the most efficient model under a prior assuming a negative effect for a5, and then re-evaluate the model under a positive prior assumption for a5.

Using the BRPN procedure, all variables in the 2nd and 6th columns were included in step (a), and a8b and a5 were removed due to opposite effects. Using the FBRPN procedure, significant variables in the 2nd and 6th columns were included in step (a), and a8b was removed due to its opposite effect. The final fitted BRPN and FBRPN models are presented in Table 4.

The BRPN model reduces the deviance by 1537.534 (from 75708.437 to 74170.903) using 9 parameters (df: 69705 vs. 69696). Similarly, the FBRPN model reduces the deviance by 1533.785 (from 75778.373 to 74244.588) using 8 parameters (df: 69756 vs. 69748). A likelihood ratio test comparing the two models yields a p-value

of 0.053 ($\chi^2 = 3.749$, $df = 1$). Therefore, the FBRPN model is slightly more efficient than the BRPN model.

Note that the BRPN and FBRPN models are fitted on different sample sizes (69706 and 69757, respectively). When the FBRPN model is refitted using the same subset of data as the BRPN model, the deviance is reduced by 1533.189 (from 75708.437 to 74175.248) using 8 parameters (df : 69705 vs. 69697). A likelihood ratio test yields a p-value of 0.037 ($\chi^2 = 4.345$, $df = 1$), which is consistent with the Z-value (2.07) in the 3rd column of Table 4. That is, the BRPN model is obtained via backward selection, while the FBRPN model is obtained via forward selection. To retain more valid samples, we consider the FBRPN model to be the most efficient. Ethnic group (a4) is a predictor approaching the conventional significance threshold in the model.

When the Bayesian prior for religious belief (a5) is specified as positive, the fitted FBRPN+a5 model yields a positive effect for a5 (p-value=0.041); however, the reduced deviance is only 3.287, which is not significant using the $\chi^2_{(1)}$ test. The situation is similar to that for a4. That is, regardless of whether a positive or negative prior is assigned to a5, the FBRPN model remains the most efficient.

Table 4. Impact of the socio-demographic module on happiness, estimates and significance (or |Z|-values) using the Bayesian procedures

Variable	simple	BRPN (Z)	FBRPN (Z)	Variable	simple	BRPN (Z)	FBRPN (Z)
Constant		-79.42***	-79.29***	a5(1 vs.0)	-0.001		
<i>Year</i>		0.040(13.85)	0.040(13.84)	a7a(1 vs.0)	0.573***	0.336(10.96)	0.332(10.87)
a2(1 vs.0)	0.071***	0.192(10.20)	0.192(10.21)	log(1+a8a)	0.034***	0.032(12.22)	0.032(12.21)
<i>agr</i> (2vs.1)	-0.276***	-0.152(6.72)	-0.153(6.77)	log(1+a8b)	0.008***		
(3 vs.1)	0.035	0.172(6.04)	0.169(5.94)	a9(1 vs.0)	0.650***	0.283(7.08)	0.282(7.07)
a4(1 vs.0)	0.032	0.070(2.07)		a10(1 vs.0)	0.667***	0.334(8.22)	0.335(8.24)

Notes: Symbols '***', '**' and '*' represent significant at levels of 0.1%, 1% and 5%, respectively

Note that we calculated the relative change in effects with and without a4 (or a5), $R_{a4}=100 \times ((\text{FBRPN}+a4)-\text{FBRPN})/(\text{BRPN}+a4)$. We find that $|R_{a4}| < 5\%$ and $|R_{a5}| < 5\%$ for all fitted effects. That is, the impacts of the representative variables (RVs) in the FBRPN model are not much affected by the inclusion of a4 (or a5), suggesting relative independence between these added predictors and the RVs. In descending order of importance, total individual income (a8a), education level (a7a), female gender (a2), political affiliation (a10), and application for party membership (a9), together with survey year (*year*) and multi-category age groups (*agr*), constitute the RVs of the module, significantly and meaningfully affecting happiness.

5. Impact of lifestyles and social attitudes on happiness after adjusting for socio-demographics

Using the BRPN procedure, all variables from the three modules were initially included in step (a). Predictors a3010, a5, a372, a302, a421, a286, a305, a8b, a3011, a281, and a29 were sequentially removed due to opposite effects, and a304, a373, a4, a283, a3012 were sequentially removed due to non-significant effects. Using the FBRPN procedure, variables listed in the FBRPN columns of Tables 2-4 were included in step (a), and predictors a421 and a373 were removed due to opposite and non-significant effects, respectively. The final fitted models are shown in Table 5.

The BRPN model reduces the deviance by 8292 (from 67696 to 59404) with 35 parameters (df: 62979 vs. 62944). Similarly, the FBRPN model reduces the deviance by 8390 (from 68956 to 60566) with 32 parameters (df: 64117 vs. 64085). The FBRPN model is more efficient, with three fewer variables (a285, a307 and a40) than the BRPN model.

Starting with the FBRPN model, each remaining predictor was added. The largest reduction in deviance is 39.83 for a3012 ($\beta=0.056$, $Z=6.78$), followed by 37.72 for a285

($\beta=0.065$, $Z=7.00$), both of which are significant using the $\chi^2_{(1)}$ test, and the effects are significant and meaningful at the same time. In the second step, no remaining predictors could be added to the FBRPN+a3012 or FBRPN+a285 model. The concepts of a3012 (browsing the Internet) and a285 (Internet usage) are close.

In fact, model FBRPN+a3012 ($n=64075$) yields a reduced deviance of 8429 with 33 parameters, while model FBRPN+a285 ($n=64067$) yields 8427 with 33 parameters. On the other hand, model FBRPN+a3012 +a285 ($n=64034$) yields a reduced deviance of 8427 with 34 parameters, where $\beta = 0.026$ ($Z=1.770$) for a3012 and $\beta = 0.040$ ($Z = 2.411$) for a285. Using the same fixed sample size of 64034, the reduced deviance is 8421 for the FBRPN+a3012 model and 8424 for the FBRPN+a285 model. These results indicate that when searching for the most efficient model, the FBRPN+a3012 model is obtained via forward selection, while the FBRPN+a285 model is obtained via backward selection. The divergence stems from varying sample sizes.

To retain more valid samples, we considered FBRPN+a3012 (Model A) as the most efficient model; the estimates along with their absolute $|Z|$ -values are shown in Table 5. Note that in Table 2, browsing the Internet (a3012) is not selected as one of the RVs for the lifestyle module, although a3012 proved statistically significant in the fitted simple year-adjusted logistic regression. To obtain further insights, we calculated the relative changes in effects between Model A (with a3012) and FBRPN (without a3012), defined as $R_a=100 \times (\text{Model A} - \text{FBRPN}) / (\text{Model A})$, as shown in the 5th column of Table 5. Results indicate that most changes are negligible ($|R_a| < 5\%$), except for the following predictors: *year* (-32.20%), a282 (-20.10%), a284 (5.60%), a303 (-12.26%), a306 (-5.23%), a308 (-40.76%), a313 (-16.94%), a32 (-15.38%), a38 (-24.94%), a422 (-8.81%), *agr* (2 vs. 1, -48.50%), *agr* (3 vs. 1, 113.39%), a7a (-31.37%), and a8a (-7.60%). Specifically, the inclusion of browsing the Internet (a3012) weakens the negative impact

of assuming men surpass women in capability (a422) on happiness and strengthens the positive effect of TV usage (a284). Meanwhile, it reduces several positive effects on happiness, including survey year (*year*), magazine usage (a282), shopping (a303), gathering with relatives (a306), listening to music at home (a308), lifelong learning (a313), overnight stays away from home (a32), tolerance of premarital sex (a38), education level (a7a), and total individual income (a8a).

In descending order of importance, perceived fairness of society (a35), trustworthiness of most people (a33), participating in physical exercises (a309), ideal number of children (a371), relaxation (a312), acceptance of extramarital sex (a39), socializing (a311), gathering with relatives (a306), TV usage (a284), total individual income (a8a), being taken advantage of if not vigilant (a34), female gender (a2), browsing the Internet (a3012), watching TV or DVDs (a301), prioritizing layoffs of females (a424), equal sharing of household chores (a425), lifelong learning (a313), application for party membership (a9), political affiliation (a10), overnight stays away from home (a32), education level (a7a), listening to music at home (a308), prioritizing marriage over career (a423), assuming men surpass women in capability (a422), shopping (a303), magazine usage (a282), and tolerance of premarital sex (a38), together with *year*, age groups (*agr*), and elderly care responsibilities (a41), constitute the RVs of the three modules, significantly and meaningfully affecting happiness.

When the Bayesian prior for religious belief (a5) is treated as positive, the effects (with $|Z|$ -values) of the fitted Model B (=Model A+a5) are shown in the 6th column of Table 5. The effect of a5 is significantly positive, and the deviance is reduced by 7.57 when adding one parameter, which is significant using the $\chi^2_{(1)}$ test. The relative changes in effects with and without a5, defined as $R_b = 100 \times (\text{Model B} - \text{Model A}) / (\text{Model B})$, are presented in the 7th column of Table 5, with all $|R_b| < 5\%$. This indicates that the impacts

of RVs in Model A are barely affected by the inclusion of religious belief (a5), suggesting relative independence between the added predictor (a5) and the RVs in Model A. Excluding the multi-category predictors, the order of importance in Model B is largely consistent with that in Model A, and a5 ranks as the fifth least important.

Table 5. Impact of lifestyles and social attitudes on happiness after adjusting for socio-demographics, estimates and significances (or $|Z|$ -values) using the Bayesian procedures

Variable	BRPN	FBRPN	Model A($ Z $)	R _a (%)	Model B($ Z $)	R _b (%)
Constant	-57.34***	-73.49***	-56.62***	-29.79	-57.38***	1.32
survey year (<i>year</i>)	0.027***	0.035***	0.026(7.1)	-32.20	0.027(7.1)	1.41
magazine usage (a282)	0.034*	0.043**	0.036(2.6)	-20.10	0.037(2.7)	2.17
TV usage (a284)	0.089***	0.089***	0.094(7.3)	5.60	0.095(7.3)	0.55
Internet usage (a285)	0.063***					
watching TV or DVDs (a301)	0.090***	0.084***	0.085(6.7)	1.42	0.085(6.8)	0.29
shopping (a303)	0.024*	0.030**	0.027(2.7)	-12.26	0.027(2.6)	-1.21
gathering with relatives (a306)	0.108***	0.126***	0.120(7.8)	-5.23	0.119(7.8)	-0.44
gathering with friends (a307)	0.034*					
listening to music at home (a308)	0.030***	0.040***	0.028(3.3)	-40.76	0.028(3.2)	-2.61
participating in physical exercises (a309)	0.093***	0.096***	0.093(11.6)	-3.44	0.093(11.6)	-0.11
browsing the Internet (a3012)			0.056(6.8)	100	0.057(6.8)	0.91
socializing (a311)	0.078***	0.087***	0.087(8.5)	0.77	0.088(8.5)	0.47
relaxation (a312)	0.100***	0.103***	0.101(9.1)	-1.16	0.101(9.1)	-0.17
lifelong learning (a313)	0.057***	0.070***	0.060(4.6)	-16.94	0.059(4.5)	-1.31
overnight stays away from home (a32)	0.029***	0.036***	0.031(3.6)	-15.38	0.031(3.6)	-0.70
trustworthiness of most people (a33)	0.197***	0.197***	0.199(19.7)	1.04	0.200(19.7)	0.19
being taken advantage of if not vigilant(a34)	-0.066***	-0.067***	-0.068(6.9)	1.36	-0.068(6.9)	-0.23
perceived fairness of society (a35)	0.541***	0.537***	0.539(52.8)	0.52	0.540(52.9)	0.14
ideal number of children(log(1+a371))	0.336***	0.319***	0.332(9.5)	3.95	0.326(9.3)	-2.07
tolerance of premarital sex (a38)	0.030**	0.028**	0.022(2.1)	-24.94	0.022(2.1)	0.30
acceptance of extramarital sex (a39)	-0.138***	-0.146***	-0.147(9.1)	0.25	-0.147(9.1)	0.01
acceptance of homosexual (a40)	-0.042**					
elderly care responsibilities, a41(2 vs.1)	0.460***	0.449***	0.455(13.5)	1.23	0.454(13.4)	-0.17
a41(3 vs.1)	0.511***	0.500***	0.511(9.5)	2.21	0.513(9.6)	0.27
a41(4 vs.1)	0.240***	0.243***	0.244(7.0)	0.34	0.246(7.0)	0.65
assuming men surpass women in capability(a422)	-0.027**	-0.028**	-0.026(2.7)	-8.81	-0.027(2.8)	3.49
prioritizing marriage over career(a423)	-0.032***	-0.031**	-0.031(3.3)	1.77	-0.032(3.3)	0.90
prioritizing layoffs of females (a424)	-0.071***	-0.074***	-0.071(6.4)	-4.51	-0.071(6.4)	-0.64
equal sharing of household chores (a425)	0.055***	0.055***	0.056(5.6)	1.20	0.056(5.6)	-0.04
female gender, a2(1 vs.0)	0.165***	0.146***	0.150(6.9)	2.95	0.146(6.7)	-3.29
age groups, <i>agr</i> (2 vs.1)	-0.095***	-0.160***	-0.108(3.8)	-48.50	-0.108(3.8)	-0.28

<i>agr</i> (3 vs.1)	0.092*	-0.010	0.077(2.0)	113.39	0.076(2.0)	-2.11
religious belief, a5 (1 vs. 0)					0.102(3.1)	100
education level, a7a (1 vs. 0)	0.128***	0.167***	0.127(3.5)	-31.37	0.131(3.6)	2.47
total individual income (a8a)	0.021***	0.023***	0.022(7.1)	-7.60	0.021(7.1)	-0.61
application for party membership, a9(1 vs.0)	0.185***	0.189***	0.185(4.2)	-2.31	0.186(4.2)	0.18
political affiliation, a10 (1 vs.0)	0.191***	0.186***	0.186(4.1)	-0.15	0.188(4.1)	1.13

Notes: Symbols '***', '**' and '*' represent significant at levels of 0.1%, 1% and 5%, respectively

6. Discussions and conclusions

6.1 The choose for the logistic regression model

Although the response variable for question A36 consists of five ordered categories, we employed the logistic regression model (1) instead of the cumulative logit model (Agresti, 2007) typically used for ordinal responses. We consider a cumulative logit model with the following format:

$$\text{logit}[P(a36 \leq j)] = \alpha_j + \beta X, \quad j = 1, 2, 3, 4 \quad (2)$$

where X is a continuous variable, and α_j , $j=1,2,3,4$ and β are unknown parameters. If $\hat{\beta}$ in model (2) is significantly positive, it implies that as X increases, a36 is likely to take smaller values on the unhappy direction. Thus, predictor X is negatively associated with happiness (i.e., X is a risk factor for happiness). This conclusion regarding protective or risk factors is similar to that drawn from the corresponding logistic regression model (1), where the response variable is binary (happy vs. unhappy).

Model (2) expands the study to explore relationship between the five levels of happiness while identifying important and meaningful protective or risk factors. However, this approach adds unnecessary complexity. Therefore, we selected model (1) for a more direct analysis.

6.2 The concepts of endogenous and exogenous

It is worth noting that the Bayesian BRPN and FBRPN procedures provide a path for generalizing analyses from balanced orthogonal design data to observational data. These procedures are not required when analyzing data from balanced orthogonal designs. For instance, when analyzing data from an 18-run, 6-factor, 3-level balanced orthogonal design ($L_{18}(3^6)$), we can fit various linear regression models using different combinations of all six factors. Notably, the estimated effects and sum of squares for each factor remain invariant across these models due to the orthogonality and balance in the $L_{18}(3^6)$ design. In other words, the orthogonality in the design ensures that predictors are uncorrelated, and the balance ensures unbiased effect estimates, which facilitates precise estimation and strengthens the basis for causal inference. In contrast, covariates in observational data are neither orthogonal nor balanced. However, by employing the Bayesian BRPN and FBRPN procedures in modelling, we can sequentially derive relatively uncorrelated representative variables and simultaneously obtain effect estimates after adjusting for possibly unbalanced covariates. Theoretically, these adjusted estimates are closer to the true effects than unadjusted ones. We illustrate this point using religious belief (a5) and the well-known Simpson's paradox dataset from Agresti (2007).

Regarding religious belief (a5), many individuals once adhered to Buddhism, Daoism, Christianity, or Islam, though these practices were severely disrupted during the Cultural Revolution (1966 – 1976). Nowadays, the proportion of respondents with a5=1 is disproportionately low among those with a10=1 (or a9=1), a subgroup prone to being happy, which contributes to unbalanced covariate structures. In Chinese, 'religious belief' translates to 信仰 (xìn yǎng). The essence of these two characters is a belief that humanity will recover, even if it has been observed to decline. Ultimately,

one needs to look upward to track its recovery trajectory and refuse to join others in wrongdoing. Therefore, a positive association between religious belief and happiness (a5) is logically sound in China, yet unbalanced covariate structures exist.

In fact, the estimated effect of religious belief (a5) is negative and non-significant (-0.001) in simple year-adjusted logistic regression, yet becomes positive and significant (0.102) in the multiple year-adjusted logistic regression (Model B). The situation resembles Simpson's paradox, as discussed in Zhao (2026), where the response is *death* penalty (1=yes vs. 0=no), and the two predictors are *victim* (1-black vs. 0-white) and *defendant* (1-black vs. 0-white). In separate simple logistic regression models, the effect was -1.705 ($p=0.001$) for *victim*, and -0.369 ($p=0.23$) for *defendant*; in the main-effects model, the effect was -2.404 ($p<0.001$) for *victim*, and 0.868 ($p=0.018$) for *defendant*. The interaction term between *victim* and *defendant* in the full model was not significant. That is, after controlling for the victims' race, the probability of receiving the death penalty is significantly higher for black defendants than for white defendants; without such control, the probability is lower for black defendants. The disproportionately low proportion of black defendants (*defendant*=black) among cases with white victims (*victim*=white), a subgroup more prone to receiving the death penalty, is the source of the unbalanced covariates structure.

Overall, due to unbalanced covariate structures in observational data, a truly significant positive effect can be estimated as non-significantly negative (or positive) in a simple regression model; however, this bias can be adjusted for through a multiple linear regression. No additional variable needs to be derived to improve the model.

In contrast to the endogeneity issues, the variable humidity (*humid*) in the bike-sharing data in Zhao (2026) is exogenous. In simple linear regression, the effect of humidity (*humid*) is negative but not significant, aligning with the general intuition that

humidity does not affect the number of bike renters. However, it becomes significantly negative in the most efficient model (a multiple linear regression). As noted in that paper, this phenomenon arises because an important external variable (not in the bike-sharing data), the “forecasted probability of rain (*fcrain*)”, which is proxied by the highly correlated humidity (*humid*), and should be further observed and incorporated to improve the model.

The concepts of exogenous and endogenous variables originate in observational data analysis, not in balanced orthogonal designs (such as $L_{18}(3^6)$). Etymologically, exogenous (from *exo-* ‘outside’ + *-genous* ‘produced’) implies determination by external factors, while endogenous (from *endo-* ‘within’ + *-genous* ‘produced’) implies determination within the system. Specifically, the significant negative effect of humidity (*humid*) in the most efficient model is likely spurious, driven by its strong correlation with a not-included and highly correlated predictor, forecasted probability of rain (*fcrain*), that was not recorded in the current bike-sharing data. In contrast, the significant positive effect of *a5* (or *defendant*) in the most efficient model is intrinsic, whereas the negative and non-significant effect in the simple logistic regression model is attributed to the unbalanced covariate structure. These two terms are frequently confused in the literature, without reference to the etymology of the words.

6.3 Significant, statistically significant, and Bayesian prior

In this paper, we have been emphasizing both important (i.e., practically significant) and meaningful protective and risk factors, while avoiding misconceptions about the term ‘statistically significant’. Without emphasizing prior knowledge of protective or risk factors of the response, the BRPN and FBRPN procedures would degenerate into the backward (or stepwise) method described in standard statistics textbooks. For example,

using the first city's real estate data (Zhao, 2026), the effect of number of bedrooms on house price is 44965.8 ($p < 0.001$) in simple regression, but -8830.0 ($p < 0.001$) when estimated via stepwise regression. Should we describe the latter effect as 'statistically significant'? Is this effect meaningful? If meaningful protective or risk factors are not guaranteed, what is the point of searching for 'important' ones among the factors?

Etymologically, statistics derived from stat-'status'. Strictly speaking, as a statistical term, should not the 'statistically significant' mean that a predictor ranks as an important one while maintaining its 'status' as a meaningful protective or risk factor'? If the status as a protective or risk factor has been checked while building a model, how can the dispute between the Bayesian and Frequentist schools arise?

In recent years, whenever the term 'Bayesian prior' is used, some statisticians equate it exclusively with 'prior distribution' rather than 'prior knowledge'. They seek explicit prior and posterior distributions; without these, grants are often denied and manuscripts rejected. In the case of model selection, the simplest form of Bayesian prior knowledge regarding unknown parameters is a constraint on their signs depending on protective or risk factors of the response, a task that can easily be accomplished using the sequential BRPN and FBRPN procedures. The use of priors on protective or risk factors is highly practical and aligns with the Bayesian framework established by Thomas Bayes (1702 – 1761), the British theologian, statistician, and philosopher, best known for Bayes' theorem.

6.4 Stackability

Comparing the RVs of the modules individually and collectively in Tables 2–5, we find that the RVs of the three combined modules are: (the RVs of the lifestyle module) + (the RVs of the social attitude module—a421–a373) + (the RVs of the socio-

demographic module) + a3012+a5. In other words, when considering the three modules together, except for ideal number of daughters (a373) and men working and women keeping house (a421), all other RVs from the lifestyle, social attitude, and socio-demographic modules, together with a3012, constitute the RVs of the combined modules. The modules are nearly stackable because there is no overlap in the aspects surveyed across them.

To obtain further insights, we calculated the relative changes of the effects in Model B with and without a421 (alternatively a371), defined as $R_{a421} = 100 \times ((\text{Model B} + a421) - \text{Model B}) / (\text{Model B} + a421)$. The results show that all $|R_{a421}| < 5\%$, except for a422 (15.85%) and a423 (6.38%), which means that the effect of a421 is mainly absorbed by that of a422 and a423. In addition, Pearson correlation coefficients between a421 and the predictors in Model B were calculated. Given that the two largest correlations are 0.49 (with a422) and 0.35 (with a423), and that a421, a422, and a423 all belong to the construct of gender equality, a421 (men working and women keeping house) is less representative than a422 (assuming men surpass women in capability) and a423 (prioritizing marriage over career).

Similarly, we find that all $|R_{a373}| < 5\%$, except for a282 (24.94%), a303 (-23.02%), a308 (-34.16%), a32 (-11.95%), a38 (20.78%), a41 (2 vs. 1, -6.19%), a41 (3 vs. 1, -5.69%), a41 (4 vs. 1, -17.15%), a422 (-19.63%), a423 (5.09%), a2 (-6.54%), agr (2 vs. 1, -19.98%), a8a (-12.81%), a9 (15.28%), a10 (-20.87%). That is, the effect of a373 is mainly absorbed by that of magazine usage (a282), shopping (a303), listening to music at home (a308), overnight stays away from home (a32), tolerance of premarital sex (a38), elderly care responsibilities (a41), assuming men surpass women in capability (a422), prioritizing marriage over career (a423), female gender (a2), age groups (agr), total individual income (a8a), application for party membership (a9), and political

affiliation (a10), all of which are gender-related. In addition, Pearson correlation coefficients between a373 and the predictors in Model B were calculated; the highest correlation was 0.71 (with a371). Since both a371 (ideal number of children) and a373 (ideal number of daughters) pertain to the offspring domain, and given that a371 captures a more fundamental aspect of fertility preference, a373 is considered less representative as a standalone predictor. Conceptually, the ideal number of daughters (a373) integrates two components: offspring quantity and gender preference, which were explicitly disentangled in the analysis.

With the rapid development of the Internet, browsing the Internet (a3012) affects people's lifestyles, social attitudes, and socio-demographics, which in turn affects their happiness. Though the impact of browsing the Internet (a3012) on happiness is statistically significant in simple age-adjusted logistic regression, it is not strong enough among numerous competing variables to become a RV in the lifestyle module. However, it becomes one of the RVs of the three modules due to its cumulative impacts on happiness.

Studies with large sample sizes can accommodate more predictors from diverse domains, thereby enabling the selection of more important and meaningful protective or risk predictors. For instance, Model B ($n=64068$), integrating three modules, yielded a reduced deviance of 8437 using 34 parameters. The efficiency surpasses that of the most efficient models for individual modules: the life style module ($n=75461$, reduced deviance of 3003, 12 parameters), the social attitude module ($n=59380$, reduced deviance of 5718, 16 parameters), and the socio-demographic module ($n=69757$, reduced deviance of 1533, 8 parameters). By incorporating all modules from the CGSS phase II data in future analyses, we can identify statistically significant predictors of happiness from a more comprehensive perspective.

The developed model-building strategies are broadly applicable and highly practical, and can be applied in many research fields, such as business, economics, engineering, biology, medicine, education, and the social sciences. Regarding the CGSS data, the modular design provides a good example of how to conduct comprehensive questionnaire surveys, allowing experts from different fields to participate in the design process and to improve the questionnaire based on results of analyses from previous surveys. The modular approach facilitates studying the impacts of several modules individually or collectively on a given outcome variable. Furthermore, Bayesian FBRPN and BRPN procedures can be employed to identify the most efficient model containing important (i.e., practically significant) and meaningful representative variables, regardless of the severity of multicollinearity.

Reference:

- [1] Agresti A. (2007). *An Introduction to Categorical Data Analysis* (2nd). John Wiley & Sons, Inc.
- [2] Chen M.H., Ibrahim J.G., and Yiannoutsos C. (1999). Prior elicitation, variable selection and Bayesian computation for logistic regression models. *Journal of the Royal Statistical Society, Series B*. 61(1), 223-242.
<https://doi.org/10.1111/1467-9868.00173>
- [3] Efron B., Hastie T., Johnstone I., and Tibshirani R.(2004). Least angle regression. *The Annals of Statistics*, 32 (2):407-499.
<https://doi.org/10.1214/0090536040000000067>.
- [4] Li J., Lai D.W.L., and Chappell N.L.(2022). The older, the happier?: a cross-sectional study on the ‘paradox of aging’ in China. *Research on Aging*, 44(1), 34-

43. <https://doi.org/10.1177/0164027521994225>.
- [5] Kang S., Vogt C.A., and Lee S. (2018). Does taking vacations make people happy? A regional disparity perspective. *Asia Pacific Journal of Tourism Research*, 23(11), 1021-1033, <https://doi.org/10.1080/10941665.2018.1515089>
- [6] Mitchell T.J., and Beauchamp J.J. (1988). Bayesian variable selection in Linear regression. *Journal of the American Statistical Association*, 83(404), 1023-1032. <https://doi.org/10.1080/01621459.1988.10478694>
- [7] Shi W., Zhu X., Shi J., Xu L., and Jiang W. (2023). Physical activity, individual emotions, and prosocial attitudes: the role of citizen identity. *Social Behavior and Personality: an international journal*, 51(3), 26-39. <https://doi.org/10.2224/sbp.12224>
- [8] Su Z., Zhou Q., Ye Y., and Li D. (2021). How the media construct happiness under cultural perspective in china: through collectivistic and individualistic values. *Social Science Quarterly*, 102(6), 2619-2639. <https://doi.org/10.1111/ssqu.13073>
- [9] Tokuda Y., Fujii S., and Inoguchi T. (2010). Individual and country-level effects of social trust on happiness: The Asia barometer survey. *Journal of Applied Social Psychology*, 40(10), 2574-2593. <http://doi.org/10.1111/j.1559-1816.2010.00671.x>
- [10] Wei X., Huang S., Stodolska M., and Yu Y. (2015). Leisure time, leisure activities, and happiness in China, *Journal of Leisure Research*, 47(5), 556-576. <http://doi.org/10.18666/jlr-2015-v47-i5-6120>
- [11] World Happiness Report 2017-2023. <https://worldhappiness.report/archive/>
- [12] Xu H., Zhang C., and Huang Y. (2023). Social trust, social capital, and subjective well-being of rural residents: micro-empirical evidence based on the Chinese General Social Survey (CGSS). *Humanities and Social Sciences Communications*, 10(49), 1-13. <https://doi.org/10.1057/s41599-023-01532-1>

- [13] Xu S.Y., and Chen P. (2017). Income, social status, and happiness: fairness perception perspective (in Chinese). *Journal of Management Sciences in China*, 20(12): 99-116. <https://doi.org/10.3969/j.issn.1007-9807.2017.12.008>
- [14] Yan Y., Deng Y., Igartua J.J., and Song X. (2023). Does Internet use promote subjective well-being? Evidence from the different age groups based on CGSS 2017 data. *International Journal of Environmental Research and Public Health*, 20(4), 2897, 1-14. <https://doi.org/10.3390/ijerph20042897>
- [15] Zhao B.B. (2026). Two Bayesian variable screening procedures. *Communications in Statistics - Theory and Methods*, 55(4), 1176-1188.
<https://doi.org/10.1080/03610926.2025.2517276>
- [16] Zheng S.N., Wang S.Q., and Zheng Y.F. (2018). The number of children and the happiness of rural residents: based on gender and age differences (in Chinese). *Northwest Population*, 39(5), 30-38.
<https://doi.org/10.15884/j.cnki.issn.1007-0672.2018.05.004>