Psychology of Addictive Behaviors
2007, Vol. 21, No. 4, 441–452

0893-164X/07/$12.00 DOI: 10.1037/0893-164X.21.4.441

Inference in Regression Models of Heavily Skewed Alcohol Use Data: A Comparison of Ordinary Least Squares, Generalized Linear Models, and Bootstrap Resampling

Dan J. Neal
Kent State University

Jeffrey S. Simons
University of South Dakota

Analysis of alcohol use data and other low base rate risk behaviors using ordinary least squares regression models can be problematic. This article presents 2 alternative statistical approaches, generalized linear models and bootstrapping, that may be more appropriate for such data. First, the basic theory behind the approaches is presented. Then, using a data set of alcohol use behaviors and consequences, results based on these approaches are contrasted with the results from ordinary least squares regression. The less traditional approaches consistently demonstrated better fit with model assumptions, as demonstrated by graphical analysis of residuals, and identified more significant variables potentially resulting in theoretically different interpretations of the models of alcohol use. In conclusion, these models show significant promise for furthering the understanding of alcohol-related behaviors.

Keywords: generalized linear models, bootstrapping, ordinary least squares, base rate behaviors, alcohol use data

Over the last decade, the importance of bringing relatively modern statistical approaches to the psychological sciences has been stressed (e.g., Grissom, 2000; Wilcox, 1995, 1998). Despite this, the majority of published research on addictive behaviors continues to report analyses based on ordinary least squares (OLS), such as analysis of variance (ANOVA) and multiple regression. Such an approach can be problematic when used in analyses of data on alcohol consumption and related low base rate behaviors. In particular, models based on OLS require the assumptions of normally distributed errors with constant variance. These assumptions are questionable because (a) data on alcohol use and related risk behaviors typically show significant heteroskedasticity, that is, errors are more widely distributed at higher levels of the behaviors and (b) the lower limit of zero inherent in these typically skewed data generally precludes the possibility of normally distributed residuals.

The use of OLS regression when model assumptions are violated can be problematic, particularly with respect to inference regarding the estimated parameters. Nonnormally distributed data and heterogeneous variances can lead to test statistics that are not t or F distributed, and therefore p values based on these test statistics may be biased (Grissom, 2000; Hamilton, 1992). One common solution when dealing with violations of model assumptions is to transform the data to improve the normality of residuals (e.g., square-root or natural log transformations). Although these approaches do tend to improve the distribution of residuals of the model, interpretation of analyses based on transformed variables can be difficult. Perhaps more importantly, it is not uncommon to be unable to find a transformation that improves the normality of residuals. In many cases, there is no transformation that is appropriate for the data.

The development of theoretical models of alcohol use (or other risk behaviors) is, to a large extent, guided by inference tests of estimated parameters. It is thus important to identify alternative statistical methods that may best approximate the distributions observed in alcohol, drug, and other risk behavior research to facilitate scientific progress in the addictions and related fields. Failing that, the use of statistical methods that free the researcher from the limitation of normal-theory assumptions may provide an alternative context for statistical inference. To address these issues, one may choose to implement analyses based on generalized linear models (GzLM) or bootstrapping.

This article highlights GzLMs and bootstrapping and compares and contrasts results from traditional analyses with results from these alternative analyses. Differences in the substantive conclusions resulting from the analyses are highlighted. Of note, although we focus on alcohol use data as the primary example, the identified problems and their resolution is a function of the underlying distributions and are thus relevant to a wider range of research topics.

Alcohol Use Data

Alcohol use data can be classified into two categories: continuous data and count data. *Continuous data* can take on any real value and, in terms of alcohol use data, usually have a lower limit of zero. Examples of these data include blood alcohol concentra-

Dan J. Neal, Department of Psychology, Kent State University; Jeffrey S. Simons, Department of Psychology, University of South Dakota.

This research was supported in part by National Institute on Alcohol Abuse and Alcoholism Grants R01-AA014573 to Jeffrey Simons and 5T32-AA07471 to Rueben Gonzalez. We would like to thank Catherine Caska for her assistance in collecting and entering the data.

Correspondence concerning this article should be addressed to Dan J. Neal, Department of Psychology, P.O. Box 5190, Kent State University, Kent, OH 44242-0001. E-mail: dneal2@kent.edu

tion (BAC) or a composite measure of drinks per drinking day (sum of drinks per week divided by the number of drinking occasions) generated from the Daily Drinking Questionnaire (Collins, Parks, & Marlatt, 1985). *Count data* represent the number of times that a specific behavior occurred and are further restricted to integer values only. Number of discrete drinking episodes and number of binge episodes (e.g., Wechsler, Dowdall, Davenport, & Rimm, 1995) are two examples of count data. Furthermore, data that represent the summation of items on a questionnaire may have distributions that approximate count distributions. An alcohol-related consequences measure, such as the Rutgers Alcohol Problem Index (White & Labouvie, 1989), which assesses relatively infrequent consequences and typically results in positively skewed integer data, is one example.

Both count data and continuous data often follow well-defined probability distributions other than the normal. Exploration of the GzLM (McCullagh & Nelder, 1989) provides some insight into these other distributions.

GzLMs

The GzLM (McCullagh & Nelder, 1989) is a broad class of linear models in which the distribution of errors can be assumed to follow a variety of probability functions beyond the normal distribution.[1] The GzLM allows for specification of error distributions that include but are not limited to the normal, binomial, Poisson, negative binomial, and gamma distributions. For example, both OLS regression and logistic regression (e.g., Hosmer & Lemeshow, 2000) are special cases of the GzLM based on the normal distribution and the binomial distribution, respectively.

Particularly relevant to alcohol use data are three distributions: the Poisson, the negative binomial, and the gamma. The Poisson and the negative binomial distributions are employed to model count data, as they require non-negative integer values. However, the non-negative integer nature of the data may also make these distributions useful for heavily skewed questionnaire data scored as totals. Although such data are not truly distributed as Poisson or negative binomial, inference in Poisson and negative binomial regression is still valid given that the conditional mean and variance functions are consistent with the model specification (Cameron & Trivedi, 1998).

The Poisson distribution is unusual in that it is completely defined by a single parameter, lambda (λ), that defines the mean and the variance. Examples of the Poisson distribution, with $\lambda = 1$, $\lambda = 2$, and $\lambda = 4$, are shown in the top row of Figure 1. With lower mean values, the Poisson is positively skewed; but, at higher mean values, the Poisson converges to the normal distribution. The single parameter assumption, however, is often overly restrictive, especially with data specific to alcohol consumption. Thus, the negative binomial distribution (a generalization of the Poisson) is often more practical in practice. The negative binomial distribution is defined by two parameters, lambda (λ, the mean) and alpha (α, the dispersion), and requires that the mean is proportional to the variance. As λ increases, the variance also increases and is constrained to be equal to $\lambda + \alpha * \lambda^2$. Examples of the negative binomial distribution, with $\lambda = 1$, $\lambda = 2$, and $\lambda = 4$, and $\alpha = 1$ (yielding variances = 2, 6, and 20) are shown in the middle row of Figure 1. Visual inspection of these distributions confirms that they may approximate observed alcohol consumption data well; both the nature of the data (non-negative integers) and the shape of the distribution (positive skew with most data at or near zero) are consistent with many distributions of alcohol use.

Just as the Poisson and negative binomial distributions are relevant to count data, the gamma distribution is relevant to continuous data that demonstrate significant skew and are constrained to positive values. Like the Poisson, the gamma distribution is represented by a single parameter, gamma (γ), which is the mean; the variance is equal to γ^2. The gamma distribution, with parameters $\gamma = 1$, $\gamma = 2$, and $\gamma = 4$, can be seen in the bottom row of Figure 1; as with the negative binomial distribution for count data, the gamma distribution is consistent with many observed distributions of alcohol data measured on a continuous metric.

Complete reviews of the GzLM can be found in several books, including McCullagh and Nelder (1989) and Hardin and Hilbe (2001). Gardner, Mulvey, and Shaw (1995) reviewed the theory and use of Poisson and negative binomial regression, presenting the fundamental mathematical theory, estimation and inference procedures, and interpretation of results. The use of GzLM based on gamma distributions has not been reviewed in a psychological context.

Statistical Models

Although OLS regression and GzLM regression share many similarities, it is the differences between the two models that make the GzLM more appropriate for alcohol use data. The basic OLS regression model posits that an individual score (y_i) is a function of a set of p explanatory variables (x_{1i}, x_{2i}, . . . x_{pi}) and an unobserved error, which is normally distributed with mean 0 and variance σ^2. A set of p regression coefficients (β_1, β_2, . . . β_p) is multiplied with the set of explanatory variables to yield a predicted estimate of the individual score, $\hat{Y}_i$. That is,

$$\hat{Y}_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ \ldots \ + \beta_p X_{pi}. \quad (1)$$

A mathematically identical interpretation of $\hat{Y}_i$ is as a conditional mean (i.e., the predicted mean conditional on the predictor variables); the regression equation yields the conditional mean of scores that are normally distributed with variance equal to σ^2.

The basic model for the GzLM is

$$\eta_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ \ldots \ + \beta_p X_{pi}, \quad (2)$$

[1] The generalized linear model should not be confused with the general linear model (GLM). The GLM requires the assumption of univariate (e.g., ANOVA, multiple regression) or multivariate (*e.g.*, multivariate analysis of variance, discriminant analysis) normality. The generalized linear model (variously abbreviated as GzLM, GLIM, and GLZ) provides for the use of multiple error distributions. Confusion regarding the GLM and the GzLM may result in part from the abbreviation "GLM." In statistical programs, such as SPSS and SAS, GLM refers to the general linear model; but, in other programs, such as Stata, GLM refers to the generalized linear model. Both SAS and Stata have implemented full algorithms for estimation of generalized linear models. In Stata, the algorithm is -glm- and was introduced in 1992, whereas in SAS the algorithm is PROC GENMOD and was introduced in 1994. To date, there is no algorithm for estimation of generalized linear models in SPSS; however, SPSS has announced that generalized linear models will be available in the upcoming release of SPSS 15.

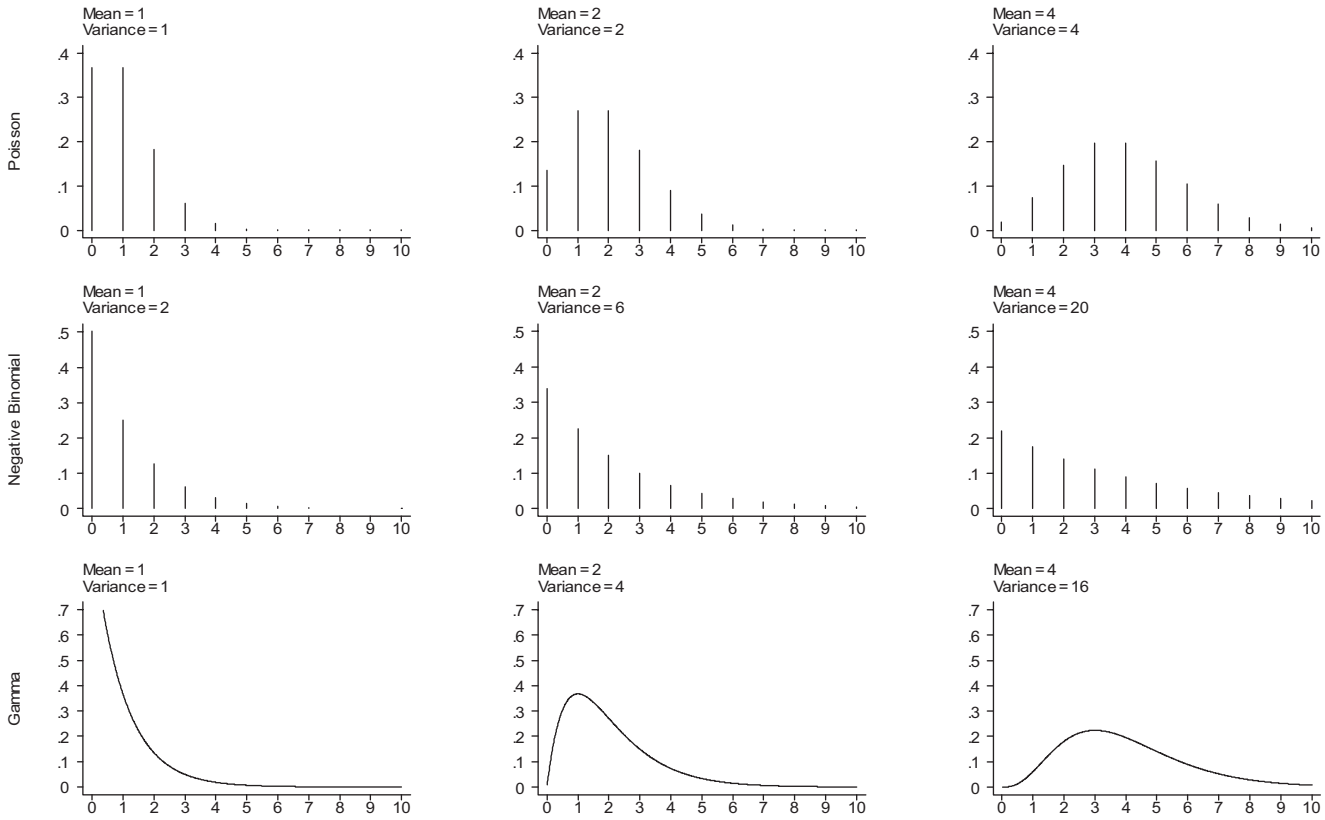

Figure 1. Examples of distributions. The top row is the Poisson distribution with $\lambda = 1$, $\lambda = 2$, and $\lambda = 4$. The middle row is the negative binomial distribution, with $\lambda = 1$, $\lambda = 2$, and $\lambda = 4$, and $\alpha = 1$ (yielding variances = 2, 6, and 20). The bottom row is the gamma distribution with $\gamma = 1$, $\gamma = 2$, and $\gamma = 4$.

where η_i is the linear combination of the p regression coefficients (β_1, β_2, . . . β_p) and the p explanatory variables (x_{1i}, x_{2i}, . . . x_{pi}). Thus far, this is identical to OLS regression; however, the differences between the GzLM and OLS regression stem from the fact that the conditional mean μ_i is not necessarily equal to η_i. Instead, the conditional mean must be recovered through the use of a link function (McCullagh & Nelder, 1989). For both Poisson and negative binomial regression, the link function typically used is the natural log function, whereas for the gamma distribution it is the reciprocal function, although the natural log is also frequently employed. Therefore, for the log link the conditional means can be modeled as

$$\ln(\mu_i) = \eta_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ \ldots \ + \beta_p X_{pi}. \qquad (3)$$

And μ_i can be recovered by exponentiating the linear combination

$$\mu_i = e^{\eta i} = e^{\beta_1 X_1 + \beta_2 X2i + \ \ldots \ + \beta_p X_{pi}.} \qquad (4)$$

One of the benefits of modeling the natural log of the conditional means, as opposed to modeling the conditional means directly, is that a predicted conditional mean can never be less than zero. Thus, the log-linear models can avoid predicting nonsensical values; that is, the statistical model would never predict that an individual would consume fewer than zero drinks.[2]

Once the conditional mean μ_i is estimated, the other major difference between OLS regression and GzLM becomes apparent. In OLS regression, the residuals are assumed to have a normal distribution of $\mu = 0$ and variance $= \sigma^2$ for all values of $\hat{Y}$. This is not the case in GzLM models. For the Poisson, negative binomial, and gamma distributions, the variances are all functions of the mean, and thus the error variances vary with the conditional means (i.e., predicted values). Because the variance of the distributions is either equal to or proportional to the mean, the error distributions are inherently heteroskedastic. For each conditional mean μ_i, the assumed conditional distribution would have a mean of μ_i and variance μ_i for Poisson regression, mean of μ_i and variance of $\mu_i + \alpha^* \mu_i^2$ for the negative binomial, and mean μ_i and variance μ_i^2 for the gamma. Therefore, in both cases, larger values of μ_i correspond to larger values of both the conditional mean and the conditional variance. Thus, the heteroskedasticity that is typi-

[2] Of note, the use of the GzLM with the log link is not the same as using OLS regression on a log transformed variable. With the GzLM, the conditional mean μ_i can be directly recovered via the inverse link function, but this is not true with OLS regression on a log-transformed variable. The primary difference is that the GzLM models the log of the predicted mean of Y, whereas OLS on a transformed variable models the predicted mean of the log of Y, which includes transformation of the error term.

cally observed with alcohol use data is incorporated directly into the model.

Additionally, a high degree of skew is expected with these data, especially for small values of μ_i. That is, as μ_i approaches zero, the conditional distribution must be positively skewed as values less than zero are not possible for count or non-negative continuous distributions. This is also consistent with many observed distributions of alcohol use data, where heavier mean levels of consumption may yield distributions with less skew, but relatively lighter mean levels of consumption yield heavily skewed distributions with a majority of observations falling at or near zero. In community samples, most participants will report few, if any, heavy drinking days or drinking-related problems and only a small proportion will report frequent heavy consumption and related problems. This results in a heavily skewed distribution. Dependent upon type of sample, frequency of greater perceived risk behaviors (alcohol problems, illicit drug use, binge drinking, drinking and driving) will often occur at a lower base rate and demonstrate greater degree of positive skew than more common behaviors of less perceived risk (drinking frequency, frequency of driving). It is the commonly observed sample characteristics of high-risk addictive behaviors that make the GzLM of importance to the field.

Analysis of Model Residuals

In OLS regression there are many tools that can be used to assess the assumption of homoskedastic normally distributed errors, but two approaches are particularly useful. To assess normality, one can examine normal-quantile plots of the response residuals (i.e., $Y_i - \hat{Y}_i$). Likewise, to assess potential heteroskedasticity, one can examine a residual-versus-fitted plot, wherein the residuals are plotted against the predicted value ($\hat{Y}_i$). Use of these regression diagnostic tools can inform one as to how tenable the normality assumption might be.

Such approaches fail when assessing residuals for the GzLM because of the heteroskedasticity inherent in the model. Therefore, one must make an adjustment of the residuals to ensure that the residuals are comparable across all observations. There are three common adjustments of residuals used in the GzLM: the Pearson residual, the deviance residual, and the Anscombe residual (McCullagh & Nelder, 1989). Each type of residual attempts to scale and normalize the response residuals so that they can be analyzed using techniques similar to those above. The Pearson residual is generally considered the least appropriate, as it does a poor job of normalizing residuals, especially when the conditional mean μ_i is small (McCullagh & Nelder, 1989). Instead, it is generally recommended to use the deviance residuals or the Anscombe residuals. Once these adjusted residuals have been computed, the same procedures used in OLS can be applied to GzLM models. The adequacy of the distributional assumption (whether Poisson, negative binomial, or gamma) can then be assessed in the observed alcohol data. If the data are distributed as assumed, then the residuals should appear to be approximately normal and homoskedastic.

Significance Testing

OLS regression is extremely flexible because a wide variety of inference testing can be conducted. Whether used to test for differences between groups (ANOVA/analysis of covariance) or for the individual or joint significance of a set of continuous predictor variables, all significance tests reduce to *F* tests between full and reduced models. However, the use of non-normal distributions precludes the possibility of applying normal-theory inferential statistics and estimation procedures. Therefore, these models must be fit via maximum likelihood estimation, and alternative inference tests appropriate for maximum likelihood estimation must be used. Choices for these tests include likelihood ratio tests and Wald Tests (Cameron & Trivedi, 1998). Likelihood ratio tests are derived from the difference in likelihood estimates between two models, resulting in a χ^2 test that is used to test the significance of one or more regression coefficients. Wald tests are derived from the estimated variance-covariance matrices of the regression coefficients and result in a χ^2 or *z* test (the *z* test is equivalent to a χ^2 test with 1 degree of freedom). Whether using likelihood ratio or Wald *z* and χ^2 tests, it should be recognized that these tests are analogous to the *F* and *t* tests from OLS regression and their interpretation is quite similar. Thus, most hypothesis tests regarding alcohol use behavior that could be tested in OLS can also be tested when using the GzLM.

Measures of Goodness of Fit

One aspect of OLS regression models that scientists rely heavily on is the coefficient of determination, that is, R^2. R^2 is a measure of variance in the dependent variable that can be explained by the regression model and is generally interpreted as the predictive power of the model. Thus, the degree to which a predictor or set of predictors accounts for variation in alcohol use can be quantified. There is, however, no direct analogous measure of variance accounted for in GzLMs. Several alternative indices similar to R^2 have been proposed, although none have gained widespread acceptance among statisticians (Hardin & Hilbe, 2001). Suggested measures include those based on the log-likelihood of the model, such as McFadden's R^2 and Cragg and Uhler's R^2 (cf. Hardin & Hilbe, 2001), as well as those that examine the correlation between the predicted values and observed values, similar to R^2 in OLS regression (Zheng & Agresti, 2000). Results from these various "pseudo R^2" measures do not necessarily provide similar results (Hardin & Hilbe, 2001). For example, the Cragg-Uhler R^2 is normed to range from 0–1, whereas McFadden's cannot equal 1 (Long & Freese, 2003). Thus, one should use them only to compare similar models (e.g., comparisons of nested models). Additionally, these measures in general cannot be used to quantify variation accounted for in alcohol use (or any other dependent measure). Finally, there have been recommendations to use indices of model fit, such as Akaike Information Criterion or the Bayesian Information Criterion, as model summary statistics. Given that these indices have less familiar interpretations associated with them, however, they are not illustrated in this article.

Bootstrapping

The GzLM relies on distributions other than the normal distribution. Similar to OLS regression, however, if a researcher assumes that alcohol use data are distributed Poisson, negative binomial, or gamma, but this assumption is not tenable, then inferential tests based on the model may be invalid. Although the

use of the Poisson, negative binomial, or gamma distribution increases the chance that the assumed distribution is reasonable, there is no guarantee that such a result is a given. For example, scores may be truncated at some higher value, or the degree to which the data conform to specific distribution may be less than ideal. Examination of marginal distributions (e.g., via histograms or spike plots) may provide some information in this regard. Therefore, it may also be useful to find other ways of implementing statistical tests and constructing confidence intervals that do not rely on potentially problematic distributional assumptions. One such approach is bootstrapping, a computer-intensive nonparametric approach for conducting hypothesis testing (Efron & Tibshirani, 1993).

The theory behind bootstrapping is deceptively simple and quite elegant. In general, the approach attempts to empirically estimate the sampling distribution of any given statistic. Because samples are random, a sample statistic will not necessarily equal the parameter value in the population. A single random sample of the population may yield a mean level of alcohol consumption of 3.2 drinks per day, and a second random sample from the same population may yield a slightly different estimate of mean alcohol consumption of 2.9 drinks per day. Repetition of this sampling scheme 1,000 times would yield 1,000 estimates of mean alcohol consumption. Plotting these estimates in a histogram would yield the sampling distribution, that is, how much the estimate of average consumption varies due to random differences from sample to sample.

In all parametric approaches (including both OLS and GzLM regression) an assumption regarding the statistical distribution of observed scores is made. Given that the distributional assumption is true, an analytical formula exists to estimate the sampling distribution. If one assumed that alcohol use was normally distributed in the population, then the mean and standard deviation of the sampling distribution can be calculated, yielding the standard error, confidence interval, and inference test for the mean. The standard error is only accurate, however, if the assumption regarding distribution of population scores is true. When this assumption is false, so too are the standard error, confidence interval, and reported p value of the inference test.

Bootstrapping can free the researcher from the need for distributional assumptions by simulating the process of repeatedly sampling from the population. An assumption is made that the single observed sample is representative of the population from which it was taken. If this assumption is true, then taking a random sample from the sample is approximately equivalent to taking a random sample from the population. Bootstrapping derives from this logic. A random sample of size n is taken, with replacement, from the original sample, yielding a single bootstrap sample. The statistic of interest is estimated in the bootstrap sample, yielding a bootstrap statistic. A second bootstrap sample is taken, and a second bootstrap statistic is estimated. This procedure is repeated anywhere from 50 to 5,000 times, yielding a distribution of bootstrap statistics. The bootstrap distribution is used to empirically estimate standard errors, confidence intervals, and so forth around the observed statistics (e.g., regression coefficients) from the original model. Thus, bootstrapping can be used to supplement existing analyses (e.g., OLS regression or GzLM regression) to provide more accurate confidence intervals and inference statistics.

Two important points must be addressed. First, bootstrapping is a powerful technique, but one that must be considered with its limitations in mind. In particular, the assumption that the original sample is entirely representative of the population is essential. If there are particular subpopulations in the population that are not represented in the sample (e.g., lighter drinkers were somehow less likely to be sampled initially) then a simple bootstrap procedure may fall short. Second, bootstrapping must be set up to approximate the actual sampling procedure as closely as possible. For simple random samples, this is not a difficult task; but, for more complex sampling techniques, such as stratified random sampling, this can be more difficult.

Current Study

The purpose of the current study was to demonstrate the use of GzLMs and bootstrapping in comparison to OLS regression on a set of real-world (as opposed to simulated) alcohol use data and to examine patterns of differential results based on the analysis conducted. The current dataset has three outcome variables of interest: the estimated BAC for a typical drinking occasion reported by each participant (continuous data), the number of binge episodes reported by each participant (count data), and scores on a measure of alcohol-related consequences (questionnaire data that are distributed similar to count data). For each outcome variable, four analyses were conducted: (a) OLS regression analysis, (b) OLS regression analysis with bootstrapped standard errors, (c) GzLM regression analysis, and (d) GzLM regression analysis with bootstrapped standard errors. For binge drinking and alcohol-related consequences, the GzLM was implemented using a negative binomial reference distribution with a log link function. For typical BAC, the GzLM was implemented using a gamma reference distribution with log link function.[3]

Method

The data presented are based on demographic, alcohol use, and individual differences variables. Participants were 206 (63.6% women) college students who participated in exchange for credit in an introductory psychology course. All participants provided informed consent to participate, and Institutional Review Board approval was granted prior to beginning data collection. Participants reported their gender, weight, and alcohol use patterns. Participants reported the amount of alcohol they typically consume and the amount of time they typically spend drinking, allowing for estimation of *typical BAC* using the Matthews and Miller (1979) equation. Participants also reported the number of times they consumed five or more drinks for men or four or more drinks for women in the past month to measure the number of *binge episodes.* The Rutgers Alcohol Problem Index (RAPI; White & Labouvie, 1989) was included as a measure of alcohol-related consequences.

Participants also completed three individual differences measures. The Impaired Control Scale assesses an individual's inten-

[3] Although the reciprocal link is the "canonical" link that is most commonly used with the gamma distribution, the log link is also frequently employed and can be used in this context. Such an approach is attractive because it provides more consistency between the gamma models and the negative binomial models estimated for the other two dependent variables.

Table 1
Descriptive Data of Sample by Gender

	Women (n = 131)		Men (n = 75)		Overall (N = 206)	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*
Alcohol use						
Typical BAC	0.08	0.07	0.07	0.06	0.07	0.06
Binge episodes	3.5	3.9	4.8	4.2	4.0	4.0
RAPI	6.5	6.9	8.9	7.2	7.4	7.1
Individual differences						
SSRQ	118.0	13.8	119.6	13.7	118.6	13.7
ICS	9.0	5.9	8.2	5.0	8.7	5.6
N-RDQ	2.6	2.1	2.6	2.3	2.6	2.2
P-RDQ	2.2	1.1	2.6	1.2	2.4	1.1

Note. BAC = blood alcohol concentration; binge = four drinks for women and five drinks for men; RAPI = Rutgers Alcohol Problem Index; SSRQ = Self-Regulation Questionnaire Short Form; ICS = Impaired Control Scale; N-RDQ = Reasons for Drinking Questionnaire–Negative Reinforcement; P-RDQ = Reasons for Drinking Questionnaire–Positive Reinforcement.

tion to limit alcohol consumption in certain situations. The scale consists of 10 items and has shown good psychometric properties, including an internal consistency of .88 (Heather, Booth, & Luce, 1998). The internal consistency of this scale in this sample was .83. The Self-Regulation Questionnaire Short Form (SSRQ; Carey, Neal, & Collins, 2004) is a 31-item measure based on the Self-Regulation Questionnaire (Brown, Miller, & Lawendowski, 1999) designed to assess capacity for self-regulation in substance abusers. The SSRQ has good psychometric properties, including an internal consistency of .92, and correlates with the longer Self-Regulation Questionnaire at .96 (Carey et al., 2004); the full version has a 2-day test–retest reliability of .91 (Brown et al., 1999). The internal consistency of the SSRQ in this sample was .91. The Reasons for Drinking Questionnaire (RDQ; Farber, Khavari, & Douglass, 1980) is a 14-item questionnaire that assesses motivations for drinking. The RDQ has two subscales, negative reinforcement motives (escaping unpleasurable stimuli) and positive reinforcement motives (gaining pleasurable stimuli). Psychometric analyses indicate a strong two-factor solution (Farber et al., 1980) and adequate internal consistency (Cronbach's alphas of .84 and .65; Carey & Correia, 1997). In this sample, the internal consistency of the negative reinforcement factor was .71 and the internal consistency of the positive reinforcement factor was .54.

Summary statistics by gender for the four individual differences measures and the two alcohol use variables are presented in Table 1. Histograms of the two alcohol use variables and the alcohol-related consequences measure, which exhibit significant skew, are presented in Figure 2.

Results

Analysis Strategy

Three sets of analyses were conducted in order to examine the correlates of "typical BAC," "binge drinking episodes," and "RAPI." For typical BAC and binge episodes, five predictor variables were used: gender (0 = female, 1 = male), impaired control (continuous variable, range = 0–26), self-regulation capacity (continuous variable, range = 72–156), negative reinforcement drinking motives (continuous variable, range = 0–9), and positive reinforcement motives (continuous variable, range = 0–5). For the RAPI, seven predictor variables were used: gender, impaired control, self-regulation, negative reinforcement motives, positive reinforcement motives, typical BAC, and binge episodes. Exploratory analyses on the predictor variables did not identify any significant problems with outliers or leverage in the regression models.

For typical BAC, four analyses were conducted: In Model 1, a simple OLS regression model was computed. In Model 1a, the same regression model was computed, but standard errors were estimated using a bootstrap procedure with 5,000 bootstrap replications. In Model 2, a GzLM with a gamma reference distribution and log link was computed. In Model 2a, the same GzLM model was computed, but standard errors were estimated using a bootstrap procedure with 5,000 bootstrap replications.

For binge episodes and RAPI scores, four analyses were computed for each dependent variable. Models 1 and 1a were identical to the models for typical BAC. In Model 2, a GzLM model with a negative binomial reference distribution and log link was computed. In Model 2a, the same GzLM model was computed, but standard errors were estimated using a bootstrap procedure with 5,000 bootstrap replications.

All analyses were conducted in Stata 7.0 (Stata Corporation, 2001). OLS regression was conducted using the -regress- command. GzLM regression with the gamma reference distribution were computed using the -glm, family(gamma) link(log)- command, and GzLM regressions with the negative binomial reference distribution were computed using the -nbreg- command.[4] For comparative purposes, measures of predictive power of the models included R^2 for OLS regression and McFadden's R^2, the Cragg-Uhler R^2, and the squared correlation between predicted and observed scores (Zheng and Agresti's R^2; Zheng & Agresti, 2000) for GzLMs. Bootstrapping was implemented using the -bs- command.

[4] Although negative binomial regression models could have been computed using the -glm, family(nb) link(log)- command, estimation using -nbreg- is preferable as it provides an estimate of alpha, whereas -glm- requires alpha to be specified by the user.

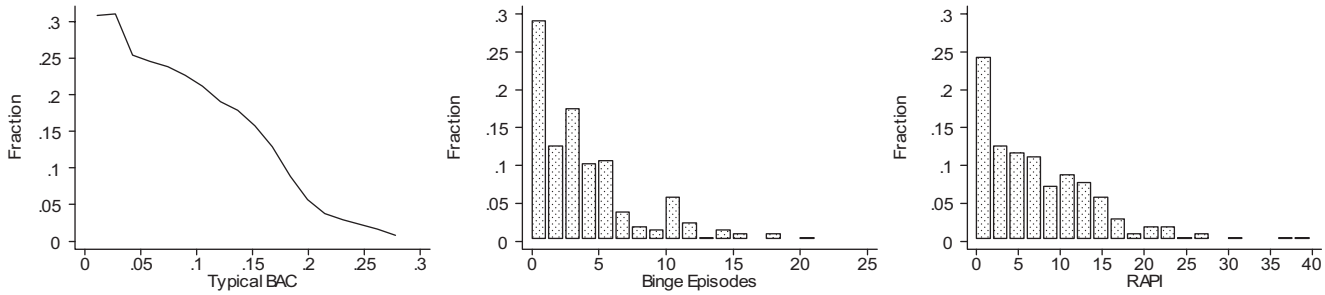

Figure 2. Histograms of the two alcohol use variables and the alcohol-related consequences measure, which exhibit significant skew. BAC = blood alcohol concentration; RAPI = Rutgers Alcohol Problem Index.

Typical BAC

Full results are presented in Table 2. Overall, Model 1 (OLS regression) was significant, $F(5, 196) = 9.08$, $p < .0001$, $R^2 = .19$. Impaired control over alcohol and negative reinforcement motives were significant and positively associated with typical BAC. Additionally, self-regulation was marginally significant. Similarly, analysis of Model 1a (OLS regression with bootstrapped standard errors) yielded three significant predictors, including impaired control over alcohol, self-regulation, and negative reinforcement motives, which were all positively associated with BAC.

Overall, Model 2 (GzLM with gamma distribution) was significant, $\chi^2(5, N = 206) = 37.94$, $p < .0001$, McFadden's $R^2 = .03$, Cragg-Uhler $R^2 = .16$, Zheng and Agresti's $R^2 = .14$. However, in contrast to Models 1 and 1a, Model 2 yielded four significant predictor variables. Specifically, impaired control over alcohol, self-regulation, negative reinforcement motives, and positive reinforcement motives were all positively associated with BAC. Analysis of Model 2a (GzLM model with bootstrapped standard errors) yielded similar results. Impaired control over alcohol, self-regulation, negative reinforcement motives, and positive reinforcement motives were all positive significant predictors of BAC.

Examination of the residual normal-quantile plots (see Figure 3, top row), as well as the residual-versus-fitted plots (see Figure 4, top row) comparing Models 1 and 2 indicated an interesting trend. The deviance residuals in the GzLM model appear to be more normally distributed than the residuals in the OLS model; as such, the GzLM model may more accurately represent the distribution of typical BAC. However, both models appear to demonstrate heteroskedasticity: For OLS, wider variation in residuals appears at higher levels of predicted typical BAC; and, for the GzLM, the wider variation appears at lower levels of predicted values. In particular, the GzLM seems to lack positive residuals at higher predicted values, that is, the model is overestimating BAC at this level.

Thus, use of the gamma distribution identified four significant predictors compared to only two in the OLS regression. Furthermore, the results from the bootstrapped standard errors were consistent within the GzLM regression but not within the OLS regression where self-regulation had p values of .054 versus .024. Although more significant predictors are not necessarily better, the consistency across models with and without bootstrapped errors, as well as visual inspection of the residual plots, suggests that the gamma model is performing better.

Binge Drinking

Results are presented in Table 2. Overall, Model 1 (OLS regression) was significant, $F(5, 196) = 7.83$, $p < .0001$, $R^2 = .17$. Three of the five explanatory variables were significant. Men reported more frequent binge episodes compared to women, and impaired control over alcohol and negative reinforcement motives were positively associated with binge frequency. Analysis of Model 1a (OLS regression with bootstrapped standard errors) yielded similar results, with gender, impaired control over alcohol, and negative reinforcement motives all significant.

Overall, Model 2 (GzLM with negative binomial distribution) was significant, $\chi^2(5, N = 206) = 43.04$, $p < .0001$, McFadden's $R^2 = .04$, Cragg-Uhler $R^2 = .19$, Zheng and Agresti's $R^2 = .12$. However, in contrast to Models 1 and 1a, Model 2 yielded four significant predictor variables. Men reported more frequent binge episodes than women; and impaired control over alcohol, negative reinforcement motives, and positive reinforcement motives were all positively associated with binge frequency. Analysis of Model 2a (GzLM model with bootstrapped standard errors) yielded similar results. Gender, impaired control over alcohol, and negative reinforcement motives were all significant. Positive reinforcement motives, however, were only marginally significant.

Examination of the residual normal-quantile plots (Figure 3, middle row), as well as the residual-versus-fitted plots (Figure 4, middle row) comparing Models 1 and 2, indicates that the deviance residuals in the GzLM model appear to be more normally distributed than the residuals in the OLS model. Although neither plot appears ideal, the data support the idea that the GzLM model may more accurately represent the distribution of binge drinking. Additionally, both models appear to demonstrate heteroskedasticity, although the data for OLS appear to suffer worse from this problem whereas the GzLM again appears to have a dearth of positive residuals at higher predicted values.

Thus, use of the negative binomial distribution identified four significant predictors, compared to three in the OLS regression. However, the results from the bootstrapped standard errors were not entirely consistent within the GzLM regression; positive reinforcement motives demonstrated $p = .046$ with the GzLM and $p = .052$ with the bootstrapped standard errors. Although using a strict alpha rule for significance would lead to different interpretations, there is relatively little practical difference between these two p values, unlike the p values associated with the regression coefficient for self-regulation in the OLS model for typical BAC above.

Table 2
Results from Ordinary Least Squares (OLS) and Generalized Linear Models (GzLM) Regression Models With and Without Bootstrapped (BS) Standard Errors

	OLS regression							GzLM–gamma						
		Without bootstrapping			With bootstrapping				Without bootstrapping			With bootstrapping		
Variable	*b*	*SE*	*t*	*p*	BS *SE*	BS *z*	*p*	*b*	*SE*	*z*	*p*	BS *SE*	BS *z*	*p*
BAC														
Gender	−0.0073	0.0087	0.84	.401	0.0083	0.88	.379	−0.10	0.14	−0.70	.485	0.15	−0.66	.509
ICS	0.0033	0.00094	3.47	.001	0.00097	3.35	.0008	0.049	0.015	3.18	.001	0.015	3.21	.001
SSRQ	0.0007	0.00036	1.94	.054	0.00031	2.26	.024	0.013	0.0061	2.05	.040	0.0063	2.01	.044
N-RDQ	0.0073	0.0023	3.15	.002	0.0021	3.45	.0006	0.13	0.041	3.04	.002	0.037	3.43	.0006
P-RDQ	0.0056	0.0041	1.35	.178	0.0034	1.62	.105	0.20	0.083	2.42	.016	0.093	2.15	.032

	OLS Regression							GzLM–Negative binomial						
		Without bootstrapping			With bootstrapping				Without bootstrapping			With bootstrapping		
Binge	*b*	*SE*	*t*	*p*	BS *SE*	BS *z*	*p*	*b*	*SE*	*z*	*p*	BS *SE*	BS *z*	*p*
Gender	1.29	0.56	2.33	.021	0.59	2.18	.029	0.37	0.14	2.68	.007	0.15	2.54	.011
ICS	0.13	0.060	2.23	.027	0.049	2.72	.007	0.041	0.016	2.61	.009	0.014	2.84	.005
SSRQ	0.037	0.023	1.59	.113	0.022	1.71	.087	0.011	0.0058	1.83	.066	0.0064	1.66	.097
N-RDQ	0.51	0.15	3.47	.001	0.15	3.38	.0007	0.14	0.039	3.73	.0002	0.038	3.75	.0002
P-RDQ	0.28	0.26	1.08	.280	0.25	1.15	.250	0.14	0.071	2.00	.046	0.073	1.94	.052

	OLS Regression							GzLM–Negative binomial						
		Without bootstrapping			With bootstrapping				Without bootstrapping			With bootstrapping		
RAPI	*b*	*SE*	*t*	*p*	BS *SE*	BS *z*	*p*	*b*	*SE*	*z*	*p*	BS *SE*	BS *z*	*p*
Gender	2.52	0.83	3.04	.003	0.84	2.98	.003	0.43	0.12	3.72	.0002	0.12	3.59	.0003
BAC	24.41	6.88	3.55	.0005	7.29	3.35	.0008	4.43	0.99	4.47	<.0001	1.09	4.06	<.0001
Binge	0.27	0.11	2.51	.013	0.13	2.13	.033	0.044	0.015	2.97	.003	0.014	3.08	.002
ICS	0.21	0.091	2.35	.020	0.082	2.59	.010	0.045	0.013	3.41	.001	0.012	3.86	.0001
SSRQ	−0.11	0.034	3.15	.002	0.035	3.09	.002	−0.018	0.0051	3.47	.001	0.0057	3.12	.002
N-RDQ	0.56	0.23	2.48	.014	0.25	2.28	.022	0.066	0.031	2.11	.035	0.030	2.33	.020
P-RDQ	0.059	0.37	0.15	.879	0.32	0.18	.857	0.13	0.060	2.23	.026	0.059	2.27	.023

Note. BAC = blood alcohol concentration; ICS = Impaired Control Scale; SSRQ = Self-Regulation Questionnaire Short Form; N–RDQ = Reasons for Drinking Questionnaire–Negative Reinforcement; P–RDQ = Reasons for Drinking Questionnaire–Positive Reinforcement; binge = four drinks for women and five drinks for men; RAPI = Rutgers Alcohol Problem Index.

Furthermore, as the residuals in the GzLM appear to show better model fit, the GzLM does appear to be performing better than OLS in this instance. The use of the bootstrapped standard errors would be recommended, however.

Rutgers Alcohol Problem Index

Results are presented in Table 2. Overall, Model 1 (OLS regression) was significant, $F(7, 194) = 20.66$, $p < .0001$, $R^2 = .43$. Six of the seven explanatory variables were significant. Men reported higher scores on the RAPI, and typical BAC, binge episodes, impaired control over alcohol, and negative reinforcement motives were positively associated with RAPI scores; self-regulation was negatively associated with RAPI scores. Analysis of Model 1a (OLS regression with bootstrapped standard errors) yielded similar results; gender, typical BAC, binge episodes, impaired control over alcohol, self-regulation, and negative reinforcement motives were all significant.

Overall, Model 2 (GzLM with negative binomial distribution) was significant, $\chi^2(7, N = 206) = 129.15$, $p < .0001$, McFadden's $R^2 = .10$, Cragg-Uhler $R^2 = .47$, Zheng and Agresti's $R^2 = .35$. However, in contrast to Models 1 and 1a, Model 2 yielded seven significant explanatory variables. Men reported higher RAPI scores, and typical BAC, binge episodes, impaired control over alcohol, negative reinforcement motives, and positive reinforcement motives were all positively associated with RAPI scores; self-regulation was negatively associated with RAPI scores. Analysis of Model 2a (GzLM regression with bootstrapped standard errors) yielded similar results; gender, typical BAC, binge episodes, impaired control over alcohol, self-regulation, negative reinforcement motives, and positive reinforcement motives were all significant.

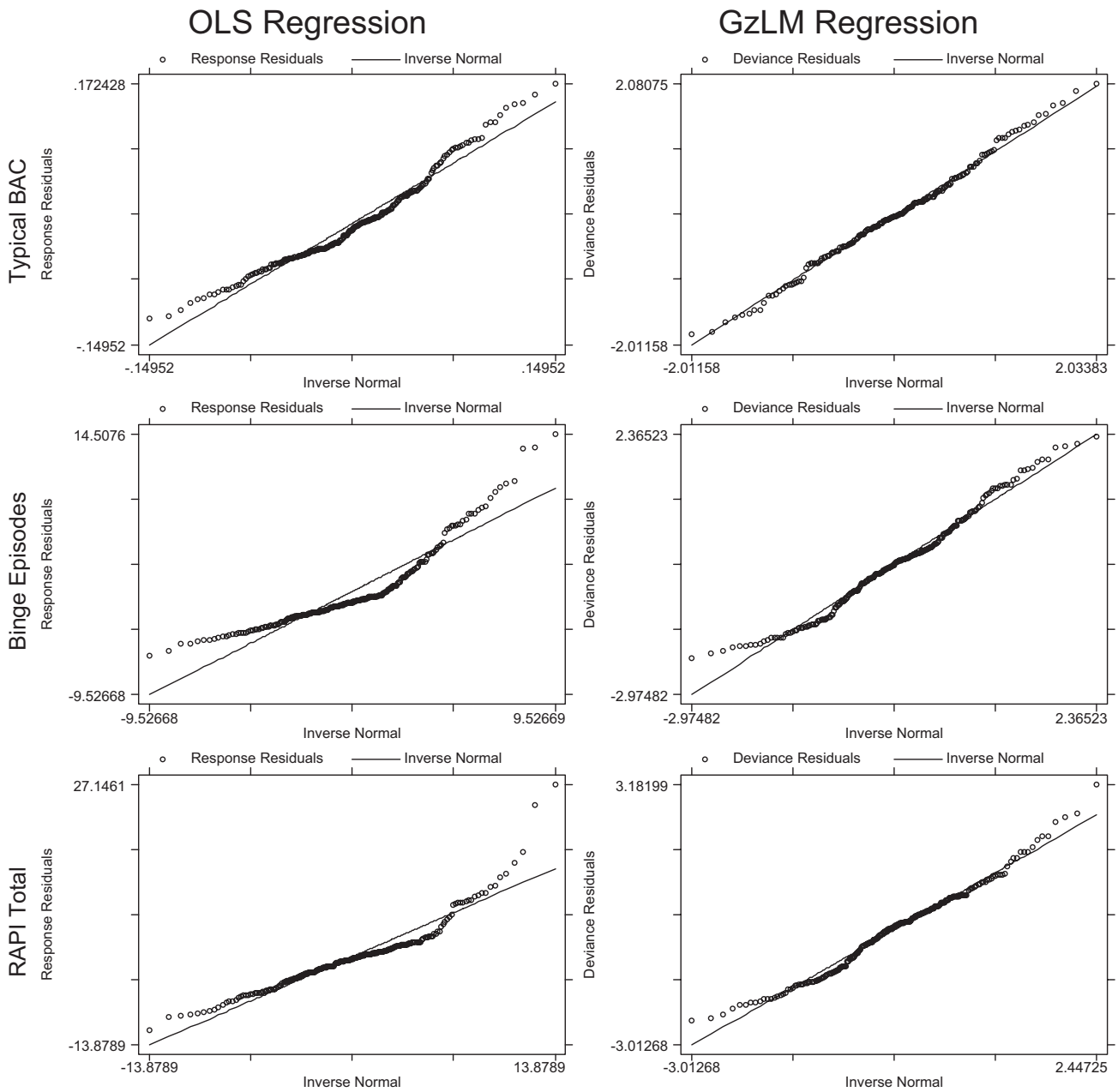

Figure 3. Normal-quantile plots examining normality of the residuals from Model 1 (ordinary least squares [OLS] analyses) in the first column and deviance residuals from Model 2 (generalized linear models [GzLM] analyses) in the second column. Top row = typical blood alcohol concentration (BAC); middle row = binge drinking; bottom row = Rutgers Alcohol Problem Index (RAPI) analyses.

Examination of the residual normal-quantile plots (Figure 3, bottom row), as well as the residual-versus-fitted plots (Figure 4, bottom row) comparing Models 1 and 2 indicate that the deviance residuals in the GzLM model appear to be more normally distributed than the residuals in the OLS model, although again the GzLM model does not appear to be ideal. Thus, the data support the idea that the GzLM model may more accurately represent the distribution of RAPI scores. Examination of the residual-versus-fitted plots also indicates a significant degree of heteroskedasticity in both models; and, for this variable, the degree of heteroskedasticity does not appear to differ between the two methods. The OLS model does predict negative RAPI scores, which can be problematic from an interpretation standpoint; the GzLM model does not have this problem.

In sum, use of the negative binomial distribution identified seven significant predictors, compared to six in the OLS regression. Although both models demonstrated some degree of heteroskedasticity, the GzLM residuals were approximately normally distributed. Combined with the fact that

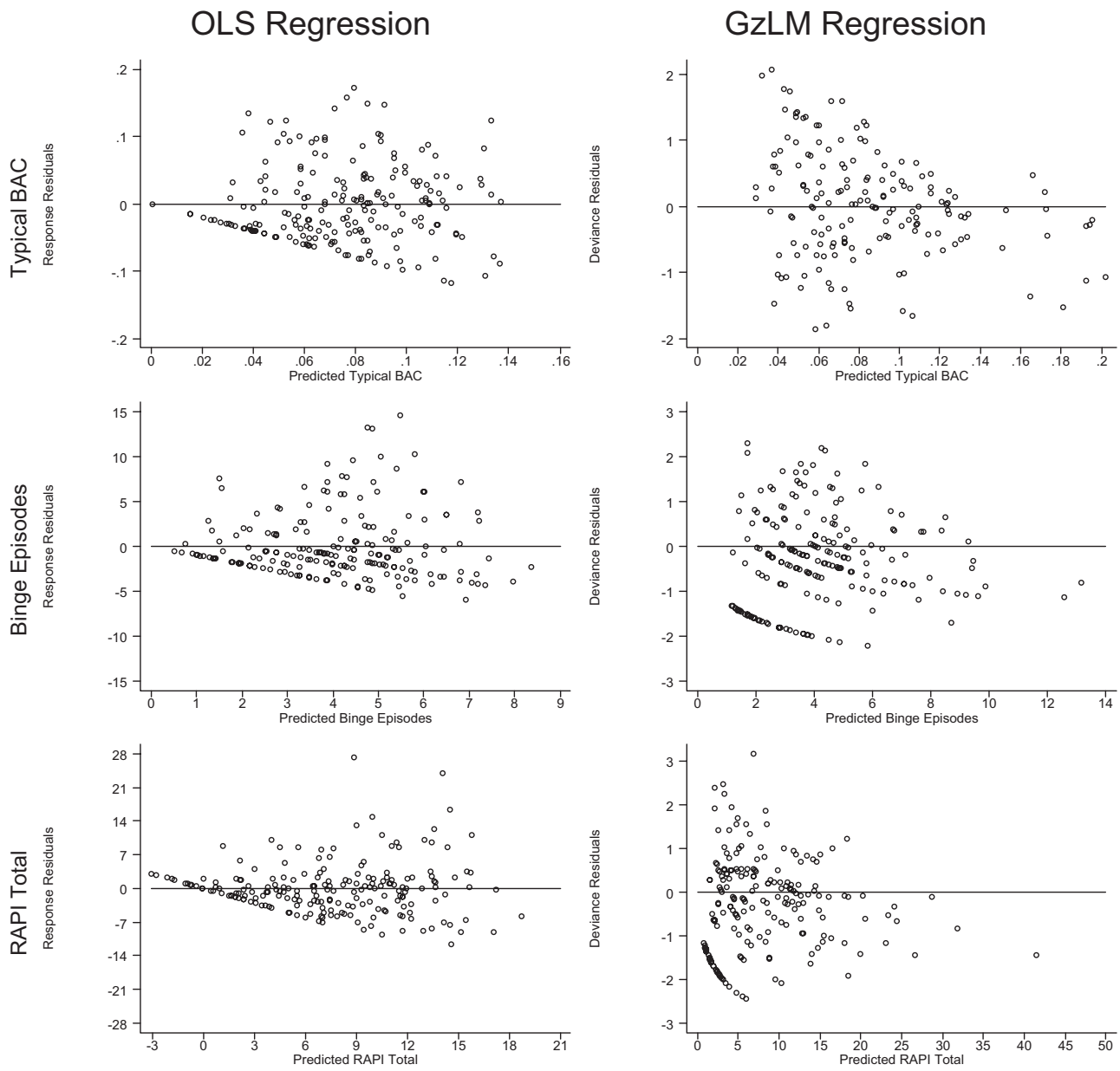

Figure 4. Residual-versus-fitted plots examining homoskedasticity of the residuals from the ordinary least squares (OLS) analyses (first column) and deviance residuals from the generalized linear models (GzLM) analyses (second column). Top row = typical blood alcohol concentration (BAC); middle row = binge drinking; bottom row = Rutgers Alcohol Problem Index (RAPI) analyses.

this model had an extra significant predictor variable and was consistent with the bootstrapped standard errors, it may be functioning slightly better than the OLS model. As with the binge drinking variable, however, it is likely that the GzLM model should be supplemented with bootstrapped standard errors to ensure consistency in the interpretation of the model.

Discussion

The purpose of the current study was to demonstrate the use of GzLMs and bootstrapping in comparison to OLS regression on a set of data related to alcohol use. Analyses were conducted on three dependent variables, typical BAC, binge episodes, and RAPI scores; and, for each variable, analyses were conducted using OLS regression, OLS regression analysis with bootstrapping, GzLM regression, and GzLM regression analysis with bootstrapping.

It was demonstrated that the GzLM, although mathematically more complex than OLS regression, can be implemented and interpreted in a manner quite similar to the latter approach. Procedures for hypothesis testing, assessment of model assumptions, and summarizing predictive power are available and easily acces-

sible. Likewise, bootstrapping is a simple and elegant approach to empirically estimating standard errors for OLS and GzLM regression models in the presence of violation of model assumptions. Researchers who are familiar with OLS should not have difficulty progressing to the GzLM or implementing bootstrap procedures.

Comparison of OLS Regression and GzLMs

In most cases, results based on GzLM and OLS regression were consistent with each other, demonstrating that OLS regression is a fairly robust technique even when distributional assumptions are not met. Importantly, however, there were substantive differences between the results of OLS regression and GzLM regression on all three of the alcohol use indices. The GzLM regressions yielded more significant predictors compared to the OLS regressions. Specifically, positive reinforcement motives (e.g., drinking to celebrate, drinking to be sociable) were significant predictors of typical BAC, binge drinking, and alcohol-related problems when analyzed via GzLM but were nonsignificant in the regression models based on OLS. Such differences are not the result of slight changes in reported p values such that one test barely rejects the null hypothesis (e.g., $p = .04$) and the other test barely fails to reject the null hypothesis (e.g., $p = .06$). Rather, the differences appear to be substantial, with p values for the OLS models ranging from .178 to .879 and p values for the GzLM models ranging from .016 to .046. The differences in p values for self-regulation in predicting typical BAC were less meaningful; however, those researchers adhering to a traditional approach to hypothesis testing would have drawn different conclusions between the GzLM and the OLS regression models.

For the purpose of identifying relevant predictors of drinking behavior and consequences, the traditional approach of OLS would have provided potentially misleading results. The importance of such a result cannot be overlooked; the misidentification of predictors of drinking can have a negative effect on the development of scientific models of alcohol use behavior. Because these analyses were conducted on a set of real-world (i.e., not simulated) data, there is no gold standard to determine which model provided the correct results. Examination of the deviance results (GzLM) and response residuals (OLS) provides some insight, however. In general, the GzLM approach provided a better fit to the observed distributions. The deviance residuals from the GzLM all appeared more normally distributed than the response residuals from OLS regression, and heteroskedasticity appeared less problematic for the typical BAC and binge drinking in the GzLM compared to OLS regression. Given that the distributional assumptions of the GzLM analyses were more tenable, it is likely that the real, albeit small, effect of positive reinforcement drinking motives was accurately detected in the GzLM. In contrast, the significance of this small effect was obscured by residual variance due to the misspecification of the OLS model. Correct specification of the error variance in the GzLM analyses allowed the effect to emerge.

Determining which model fit better or explained more variation in the three indices of alcohol use is impossible. The R^2 values from OLS regression were quite comparable to the Cragg-Uhler R^2. Additionally, both R^2 and the Cragg-Uhler R^2 were somewhat higher than the Zheng and Agresti R^2 values and considerably higher than the McFadden R^2 values. Comparing the R^2 estimates and the Cragg-Uhler R^2 estimates, R^2 was slightly higher for analysis of typical BAC, and the Cragg-Uhler R^2 estimates were slightly higher for binge episodes and alcohol-related consequences. It is not justified to conclude, however, that the higher R^2 implies that OLS is a better analytic strategy for BAC and GzLM is a better analytic strategy for binge episodes and alcohol-related consequences. The various pseudo R^2 indices do not have equivalent meanings and are not directly comparable (Hardin & Hilbe, 2001). What these results primarily demonstrate is that the selection of R^2 measures can have a dramatic impact on the substantive interpretation of the model. In particular, use of these indices as immutable measures of predictive power irrespective of the context of the model and the specific index is problematic. That being said, they may be quite useful for examining the relative strength of nested or otherwise comparable models.

Comparison of Parametric Inference and Bootstrap Inference

Comparison of the results based on the parametric standard errors (i.e., the standard errors derived from OLS estimates and GzLM estimates) to results based on the bootstrapped standard errors yields a consistent pattern. In general, the parametric standard errors from each model did not differ substantially from bootstrapped standard errors for the same model. The two notable exceptions were the OLS model of typical BAC and the GzLM for binge drinking; in the former case, the bootstrap method identified an additional significant variable, and in the latter case the bootstrap method identified one less significant variable. Of note, the relative difference in p values for SSRQ in the BAC model (.054 for OLS, .024 for bootstrapping) was quite substantial, but the relative difference in p values for positive reinforcement motives in the binge drinking model were approximately equal (.046 for GzLM, .052 for bootstrapping) and just happened to fall on opposite sides of the (somewhat arbitrary) alpha criterion of .05.

Although the statistical model that is more appropriate should, in theory, have bootstrapped standard errors that more closely align with the parametric standard errors, this was not observed in the current data. Though the results seem to suggest that the combined use of bootstrapping with other regression models (OLS or GzLM) is not valuable, the use of bootstrap procedures with alcohol data may still provide significant benefit. Boostrapped standard errors can evaluate the relative reliability of the parametric standard errors; that is, to examine whether the parametric and nonparametric approaches provide similar estimates. This may be especially beneficial when the results of inference testing may be difficult to interpret due to significant violations of model assumptions.

Furthermore, it is important to recognize that bootstrapping is not limited to supplementing existing models with empirically based standard errors. One can use bootstrapping to estimate standard errors and confidence intervals for any statistic, and it may be particularly useful when alternative probability distributions are unavailable. For example, Miller et al. (2002) used bootstrapping to compare the strength of test–retest reliabilities of several alcohol use indices in the presence of significant deviations from bivariate normality. Bootstrapping may also be applied when analytic formulas for standard errors are unavailable. For example, Neal and Carey (2005) used bootstrapping to estimate the standard errors and 95% confidence intervals for eigenvalues in a factor

analysis, and Roberts, Neal, Kivlahan, Baer, and Marlatt (2000) used bootstrapping to determine 95% confidence intervals of a cut score between a normative drinking group and a heavy drinking group of college students.

Extension of GzLM

The analyses presented here are relatively simple models on a set of cross-sectional data. Just as ANOVA and OLS regression have longitudinal (repeated-measures) extensions, so does the GzLM. Researchers wishing to use mixed or multilevel models may find the generalized linear mixed model (e.g., Diggle, Liang, & Zeger, 1994) useful. The generalized linear mixed model combines the multiple reference distributions available in the GzLM with the addition of both fixed and random effects as estimated in mixed linear models. Thus, the generalized linear mixed model can be used to estimate models that are analogous to the random intercept, random slopes, and random coefficients models. Likewise, longitudinal models can also be estimated using generalized estimating equations (e.g., Hardin & Hilbe, 2003). Although the generalized estimating equations model does not allow for estimation of random effects, these are often more easily estimated with complex data.

In sum, although the underlying mathematical structure of the GzLM model is more complex than OLS (due to maximum likelihood estimation and the implementation of a link function), researchers with familiarity with OLS regression should easily be able to implement and interpret results based on GzLM. That the GzLM models allow for analysis of highly skewed continuous data, such as BAC, as well as highly skewed count and pseudocount data, such as binge episodes and RAPI scores, underscores the usefulness of these approaches. These alternative analysis approaches have a significant amount to offer applied researchers.

References

Brown, J. M., Miller, W. R., & Lawendowski, L. A. (1999). The self-regulation questionnaire. In L. VandeCreek & T. L. Jackson (Eds.), *Innovations in clinical practice: A source book* (pp. 281–292). Sarasota, FL: Professional Resource Press/Professional Resource Exchange, Inc.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data.* Cambridge, UK: Cambridge University Press.

Carey, K. B., & Correia, C. J. (1997). Drinking motives predict alcohol-related problems in college students. *Journal of Studies on Alcohol, 58,* 100–105.

Carey, K. B., Neal, D. J., & Collins, S. E. (2004). A psychometric analysis of the self-regulation questionnaire. *Addictive Behaviors, 29,* 253–260.

Collins, R. L., Parks, G. A., & Marlatt, G. A. (1985). Social determinants of alcohol consumption: The effects of social interaction and model status on the self-administration of alcohol. *Journal of Consulting and Clinical Psychology, 53,* 189–200.

Diggle, P. J., Liang, K. Y., & Zeger, S. L. (1994). *Analysis of longitudinal data.* Oxford, UK: Oxford University Press.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Farber, P. D., Khavari, K. A., & Douglass, F. M. (1980). A factor analytic study of reasons for drinking: Empirical validation of positive and negative reinforcement dimensions. *Journal of Consulting and Clinical Psychology, 48,* 780–781.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin, 118,* 392–404.

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology, 68,* 155–165.

Hamilton, L. C. (1992). *Regression with graphics.* Pacific Grove, CA: Brooks/Cole.

Hardin, J., & Hilbe, J. (2001). *Generalized estimating equations.* Boca Raton, FL: Chapman & Hall.

Hardin, J., & Hilbe, J. (2003). *Generalized linear models and extensions.* College Station, TX: Stata Press.

Heather, N., Booth, P., & Luce, A. (1998). Impaired Control Scale: Cross-validation and relationship with treatment outcome. *Addiction, 93,* 761–771.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression.* New York: Wiley.

Long, J. S., & Freese, J. (2003). *Regression models for categorical dependent variables using Stata* (2nd ed.). College Station, TX: Stata Press.

Matthews, D. B., & Miller, W. R. (1979). Estimating blood alcohol concentration: Two computer programs and their applications in therapy and research. *Addictive Behaviors, 4,* 55–60.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models.* New York: Chapman & Hall.

Miller, E. T., Neal, D. J., Roberts, L. J., Baer, J. S., Cressler, S. O., Metrik, J., & Marlatt, G. A. (2002). Test-retest reliability of alcohol measures: Is there a difference between Internet-based assessment and traditional methods? *Psychology of Addictive Behaviors, 16,* 56–63.

Neal, D. J., & Carey, K. B. (2005). A follow-up psychometric analysis of the self-regulation questionnaire. *Psychology of Addictive Behaviors, 19,* 414–422.

Roberts, L. J., Neal, D. J., Kivlahan, D. R., Baer, J. S., & Marlatt, G. A. (2000). Individual drinking changes following a brief intervention among college students: Clinical significance in an indicated preventive context. *Journal of Consulting and Clinical Psychology, 68,* 500–505.

Stata Corporation. (2001). Stata 7.0 [Computer software]. College Station, TX: Author.

Wechsler, H., Dowdall, G. W., Davenport, A., & Rimm, E. B. (1995). A gender-specific measure of binge drinking among college students. *American Journal of Public Health, 85,* 982–985.

White, H. R., & Labouvie, E. W. (1989). Towards the assessment of adolescent problem drinking. *Journal of Studies on Alcohol, 50,* 30–37.

Wilcox, R. R. (1995). ANOVA: A paradigm for low power and misleading measure of effect size? *Review of Educational Research, 65,* 51–77.

Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist, 53,* 300–314.

Zheng, B., & Agresti, A. (2000). Summarizing the predictive power of a generalized linear model. *Statistics in Medicine, 19,* 1771–1781.

Received April 25, 2006
Revision received December 19, 2006
Accepted December 21, 2006 ■