Healthy and Non-Healthy Habits, and How The Habits Change in Different Economic Times

Qingya Zhu, Li Xu, Paul D. Berger

Abstract— Insurance companies usually conduct health-background checks on people who seeks insurance plans, especially life insurance and, depending on circumstances, health insurance. Thus, it is interesting for both insurance companies and consumers to know which factors impact human health. Human health is obviously a big issue in many people's life and is significant for the growth of society. What determines the general health of people? This paper focuses on selected factors that potentially have an impact on the general health of people. These include, among others, drinking habits, smoking habits, exercise habits, and the eating of vegetables and fruit. We find that many factors have a significant effect on general health. Increased exercise and drinking turn out to be healthy habits, while smoking and having a personal doctor are negatively associated with health status. We discuss these results, some of which require further elaboration, and one of which is counter-intuitive. In addition, we perform t tests for differences in mean to determine if the level of key factors, such as people's drinking habits, smoking habits, and exercise habits, changed from before the economic crisis in 2008 to after the economic crisis in 2008.

Index Terms— People's Health, Multiple-regression analysis, Ttests for differences in means.

I. INTRODUCTION

One's health status is closely monitored by many people, and examined often by insurance plan personnel. General health status is a complicated combination of numerous factors, such as exercise habits, smoking habits, drinking habits, degree to which one eats vegetables and fruit, other dietary considerations, and many other factors. Our objective in this paper is firstly to see if selected factors that allegedly affect people's health are as expected.

Also, we wish to study whether some potentially key variables, such as smoking habits, drinking habits, and exercise habits change as a function of the economic climate. We considered 2008 as a watershed year, separating two distinct periods of time in terms of economic status. We take 2005 to 2007 as one period of time when the unemployment rate in the United States was relatively low, and the economy of the United States, and most of the world, was considered to be generally good, if not "very good." We consider 2009 to 2011 as a second period of time, a time when there was a high unemployment rate in the United States, and the U.S. economy, and that of most of the world, was considered to be generally "not good," indeed, perhaps "poor."

Qingya Zhu, MBA Scholar, Bentley University , USA
Li Xu, MBA Scholar, Bentley University , USA
Paul D. Berger, Professor of Marketing , Bentley University, USA

As will be seen, in our multiple-linear-regression results, all of the variables we considered have a significant effect on general health at p < .01. Our analysis shows some interesting results which are not consistent with the expectations based on "prevailing-wisdom," including the impact on a person's health of a person's drinking habits and the effect of his/her eating more vegetables and fruits, and having a health plan. We then attempt to explain the reason for these directionally-unexpected results.

II. LITERATURE REVIEW

In "Healthy Living in Hard Times," published in the *Journal of Health Economics* [1], Christopher J. Ruhm shows that smoking and excess-weight each declines during temporary economic downturns. He uses micro-data for adults from the years 1987 to 2000 of the Behavioral Risk Factor Surveillance System, and ran regression analyses. However, in that paper, he does not study how these changes affect health and how the effects will differ in good versus bad economic times.

In "Health, Inequality, and Economic Development," Dr. Angus Deaton [2] also points out that, despite the bad effects of economic downturns on healthcare plans and other things, people are generally healthier in bad-economic times. Dr. Deaton believed that having more time for exercising, changing to a better lifestyle, and having less work pressure are the three main reasons.

The World Health Organization (WHO) has had a series of articles entitled, *Health Impact Assessment,* going back over 25 years. These articles discuss many different factors involved in the health assessment process [3]. Several of these articles discuss how many factors combine together to affect the health of individuals. The three most significant determinants are, in somewhat macro-terms: 1) the social and economic environment; 2) the physical environment; 3) the person's individual characteristics and behaviors. Since the health data we use focus on individuals, our model uses a particular person's individual characteristics and behaviors as independent variables. The WHO website is one of hundreds-of-thousands, if not millions, of articles discussing factors that affect human health status.

III. DATA

The data are collected from the Behavioral Risk Factor Surveillance System (BRFSS) to compare the influences of different factors on health status during selected time periods. Specifically, we considered the three years before the 2008 financial crisis (2005-2007) and the three years after the 2008

financial crisis (2009-2011). BRFSS offers annual surveys covering all the states in the U.S.A., and it includes over 350,000 observations every year. The surveys include more than 300 questions related to an individual's health status, health plans, medical history and living habits. Among the available variables, we choose some key factors that are anticipated to be related to health status.

The dependent variable is used to measure people's level of health. Among the 300 variables in each year's dataset, there are 3 variables which could reflect the level of health: "general health [*during the past year*]," "number of days [*out of the past 30 days*] physical health is not good," and "number of days [*out of the past 30 days*] mental health is not good." *General health* is an estimate of self-health status given by individuals on a 1-5 scale, where 1 means "Excellent" and 5 means "Poor." Compared to the other two variables, "general health" is most suitable. In part, this is because the latter two variables reflect the situation only in the past 30 days, which is less appropriate than the overall health status for a whole year; the other part is the obvious fact that the other two variables are more narrow indicators of a "type of health," rather than overall health. Note that our dependent variable is scaled such that a larger value means a worse health status. This will, of course, be important in interpreting the results of our statistical analyses.

We next considered the choosing of independent variables that are potentially related to the health status of people. It was natural to consider smoking habits, exercise habits, the level of drinking alcoholic beverages, and the health of a person's diet. Correspondingly, we chose the variables: "frequency of days now smoking," "exercise in past 30 days," "drink any alcoholic beverages in past 30 days," "summary index for fruits and vegetables calculated per day.**"** It is reasonable to use these variables even though they were reported for frequency in the past 30 days, because, by and large, they are the a person's habits to a large extent and for the vast majority of people, likely would not change materially during the course of a year. After choosing these "more obvious" variables, we scanned the data codes with the aim of finding other variables whose effect might be interesting to capture in a multiple regression. We decided to add: "have a health-insurance plan," "have a personal doctor," a measure of "exercise frequency," and "level of obesity."

The chosen independent variables are defined as follows; the name in the parenthesis is what we use in the regression model:

- Health-insurance Plan (plan): a dummy variable. The variable = 1 if the individual is included in a health-insurance plan; the variable = 0 if he/she is not. About 87.3% of the respondents reported that they had a health-insurance plan.
- Personal Doctor (doc): a dummy variable. The variable = 1 if the individual has at least one personal doctor; the variable = 0 if he/she does not. About 85.8% of the respondents reported that they had personal doctors.
- Drinking alcoholic beverages (drink): a dummy variable. The variable = 1 if the individual reported that he/she has drunk alcohol at least

once in past 30 days; the variable = 0 if he/she has not. About 54.4% of the respondents reported that they had drunk alcohol in the past 30 days.
- Exercise (exercise): a dummy variable. The variable = 1 if the individual reported that he/she has exercised at least once in past 30 days; the variable = 0 if he/she has not. About 70.5% of the respondents reported that they had exercised in the past 30 days.
- Smoker and Heavy Smoker: both separate dummy variables. "Smoker" =1 if the person reported that he/she smokes some days, but not every day, = 0 otherwise. "Heavy Smoker" = 1 if the person reported that he/she smokes every day. About 9.5% of the respondents reported that they were "Smokers," while about 27.8% of the responders reported that they were "Heavy Smokers."

- Vegetables and fruit Index (v_f_eater and v_f_lover): both separate dummy variables. Vegetables and fruit *eaters* (v_f_eater = 1) are people who reported that they eat vegetables and fruit on average at least once per day, but fewer than three times per day; v_f_eater = 0 otherwise. Vegetables and fruit *lovers* (v_f_lover = 1) are people who reported that they eat vegetables and fruit at least 3 times a day; v_f_lover = 0 otherwise. About 26.3% of the respondents indicated that they were vegetables and fruit *eaters,* while about 56.9% of the respondents indicated that they were vegetables and fruit *lovers*.
- (Overweight): dummy variable. The variable = 1 if the individual is classified as overweight, based on body mass index 25.00 $\leq$ BMI5 < 30.00.
- (Obese): dummy variable. The variable = 1 if the individual is classified as obese, based on body mass index 30.00 $\leq$ BMI5.

IV. MODELS AND DISCUSSION

A. *Ordinary Least Squares (OLS) Model*

To test the influence of the chosen variables on an individual's health status, our basic model is as follows (with "i" indexing an individual person's data values).

$$health_i = \beta_0 + \beta_1 * plan_i + \beta_2 * doc_i + \beta_3 * drink_i + \beta_4 * exercise_i + \beta_5 * smoker_i$$
$$+\beta_6 * heavysmoker_i + \beta_7 * overweight_i + \beta_8 * obese_i + \beta_9 * v\_g\_eater_i + \beta_{10} * v\_g\_lover_i + \square$$

The survey in years 2006 and 2010 did not cover the questions involving variables: v_g_eater and v_g_lover. Thus, we ran a multiple-linear-regression with the above model ("Model 1") using the data from years: 2005, 2007, 2009 and 2011. We also ran a multiple-linear-regression model ("Model 2") without the v_g_eater and v_g_lover variables, using the data from all six years. Model 1 has n = 739,712 observations and Model 2 has n = 1,092,814 observations. The results are shown in Table 1.

Table 1: OLS Multiple-linear-regression output

Dependent variable: Health	Model 1 Years: 2005, 2007, 2009, 2011		Model 2 Years: 2005-2007, 2009-2011	
	Coefficient	Std. Error	Coefficient	Std. Error
Plan	-.126***	.004	-.124***	.003
Doc	.195***	.004	.193***	.003
Drink	-.442***	.002	-.441***	.002
Exercise	-.541***	.003	-.557***	.002
Smoker	.134***	.004	.142***	.004
Heavy Smoker	.191***	.003	.190***	.002
Overweight	.107***	.003	.105***	.002
Obese	.397***	.003	.395***	.003
V_g_eater	.123***	.004		
V_g_lover	.064***	.003		
_cons	3.011***	.006	3.093	.004
Adj. R-squared	0.1484		0.1478	

*** p < .01

We can see from Table 1 that the coefficients for the 8 variables (excluding v_g_eater and v_g_lover) are very similar for both models. The "Smoker" coefficient differs by about 6%, and the Exercise coefficient differs by 3%, while all other coefficients differ by at most 1.5%. We do not view these differences as material. Also, in both models, all the coefficients have $p < .01$. However, we must admit that this "significance" is possibly an artifact of the fact that the sample sizes are so large, and one must acknowledge that there can be a difference between "statistical significance" and "practical significance."

It is important to remember, as we noted earlier, that a *negative* coefficient corresponds with *better health,* given the ("reverse") scale, in which 1 = Excellent health and 5 = Poor health. From the results in Table 1, we can see that the people who indicate that they are "more healthy" are those who have a health-insurance plan, those who have had at least one alcoholic drink within the past 30 days, and those who have exercised within the past 30 days. Those who indicate that they are "less healthy" are those who have a personal doctor, and those who are indicated to be any of the following: a smoker, a heavy smoker, overweight, obese, eater of fruit and vegetables, and lover of fruit and vegetables.

Each of these statements should be interpreted as being accompanied by a caveat of "holding the other variables in the model constant." In other words, when considering that the coefficient of "Plan" (the first variable listed in Table 1) is -.126, we would interpret the result as saying that, "*holding all other variables in the model constant*, those people indicating that they have a health-insurance plan have a level of health that is .126 lower/"better" than those indicating that

they do not have a health-insurance plan," based on the 1 to 5 aforementioned dependent-variable scale. Most of these results agree with conventional wisdom. However, there are some results that might benefit from further elaboration, and one result that appears to be totally counter-intuitive.

It may seem surprising that a beneficial effect (given a negative coefficient) on health is indicated for those who have had at least one alcoholic drink in the past 30 days. While it is somewhat well-known that *heavy* drinking of alcoholic beverages is, in general, harmful to one's health, there is an increasing amount of literature that indicates that "modest alcoholic-beverage drinking" has a positive effect on health. That is, some effects of alcohol consumption are beneficial. Although even moderate alcohol consumption increased the risk of death in younger people, it has been shown to *decrease* the risk of death for individuals ages 55+, due to decreased risk of ischemic heart disease [4]. Therefore, based on the results, it is likely the case that the majority of people in our sample who have drunk "at least one alcoholic beverage in the past 30 days," are people who imbibed alcohol to a very modest degree, and only the minority of them being heavy alcoholic drinkers.

The variable, "having a personal doctor," shows a result of a negative relationship with health status if, indeed, one indicates that he/she has a personal doctor. For explanation of this result, it needs to be noted that, since this study's time period (2005 - 2011), "having one's own doctor" has, to a large extent, changed meaning. Back in 2008 (the midpoint of the data in our study), there were two ways that people likely interpreted that question. One interpretation was to think of themselves as "having a personal doctor" if they simply had a specific primary-care physician; a second interpretation was to think of themselves as "having a personal doctor" only if they had key specialist, perhaps a hematologist or oncologist. The second interpretation is likely why "having a personal doctor" (if affirmative) has a negative relationship with health status. Over the last decade, the phrase has evolved. Nowadays, there is a burgeoning field of "*concierge doctors*," essentially, personal doctors, afforded only by the upper middle-class or rich, who wish to supplement "regular" health insurance, whether private-, employer-, or government-sponsored, and have the ability to afford the "luxury" of a "personal doctor." So, nowadays, the phrase, "having a personal doctor," is more associated with having a *concierge* doctor. While those people who are better off economically have, on average, better health (which would suggest a positive effect on health of having a personal doctor), there is also a natural bias of sorts, so that people who are much healthier than average tend not to want or need a personal/*concierge* doctor, while those with more health issues (and as noted, can afford it) are more likely to engage the service of a personal/*concierge* doctor. Given this natural bias, it might suggest that having a personal doctor correlates negatively with health. We are not certain of the direction this variable would take on, if the study were to be performed using 2018 data.

A non-intuitive result is that for the two variables, V_f_eater and V_f_lover. One would have hypothesized in advance that eating (never mind *loving* [and, hence, eating even more]!!) fruit and vegetables more often would correlate with having *better health*. Yet, both variables are indicated to do the

opposite – that is, be harmful to health - although less so if a V_g_lover, than (just) a V_g_eater. The authors have no explanation for these seeming-counter-intuitive results.

In terms of the magnitude of the coefficients, the, by-far-largest coefficients (in absolute value) are those of Exercise and Drink and Obese. As noted, for Exercise and Obese, the direction is obviously sensible, while, above, we discussed the Drink variable result. The size of the V_g_lover variable, which displays the most counter-intuitive result, has, by far, the smallest coefficient (in absolute value).

B. Logit Model

In performing our linear multiple-regression model above, we assumed that the dependent variable, health status, with its 1-5 scale, was able to be treated, essentially, as an interval-scale variable, even though, technically, such a scale is "officially" ordinal. This assumption for such a scale is routine in the real world when data from such a scale is statistically analyzed. And, there is lots of evidence that little is lost by doing so, and lots is gained in terms of clarity results and given that an interval scale variable is amenable to performing more sophisticated analyses. However, we decided to also perform an analysis that did not require the assumption of an interval scale dependent variable.

We first established that a logit model fit better than a probit model (based on having a smaller AIC value). We performed an ordered logit model. We used the data from 2005, 2007, 2009, and 2011, since, as we noted earlier, the v_f_eater and v_f_lover variables were not available for 2006 and 20010.

While our procedure was somewhat detailed, with the introduction of instrumental variables, and two-step procedure, the ordered logit model results basically duplicated the results of the linear multiple-regression. While some of the signs of the variables were filtered through the signs of the instrumental variables, ultimately, all of the (same) variables were significant (all at p < .01, except for one variable at p = .012), and all of the variables indicated the same directional impact on health status as were shown and discussed in the linear multiple-regression results.

C. A Set of T-tests

We chose three key variables: smoking, drink and exercise, on which to test whether people changed their habits from before the economic crisis of 2008 to after the economic crisis of 2008.

The null hypothesis in each t test is that the (true) mean for the two time periods are equal. The t statistic below is the standard one for testing whether two means are equal or not, with the assumption of equal variances (a robust assumption, so we are confident that the assumption is sufficiently close to true so that we achieve the more powerful test formulation):

$$t - statistic = \frac{\dfrac{Xbar_1 - Xbar_2}{\sqrt{1/n_1 + 1/n_2}}}{\sqrt{\dfrac{(n_1 - 1)s_1{}^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

$Xbar_1$ and $Xbar_2$ in the t-statistic formula represent the sample mean of two samples, respectively, s_1 and s_2 the respective standard deviations, and n_1 and n_2 the respective

sample sizes. Table 2 displays the results.

Again, we need to comment that the extremely low p-values are, in part, associated with the very large sample size of our study. In essence, when you consider a sample size in the

Table 2: Results of the three t-tests

	Xbar1	s_1	Xbar2	s_2	t-statistic	p-value
smoke	0.303	0.460	0.259	0.438	51.052	< .001
drink	0.548	0.499	0.540	0.498	8.21796	< .001
exercise	0.714	0.452	0.698	0.459	17.8816	< .001

neighborhood of a million, one can, more or less, assume that the sample mean is equal to the true mean to several digits. So, we can be extremely confident that the amount of smoking, drinking of alcoholic beverages, and exercising all decreased. The smoking mean decreased by about 15% (intuitively, a non-negligible, "meaningful," amount), suggesting that 15% fewer people had smoked during that past 30 days. The other two means decreased by lesser amounts, about 1.4% for drinking and 2.2% for exercising.

So, the amount of smoking, drinking, and exercising has changed by comparing the sample of 2005 and 2007 with the sample from 2009 and 2011. One possible explanation is that *on average*, people were busy looking for jobs in the "bad economic time" and had less time to exercise, while also having less money to spend on smoking and drinking. However, it was also true that, in the U.S., there was a long-continuing trend of decreased cigarette/cigar/pipe smoking that had started well before 2005 [5], and had decreased from 20.9% of the population in 2005, to 16.8% in 2014 [6].

V. DISCUSSION

We have discussed our primary results earlier. However, there are some "other sides" to some of these findings. An interesting finding we noted was that people tend to have worse health if they have a personal doctor. Yet, it seems clear that the independent variable, having a personal doctor, does not "cause" the dependent variable, health status. It is clearly the other way around, if at all. As we remarked, it is likely that poorer health leads a person to obtain a personal doctor. Also, it is possible that what we call *moral hazard* is a portion of the explanation. People with a personal doctor may tend to be less interested in, and/or less careful of, taking good care of their health, since they know that the personal doctor will be there for them to provide "their money's worth," should ill-health arise.

It is also interesting to go into more detail about the impact of drinking alcoholic beverages. We noted earlier that people who drink alcohol tend to be healthier than those who do not, and that research actually show that moderate drinkers tend to have better health and live longer than non-drinkers – specifically, they have fewer heart attacks and strokes, and they also are generally less likely to suffer diabetes, arthritis, enlarged prostate, and several major cancers. But, what

exactly does "moderate" mean? Medical researchers generally describe moderate drinking as one to three drinks per day. If consuming less than about a half a drink per day, there are only very small health benefits. If consuming more than 5 drinks per day, it is, on average, severely bad for one's health. A standard alcoholic drink is considered a 12-ounce can of regular beer, a 5-ounce glass of dinner wine, or one shot of 80-proof-liquor, such as whiskey or vodka. Harvard's Healthy Eating Pyramid, produced by the Harvard Medical School Guide to Healthy Eating [7] was co-developed by scientists at the Harvard School of Public Health. It is based on the best available scientific knowledge, and is presented in Figure 1. Note that it recommends drinking alcohol in moderation (see arrow in Figure 1).
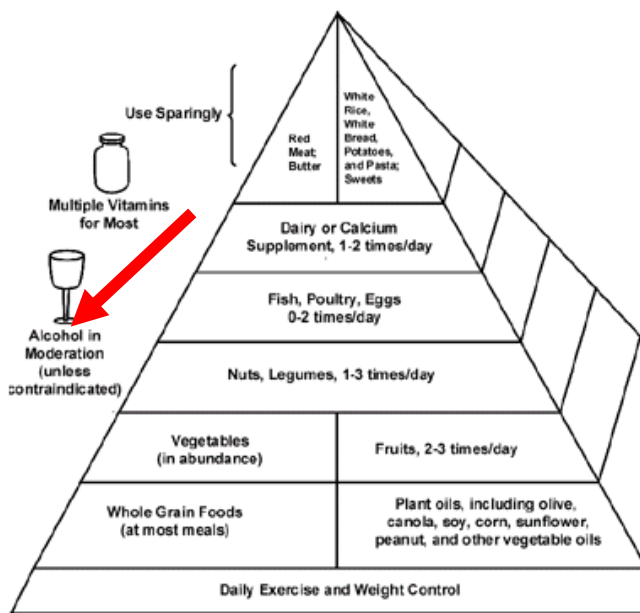

Figure 1: Harvard's healthy eating pyramid

VI. CONCLUSIONS

This paper has analyzed the effects on general health of people of various individual behavioral variables, including drinking, smoking, exercising, eating vegetables and fruit, having a health plan, having a personal doctor, and being overweight/obese. Using a linear multiple-regression analysis, with a very large sample size, we found that a person who reported exercising within the past 30 days, a person reporting drinking an alcoholic beverage with the past 30 days, and a person with a health plan, tended to have better health than those not fulfilling those conditions.

We also found that those who reported that they smoked within the past 30 days, or were overweight or obese, or had eaten fruits or vegetables at least once per day in the past 30 days, or who reported having a personal doctor, tended to have poorer health than those not exhibiting these traits or behavior. We discussed at length the logic of these findings, while agreeing that we were not able to explain the directional impact of the eating of vegetables and fruit.

We also found that three key habits related to health status changed in the years surrounding the economic crisis in the U.S. of 2008 – specifically, between the time periods of (2005 and 2007), and (2009 and 2011). Indeed, the incidence of smoking, alcoholic drinking and exercise decreased

significantly from the earlier two years to the later two years, although we pointed out that the incidence of smoking had been steadily decreasing since 1965 (and has continued to the current day, clearly independent of the 2008 economic crisis.

VII. LIMITATIONS

There are some limitations of our study. One is that the definition of some of the independent variables does not have much granularity. Having a (0, 1) variable as to whether one has exercised in the past 30 days or not leaves present a gigantic amount of variation. Surely, a variable that distinguished the number of times a person has exercised in the past 30 days (say, with a scale of 0, 1-5, 6-10, 1-15, more than 15) would be one (in this case, a 5-point scale) of many ways of obtaining more granularity in the exercise variable. This concept holds true to a lesser degree for the smoking variables (since there are *two* smoking variables!), but, perhaps, *most notably of all*, for the drinking variable.

Another limitation of our study is that we did not include any demographic variables. It is quite likely that certain of the variables we included are likely moderated by certain demographic variables. For example, gender may well moderate the drinking variable, due to different weights of men and women in general.

A third limitation is that we did not include any interaction effects in our analyses. For example, we found that smoking is harmful to health status, and that being overweight or obese is harmful to health status. And, we have an idea of the degree to which each variable affects health status by examining the coefficients of the respective variables in Table 1. However, with the omission of interaction variables, we are unable to determine whether the negative effects of smoking and overweight/obesity are (negatively) additive, or whether the presence of both factors has a larger (or smaller) negative effect on health status than the sum of the individual effects. In general, with such a very large sample size, one can statistically accommodate a large set of interaction variables without any material effect on the reliability of the results.

REFERENCES

[1] C. Ruhm , "Healthy living in bad times," *Journal of Health Economics*, 24, 2004, p. 341-363.

[2] A. Deaton, "Health, Inequality, and Economic Development," *Journal of Economic Literature* , XLI, March,2004, p. 113-158.

[3] [3] WHO website, HIA (Health Impact Assessment), 2018. http://www.who.int/heli/tools/en/

[4] I. White, I, D. Altmann,, and K. Nanchahal), "'Optimal' levels of alcohol consumption for men and women at different ages, and the all-cause mortality attributable to drinking," *Technical Report*, London: London School of Hygiene and Tropical Medicine, 2000.

[5] Centers for Disease Control, "Smokers and Tobacco Use" (2018) https://www.cdc.gov/tobacco/data_statistics/tables/trends/cig_smoking/index.htm

[6] Wikipedia, "Prevalence of tobacco consumption" (2018). https://en.wikipedia.org/wiki/Prevalence_of_tobacco_consumption

[7] Harvard Health Publishing, "Nine for 2009: Nine ways to healthier eating, Mediterranean style," 2009. https://www.health.harvard.edu/newsletter_article/Nine_for_2009_Nine_ways_to_healthier_eating_Mediterranean_style

Corresponding author is **Professor Berger.** All three authors can be reached through Professor Berger.

Qingya Zhu has a Bachelor of Science degree in Mathematics and Economics from Indiana University, a Masters degree in Actuarial Science from Boston University, and is receiving her MBA degree from Bentley University in May, 2018. She has worked at the State Street Corporation, at the Academic Technology Center and Director of PhD Programs as a Graduate Research Assistant at Bentley University, and interned at the National Fire Protection Association as a Data Scientist. She is currently employed part-time as a Marketing Research Analyst at Gangyi Cranes, Inc.

Li Xu has a Bachelor of Arts in Finance degree from Michigan State University, a Master of Arts degree in Economics from Boston University, and is receiving his MBA from Bentley University in May, 2018, with a concentration in Business Analytics. He worked at the China Great Wall Securities Co. as a Wealth Management Analyst, the Department of Economics at Boston University as a Research Analyst, American Tower as an Account Analyst, and at Bentley University as a Graduate Assistant.

Paul D. Berger is a Visiting Scholar and Professor of Marketing at Bentley University. He has his Bachelor of Science, Master of Science, and PhD. from the Sloan School of Management, at the Massachusetts Institute of Technology. He is the Director of the Master of Science in Marketing Analytics (MSMA) program at Bentley University. He has authored 7 textbooks totaling 12 editions, as well as over 200 articles published in academic journals and conference proceedings. He can be reached at pberger@bentley.edu, or Bentley University, 175 Forest St., Waltham, MA, 02452, U.S.A.