

Bayesian adjustment for measurement error in an offset variable in a Poisson regression model

Kangjie Zhang¹, Juxin Liu², Yang Liu³, Peng Zhang⁴ and Raymond J. Carroll^{5,6}

¹Everest Clinical Research, Markham, Ontario, Canada

²Department of Mathematics and Statistics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

³Upstart Network Inc., San Carlos, California, USA

⁴Department of Statistics, Zhejiang University, Hangzhou, China

⁵Department of Statistics, Texas A&M University, College Station, Texas, USA

⁶School of Mathematical and Physical Sciences, University of Technology, Sydney, Broadway, NSW, Australia

Abstract: Fatal car crashes are the leading cause of death among teenagers in the USA. The Graduated Driver Licensing (GDL) programme is one effective policy for reducing the number of teen fatal car crashes. Our study focuses on the number of fatal car crashes in Michigan during 1990–2004 excluding 1997, when the GDL started. We use Poisson regression with spatially dependent random effects to model the county level teen car crash counts. We develop a measurement error model to account for the fact that the total teenage population in the county level is used as a proxy for the teenage driver population. To the best of our knowledge, there is no existing literature that considers adjustment for measurement error in an offset variable. Furthermore, limited work has addressed the measurement errors in the context of spatial data. In our modelling, a Berkson measurement error model with spatial random effects is applied to adjust for the error-prone offset variable in a Bayesian paradigm. The Bayesian Markov chain Monte Carlo (MCMC) sampling is implemented in rstan. To assess the consequence of adjusting for measurement error, we compared two models with and without adjustment for measurement error. We found the effect of a time indicator becomes less significant with the measurement-error adjustment. It leads to our conclusion that the reduced number of teen drivers can help explain, to some extent, the effectiveness of GDL.

Key words: Conditional autoregressive (CAR) models, Graduated Driver Licensing (GDL), measurement error, offset variable, spatial random effects

Received 18 October 2020; revised 2 March 2021; accepted 18 March 2021

1 Introduction

1.1 Background

The main cause of death among teenagers in the USA is from motor vehicle accidents (Chen et al., 2014). In many developed countries, teen drivers have the highest crash

Address for correspondence: Juxin Liu, Department of Mathematics and Statistics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 5E6.

E-mail: liu@math.usask.ca



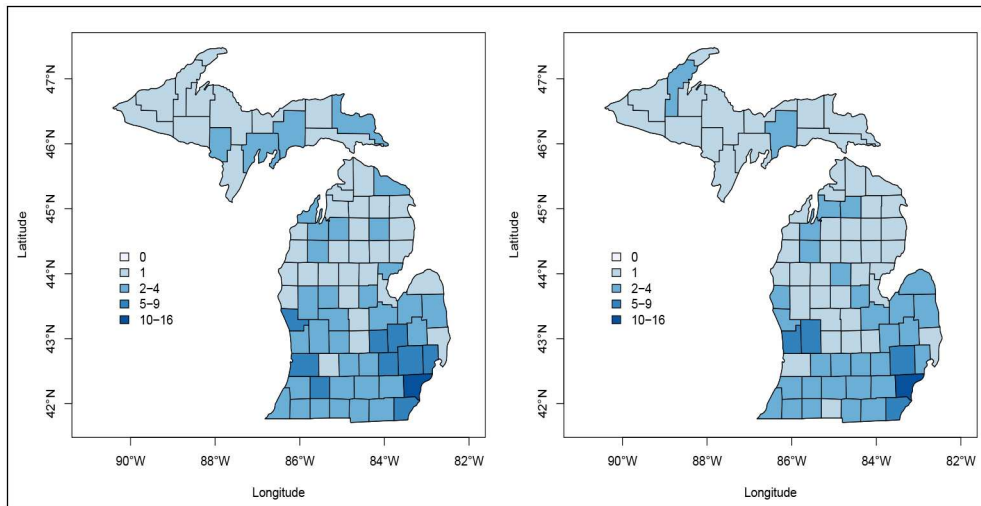


Figure 1 Reproduced figure from Chen et al. (2014): County-level teen driver fatal car crash counts in Michigan in 1996 (left) and 2004 (right)

involvement rate comparing to other driver age groups (Williams, 2003). Several approaches to reduce the number of fatal car accidents involving teenagers have been tested on and some are more successful than others. One effective policy that has been widely implemented is Graduated Driver Licensing (GDL; Shope, 2007). The three stages of GDL adopted by the state of Michigan in 1997 are as follows.

Stage I: Supervised learner period, novice teenagers are required to drive 20–60 h and hold the supervised license for a minimum length of time (6–12 months).

Stage II: Intermediate stage of GDL, teenagers are allowed to drive independently with restrictions on driving at night and passengers.

Stage III: Full-license stage of GDL, grants experienced teenager drivers full driving privileges without restrictions.

Figure 1 (reproduced plots and the original ones are in Chen et al. (2014)) presents the teen driver fatal car crash counts before and after the implementation of GDL in the Michigan State. GDL became effective in Michigan in 1997. Similar to the findings in Chen et al. (2014), our data analysis results also show that GDL is effective. But our modelling is different than that in Chen et al. (2014). The key difference is that we distinguish the population of teenagers and the population of teen drivers, while they are considered approximately the same in Chen et al. (2014). More specifically, we embed a measurement error model on the offset variable in a Poisson regression model to reflect the difference between the aforementioned two populations.

1.2 Motivating data example

We used the same data as that was analysed in Chen et al. (2014). Instead of following Chen et al.'s model, we use a Poisson regression to fit the data. The data are extracted from multiple sources, which are all publicly available. Previous studies showed that unemployment rate (Lenk, 1999), Rural–Urban Continuum index (Noland and Quddus, 2004) were associated with teenager-driver car crash counts. The data for unemployment rate in the state of Michigan were obtained from the US Bureau of Labor Statistics (<https://www.bls.gov/lau/#tables>). The Rural–Urban Continuum Codes were downloaded from the US Department of Agriculture (USDA) (<https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/>).

County-level teenager population data in the state of Michigan were extracted from the US Census database. Teenager population is used to adjust yearly counts of fatal teenager-driver crashes for variation in the size of the teenager population across counties, it can be incorporated as an offset term to the model. The population is widely considered as a surrogate for the number of licensed drivers and used for modelling the fatality rate in roadway traffic studies. This leads to a hierarchical model with measurement error, which will be discussed in following sections.

Because the recession of automobile industry in Michigan started from 2005 and this lead to changes in driving behaviour and the number of car crashes, we followed Chen et al. (2014) to collect the data seven years before (1990–1996) and after (1998–2004) the implementation of GDL. In Michigan, the most significant and recent urbanization period occurred during the auto-industrial boom that took place before the 1980s. Same as what Chen et al. (2014) did in their data analysis, we used the Rural–Urban Continuum Codes updated in 2003. This indeed does not reflect the true measurement of rurality during the study period 1990–2004. But given the code is updated only every 10 years, we consider it as an acceptable approximation.

Let i index counties in Michigan, $i = 1, \dots, 83$ and j index years (1990 to 2004 excluding 1997). The datasets consist of the following information:

n_{ij} : The number of teenager population for county i in year j ;

$X_{1_{ij}}$: Unemployment rate of county i in year j ;

X_{2_i} : Rural–Urban Continuum Codes of county i , the Rural–Urban Continuum Codes range from 1 to 9 with higher scores indicating more degree of rurality;

O_{ij} : The fatal car crash counts for county i in year j .

1.3 Measurement error in spatial modelling

There is a crucial yet implicit assumption in classical regression analysis: covariates are measured precisely. However, such an assumption is often violated in real life setting. Variables measured with error are commonly known as measurement error or errors-in-variables problem. There has been very rich literature in this area. For example, Carroll et al. (2006) provides comprehensive and systematic discussions as a classical textbook on this topic. Gustafson (2004) mainly focuses on Bayesian

methods with emphasis on applications in epidemiological studies. Yi (2017) includes an update of recent developments for a variety of settings (e.g., event history data, correlated data, multi-state event data).

Relatively less literature addressed the measurement errors in the context of spatial data (e.g., Bernadinelli et al., 1997; Xia and Carlin, 1998; Li et al., 2009; Huque et al., 2016, 2014). All of the existing work focuses on error-prone covariates. To the best of our knowledge, no work has been conducted to adjust for the measurement error in an offset variable in combination with spatial random effects.

Conditional autoregressive (CAR) modelling (Cressie, 1993) is considered to account for the spatial dependence through random effects in multi-level models. Suppose we have random variables $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_n)'$ for n geographic locations, respectively. The CAR model for these random effects is expressed via full conditional distributions:

$$\phi_i | \phi_j, j \neq i \sim N(\alpha \sum_{j=1}^n b_{ij} \phi_j, \tau_i^{-1}), \quad (1.1)$$

where τ_i is a spatially varying precision parameter, and $b_{ii} = 0$.

By Brook's Lemma (Brook, 1964), the joint distribution of $\boldsymbol{\phi}$ is:

$$\boldsymbol{\phi} \sim N(0, [A_\tau(I - \alpha B)]^{-1}), \quad (1.2)$$

where $A_\tau = \tau A$ and $A = \text{diag}(m_i)$ is an $n \times n$ diagonal matrix with m_i being the number of neighbours of location i ; I is an $n \times n$ identity matrix; α is a parameter that controls spatial dependence; $B = A^{-1}W$ (scaled adjacency matrix); W is the adjacency matrix with $w_{ii} = 0$ and $w_{ij} = 1$ if i is a neighbour of j , and 0 otherwise). Therefore, the CAR prior specification can be expressed as

$$\boldsymbol{\phi} \sim N(0, [\tau(A - \alpha W)]^{-1}). \quad (1.3)$$

The parameter α ($|\alpha| < 1$) is usually considered to be nonnegative because closer neighbourhoods tend to be more similar. In our analysis, we also include possible negative spatial dependence via a uniform prior over $(-1, 1)$. The results suggest a positive spatial dependence for the motivating data example. The results can be found in Appendix B. When α is valued 1, it is the Intrinsic Conditional Autoregressive (IAR) prior, which creates a singular precision matrix and an improper prior distribution. We also tried IAR prior but the samples, particularly for the precision parameter, have non-convergence issues. Results are not reported here.

2 Model frame and notations

To explain the difference between the models in Chen et al. (2014) and the models in our data analysis. We will first introduce their models and then our models in the following subsections.

2.1 Models in Chen et al. (2014)

Chen et al. (2014) applied a log transformation to the adjusted observed yearly county-level fatality rates as their response variable, that is, $y_{ij} = \log((O_{ij} + 1)/n_{ij})$. Normality is assumed for the response variable, that is,

$$y_{ij} | \mu_{ij}, \sigma^2 \sim N(\mu_{ij}, \sigma^2), \quad (2.1)$$

with the mean

$$\mu_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2i} + \beta_3 T_{1j} + \beta_4 T_{2j} + V_i. \quad (2.2)$$

Here X_{1ij} and X_{2i} refer to the unemployment rate and rurality, respectively. V_i denote county specific random effects and

$$\begin{aligned} T_{1j} &= \max(1997 - \text{year}_j, 0); \\ T_{2j} &= \max(0, \text{year}_j - 1997). \end{aligned}$$

It is worth noting that the adjusted log fatality rates y_{ij} are smaller than 0 because $0 < (O_{ij} + 1)/n_{ij} < 1$. Therefore, the support of the distribution of log fatality rate does not match with the support of a normal distribution.

2.2 Our proposed models

First, we introduce the naïve model *without* adjustment for the measurement error, that is,

$$\begin{aligned} O_{ij} | m_{ij} &\sim \text{Poisson}(m_{ij}), \\ \log(m_{ij}) &= \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2i} + \beta_3 T_j + \log(n_{ij}) + \phi_i, \end{aligned} \quad (2.3)$$

where $T_j = 0$ if $\text{year} < 1997$, $T_j = 1$ if $\text{year} > 1997$; ϕ_i are county specific random effects. A CAR prior for ϕ is considered. The details of the CAR prior are presented in Section 2.3.1.

In the previous model, we modelled the fatality rates in the Poisson regression by adding the offset term $\log(n_{ij})$. However, the genuine interest is to model the fatality rate for licensed teen drivers, that is, O_{ij}/D_{ij} . Here D_{ij} is the total number of licensed teen drivers for county i in year j . The same issue was raised in Chen et al. (2014) as well because the data of teen drivers in Michigan are not available at the county level from the year 1990 to 2004. Often in roadway traffic studies, the population is considered as a surrogate for the number of licensed drivers and widely used for modelling the fatality rate. In our models, we adjust for the difference between n_{ij} and D_{ij} rather than simply ignoring such difference.

Simply ignoring the measurement error in offset may cause biases in estimated regression coefficients. As shown in Chapter 3 of Zhang (2019), we can see the

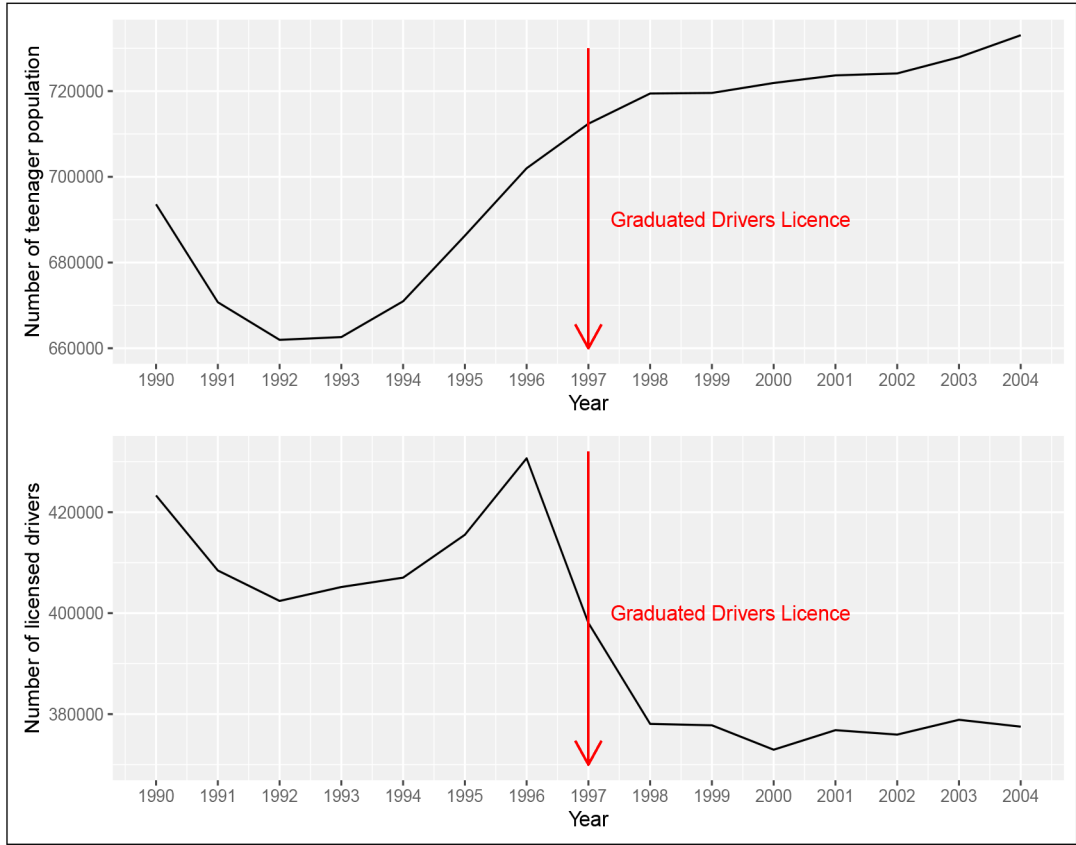


Figure 2 State level teen population and number of licensed teen driver from year 1990 to 2004

apparent discrepancy in parameter estimates between the models with and without adjustment for measurement error, particularly for regression coefficients of the predictors that are highly correlated with the number of licensed teen drivers.

The total number of licensed teen drivers in Michigan is available at the state level from year 1990 to 2004 excluding 1997. For comparison, the state level teen population and the number of licensed teen drivers are plotted in Figure 2, which show that the teen population and the number of licensed teen drivers have opposite trends after the implementation of GDL.

The actual offset variable should be $\log(D_{ij})$, but the number of licensed teen drivers at the county level is not available. We correct the measurement error in offset by introducing the rate of teen drivers denoted as R_{ij} :

$$R_{ij} = \frac{D_{ij}}{n_{ij}}, \quad (2.4)$$

take the log of both sides of this equation:

$$\log(D_{ij}) = \log(n_{ij}) + \log(R_{ij}). \quad (2.5)$$

Now embedding the measurement error model to account for the difference between D_{ij} and n_{ij} , we have proposed the following models:

$$\begin{aligned} O_{ij} | m_{ij}^* &\sim \text{Poisson}(m_{ij}^*), \\ \log(m_{ij}^*) &= \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2i} + \beta_3 T_j + \log(n_{ij}) + \log(R_{ij}). \end{aligned}$$

Because the population of licensed teen driver is obtained at state level, it is natural to consider a Berkson type of measurement error model, that is,

$$\text{logit}(R_{ij}) = \text{logit}(r_j) + \eta_i, \quad (2.6)$$

where $r_j = \sum_{i=1}^{83} D_{ij} / \sum_{i=1}^{83} n_{ij}$, represents the proportion of teen driver among teenager population at the state level; η_i for county specific random effects. The reason for using logit of R_{ij} is to avoid constrained parameters in a proposed model. We tried other possible forms of (2.6) but none of them worked. For example, we included a covariate X_1 in the measurement error model but ended up with non-convergence issue in the sampling for its regression coefficient. We also tried to add an extra random effects term independent of η_i , that is, a Besag–York–Mollié (BYM) type model considered in Chen et al. (2014). But the MCMC samples did not mix well.

Here is the summary of the differences between our proposed model and the one proposed in Chen et al. (2014). The key difference is the measurement error model we added to account for the difference between D_{ij} and n_{ij} . In addition, there are some other differences. First, we consider a Poisson model for the raw count data O_{ij} . Second, instead of using two time terms, T_1 and T_2 , we simply use an indicator variable T in the regression model. We also include the results in Appendix D for using two time terms as proposed in Chen et al. (2014), which will be summarized in our discussion part. Third, we add one more parameter α for controlling the spatial dependence when applying the CAR prior for the county specific random effects to our model, which is more flexible. In Chen et al. (2014), the authors used the IAR prior with α being fixed to 1.

We consider a Bayesian approach for model estimation. Details are presented in the following subsection.

2.3 Bayesian approach

2.3.1 Prior specification

In general, the prior distribution reflects the prior knowledge or information about the parameter of interest before collecting the data. If there is no such knowledge, then weakly-informative priors need to be assigned to the model parameters.

1. Prior for β : The regression coefficients of the Poisson regression model will be assigned independent normal priors with zero mean and standard deviation of

100. That is,

$$\beta_k \sim \text{Normal}(0, 100^2), \quad k = 0, 1, 2, 3.$$

2. A CAR prior for $\phi(\eta)$ in the model without (with) adjustment for measurement error : Let $\phi = (\phi_1, \phi_2, \dots, \phi_n)$ denote the county-level random effects either in the Poisson model or measurement error model. Recall the notations in Section 1.3 for a CAR prior, $\phi \sim N(0, [\tau(A - \alpha W)]^{-1})$. To ease the notation, we denote a CAR prior by $\phi | \alpha_\phi, \tau_\phi \sim \text{CAR}(\alpha_\phi, \tau_\phi)$, where τ_ϕ is the precision parameter. Similarly, for the random effects η in the measurement error model, $\eta | \alpha_\eta, \tau_\eta \sim \text{CAR}(\alpha_\eta, \tau_\eta)$. The CAR model is implemented in rstan following a STAN case study by Joseph (<https://mc-stan.org/users/documentation/case-studies/mbjoseph-CARStan.html>).
3. Prior for $1/\sqrt{\tau}$: Following the suggestion in Gelman (2006), we employ uniform $(0, M)$ prior for standard deviation. A value of $M = 10$ is specified as the upper limit of the uniform prior throughout this study, that is, $\sqrt{\tau} \sim \text{Unif}(0, 10)$.
4. Prior for spatial correlation α : The results of using a uniform prior for α are presented in Appendix C. As we can see from the posterior summary statistics from Appendix C, the posterior distribution of α is not much different from the prior. Therefore an informative prior is recommended if such prior information is available in practice. To explain, we consider gamma distributions for an informative prior choice. The hyperparameters, that is, the two shape parameters in a beta distribution, are chosen by utilizing the prior information.

2.3.2 Stan/rstan for Bayesian MCMC sampling

The software Stan (Gelman et al., 2015) has been becoming popular due to its applicability to a wide variety of model settings. Stan differs from BUGS and JAGS in two primary ways. First, Stan is based on a new probabilistic programming language that is more flexible and expressive than the declarative graphical modelling languages underlying BUGS or JAGS, in ways such as declaring variables with types and supporting local variables and conditional statements. Second, the MCMC techniques in Stan are based on Hamiltonian Monte Carlo, a more efficient and robust sampler than Gibbs sampling or Metropolis-Hastings especially for complex models. In our data analysis, the Bayesian MCMC sampling is conducted by using an R package ‘rstan’, which is an R interface for Stan. The source code for the data analysis can be found through <http://www.statmod.org/smij/archive.html>.

3 Data analysis

3.1 Model estimation

Given the complexity of our proposed models involving spatial random effects together with measurement error in the offset variable, we use Bayesian MCMC methods for model estimation.

The prior distributions for all parameters are listed in Section 2.3.1. We conduct the real data analysis under three different priors for the spatial correlation in the CAR model for the spatial random effects. We choose the values of the two hyperparameters to reflect the prior information on the spatial dependence.

Prior 1: Beta(17.80, 74.10) is used when the prior information suggests a weak spatial dependence. That is, 2.5% and 97.5% percentiles of the prior distribution are 0.12 and 0.28, respectively.

Prior 2: Beta(47.30, 47.30) is used when the prior information suggests a moderate spatial dependence. That is, 2.5% and 97.5% percentiles of the prior distribution are 0.4 and 0.6, respectively.

Prior 3: Beta(52.50, 12.80) is used when the prior information suggests a strong spatial dependence. That is, 2.5% and 97.5% percentiles of the prior distribution are 0.70 and 0.89, respectively.

We ran multiple chains in all cases, and inspecting the results and the trace plots show no concern about convergence of the MCMC sampler. The posterior mean, standard deviation and 95% credible interval of each parameter in two models are summarized in Tables 1, A1 and A2 under three different priors for $\alpha_\phi(\alpha_\eta)$. Figures 3, A1 and A2 present a comparison of posterior distributions of parameters under models with and without adjustment for measurement error. The results indicate that the three different choices of prior distribution for $\alpha_\phi(\alpha_\eta)$ do not affect the estimation results.

As indicated by the tables and plots, the posterior inferences for the regression coefficients of the unemployment rate and rurality are not affected by the adjustment of measurement error. While the posterior distributions of intercept and regression coefficient of time indicator have noticeable changes after adjusting for the measurement error. Moreover, the posterior distribution of the precision parameter τ_ϕ under the naïve model is obviously different than that of the precision parameter τ_η under the measurement error models. This is not surprising because they are the precision parameters with respect to different random effects distributions.

It is worth noting that the results are similar under three different prior choices for α_ϕ/α_η . This suggest that the posterior distribution of the spatial dependence parameter is dominated by the data. We only present the results under Prior 1 in the text and the results under the other two prior choices are presented in Appendix A.

3.2 Residual diagnostics for model assessment

We check the fit of our proposed model by examining the randomized quantile residuals (Dunn and Smyth, 1996). In this section, only the results under the prior 2 for $\alpha_\phi(\alpha_\eta)$ are included. The results for prior 1 and prior 3 are similar and thus not reported here.

Table 1 Posterior mean, standard deviation (SD) and 95% credible intervals (CI) of parameters for the car crash data with and without adjustment for measurement error (under prior 1)

	Without Adjustment			With Adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	-8.44	0.07	(-8.58, -8.30)	-7.88	0.06	(-8.00, -7.75)
β_1	-0.10	0.04	(-0.19, -0.02)	-0.10	0.04	(-0.18, -0.01)
β_2	0.16	0.06	(0.03, 0.28)	0.14	0.06	(0.02, 0.26)
β_3	-0.30	0.06	(-0.42, -0.18)	-0.12	0.06	(-0.24, 0.00)
τ_ϕ	1.51	0.45	(0.83, 2.57)	—	—	—
α_ϕ	0.20	0.04	(0.13, 0.29)	—	—	—
τ_η	—	—	—	0.29	0.09	(0.14, 0.50)
α_η	—	—	—	0.20	0.04	(0.13, 0.29)

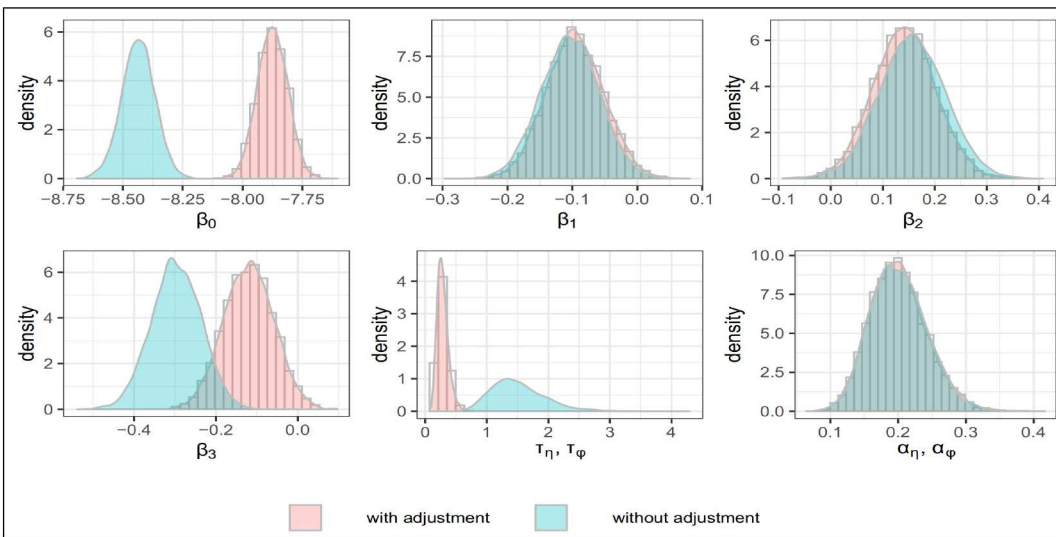


Figure 3 Posterior distributions comparisons with and without measurement adjustment (under prior 1)

Residuals are commonly used for statistical model checking. In normal linear regression model, QQ plots of residuals are commonly used as a model diagnostic tool. However, for a generalized linear model, the conventional diagnostics tool may fail as the distribution of residuals can be non-normal and heteroscedasticity even under the right model. A more general definition of residuals is proposed by Dunn and Smyth (1996), termed *randomized quantile residuals* (RQR). The key point is to add a randomization step when evaluating the fitted distribution function for a (discrete) response value to achieve continuous residuals. Basically, this is an extension from a continuous cumulative distribution function to a right-continuous distribution function. The randomized quantile residuals are (approximately) uniformly distributed over (0,1). ‘DHARMa’ (Ecology et al., 2019)

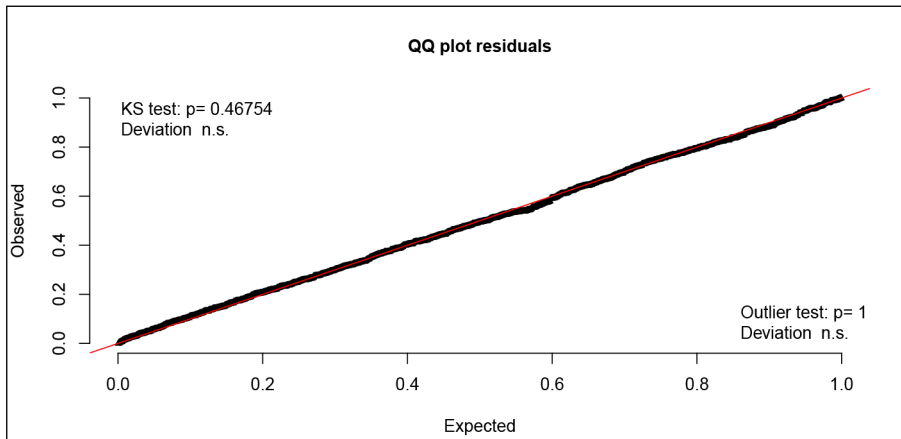


Figure 4 QQ plot of DHARMA residuals with measurement error adjustment (under prior 2)

is an R package for model diagnostics built upon the definition of RQR. In this section, we conduct residual diagnostics by using the package ‘DHARMA’ to create readily interpretable RQRs based on the 6 000 simulated (replicated) datasets of car crash counts. Figure 4 presents QQ plots based on quantile residuals from the model with adjustment for measurement error under moderate spatial dependence setting. The plot does not suggest obvious evidence for model-data misfit.

However, we acknowledge the limitations of this diagnostics method. First, it is in-sample prediction when we calculate the RQRs because the same data is used for model estimation and prediction. It is well known that this model check method can be overly optimistic. Second, in the original definition of RQR proposed in Dunn and Smyth (1996), the response variable has *independent* observations. Obviously this is not the case for the car crash counts in the motivating example.

4 Discussion

In this study, we use Poisson regression model for analysing teen driver car crash counts with adjustment for errors in an offset variable. The model is clearly different than the previous work by Chen et al. (2014) for analysing the same data. We compare the results with a naïve model without adjustment for measurement error, that is, treating the population of teenagers as the same as the population of licensed teen drivers. We find that the inferences for the regression coefficients of X_1 (unemployment rate) and X_2 (rurality) are little affected by the adjustment. However, the inferences for the intercept and the coefficient of T (coded as 0 before 1997 and 1 after 1997) are different after adjustment for measurement error. The posterior mean of the regression coefficient of T is -0.30 under the naïve model and became -0.12 under the adjustment model. Similar results can be found when using two-time terms

Table 2 WAIC values for the model with a single time covariate and two time covariates under five different CAR priors. Priors 1–3 are stated on page 9–10. Prior 4 is Uniform(−1,1) considered in Appendix B. Prior 5 is Uniform(0,1) considered in Appendix C

	Prior 1	Prior 2	Prior 3	Prior 4	Prior 5
Single time covariate					
Naïve	2648.2	2645.5	2642.6	2643.8	2643.0
ME	2648.6	2645.6	2642.6	2644.6	2644.2
Two time covariates					
Naïve	2657.9	2654.9	2652.5	2653.0	2653.0
ME	2652.2	2652.2	2647.4	2648.2	2648.2

in the model from Appendix D. We compare the WAIC values for the two models with one or two time variables under different prior choices of the spatial correlation. The results are reported in Table 2. The model with one time variable is preferred due to its consistently smaller values across different priors. Though the WAICs of the naïve model and our proposed model are very close, we argue that it is not a sensible assumption that the teen population size can be used as a good approximation for the size of teen licensed drivers (as indicated by Figure 2). Therefore, the regression coefficient of T in the naïve model cannot provide a valid interpretation for the effect of the GDL policy. As a comparison, our model takes into account the difference between teen population and teen driver population. Therefore, the fact of decreased number of licensed teen drivers after implementing the GDL policy is taken care of by the ratio between the licensed teen driver population size and the teen population size. In other words, our model allows a two-fold interpretation of GDL effect. First, the number of licensed teen drivers in the State of Michigan is decreased after the implementation of the GDL policy. Second, a negative value of the regression coefficient of T suggests reduced risk of fatal car crashes among the *licensed teen drivers*. This part is the commonly seen interpretation of effectiveness of GDL, which may be related to the increased driving experience caused by GDL. In summary, the findings of our modelling strategy provide a new perspective of interpreting the effectiveness of GDL.

To the best knowledge of our knowledge, there is no existing literature about the adjustment of mismeasured offset in the context of spatial data. In our case, we construct the measurement error model with spatial random effects. In a similar fashion, this adjustment for errors in offset variables can also be applied to other models (e.g., Zero-Inflated Poisson model, Negative Binomial model) in spatial data case. In fact, zero-inflated Poisson models are also fitted to the real data example. Please refer to the Master’s thesis of Zhang (2019) for details if interested.

Acknowledgement

We would like to thank the comments/suggestions from the anonymous reviewers that have improved the presentation of the manuscript.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: The corresponding author's research was supported by a Discovery grant from the Natural Sciences and Engineering Research Council. The last co-author's research was supported by a grant from the National Cancer Institute (U01-CA057030), and also by a TRIPODS grant from the U.S. National Science Foundation (NSF CCF-1934904).

Appendix

A Results under prior 2 and prior 3

Table A1 Posterior mean, standard deviation (SD) and 95% credible intervals (CI) of parameters for the car crash data with and without adjustment for measurement error (under prior 2)

	Without Adjustment			With Adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	-8.44	0.08	(-8.60, -8.28)	-7.87	0.07	(-8.01, -7.74)
β_1	-0.10	0.04	(-0.19, -0.02)	-0.10	0.04	(-0.18, -0.01)
β_2	0.16	0.07	(0.01, 0.29)	0.15	0.06	(0.03, 0.27)
β_3	-0.30	0.06	(-0.41, -0.18)	-0.12	0.06	(-0.24, 0.00)
τ_ϕ	1.52	0.43	(0.85, 2.50)	—	—	—
α_ϕ	0.51	0.05	(0.41, 0.61)	—	—	—
τ_η	—	—	—	0.30	0.10	(0.15, 0.53)
α_η	—	—	—	0.51	0.05	(0.40, 0.61)

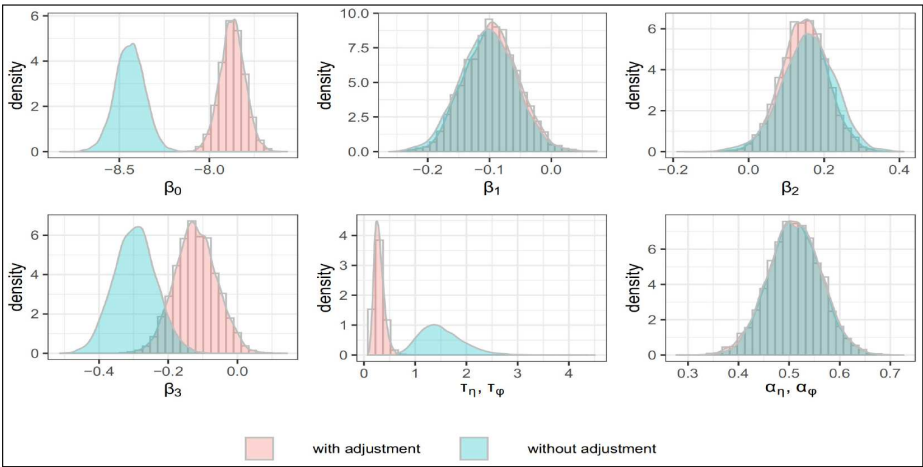


Figure A1 Posterior distributions comparisons with and without measurement adjustment (under prior 2)

Table A2 Posterior mean, standard deviation (SD) and 95% credible intervals (CI) of parameters for the car crash data with and without adjustment for measurement error (under prior 3)

	Without Adjustment			With Adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	-8.44	0.11	(-8.66, -8.22)	-7.86	0.08	(-8.02, -7.69)
β_1	-0.10	0.05	(-0.19, -0.01)	-0.10	0.04	(-0.18, -0.01)
β_2	0.16	0.08	(0.01, 0.31)	0.15	0.07	(0.02, 0.28)
β_3	-0.30	0.06	(-0.42, -0.17)	-0.12	0.06	(-0.24, 0.00)
τ_ϕ	1.70	0.50	(0.94, 2.89)	—	—	—
α_ϕ	0.82	0.05	(0.72, 0.90)	—	—	—
τ_η	—	—	—	0.35	0.12	(0.16, 0.63)
α_η	—	—	—	0.81	0.05	(0.71, 0.90)

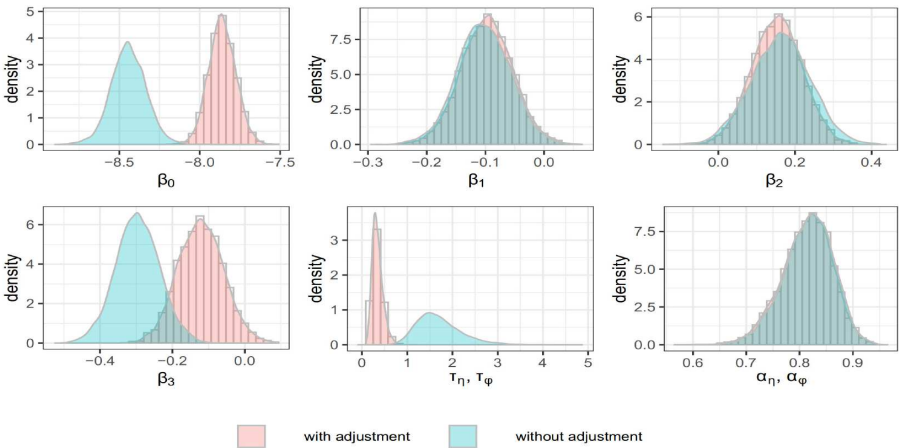


Figure A2 Posterior distributions comparisons with and without measurement adjustment (under prior 3)

B Results when using a Unif(−1, 1) prior for $\alpha_\phi(\alpha_\eta)$

In this appendix, the posterior summary are based on the same model settings as model 2.3 and model 2.6 in Section 2.2 but with a Unif(−1, 1) prior for $\alpha_\phi(\alpha_\eta)$.

Table B1 Posterior mean, standard deviation (SD) and 95% credible intervals (CI) of parameters for the car crash data with and without measurement error adjustment with a Unif(−1, 1) prior for $\alpha_\phi(\alpha_\eta)$

	Without Adjustment			With Adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	−8.46	0.14	(−8.80, −8.20)	−7.86	0.09	(−8.03, −7.68)
β_1	−0.10	0.05	(−0.19, −0.01)	−0.10	0.04	(−0.19, −0.01)
β_2	0.16	0.08	(0.01, 0.31)	0.15	0.07	(0.02, 0.28)
β_3	−0.30	0.06	(−0.41, −0.18)	−0.12	0.06	(−0.25, 0.01)
τ_ϕ	1.72	0.53	(0.92, 2.95)	—	—	—
α_ϕ	0.75	0.24	(0.08, 0.98)	—	—	—
τ_η	—	—	—	0.34	0.13	(0.15, 0.65)
α_η	—	—	—	0.70	0.27	(−0.04, 0.98)

C Results from a Unif(0, 1) prior setting for α

The following posterior summary statistics are from the same model settings as model 2.3 and model 2.6 in Sections 2.1 and 2.2. However, here we use a Unif(0, 1) prior for α instead of using an informative prior as mentioned in Section 2.3.1.

Table C1 Posterior mean, standard deviation (SD) and 95% credible intervals (CI) of parameters for the car crash data with and without measurement error adjustment with a unif(0, 1) prior for $\alpha_\phi(\alpha_\eta)$

	Without adjustment			With adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	−8.44	0.15	(−8.75, −8.14)	−7.86	0.09	(−8.03, −7.68)
β_1	−0.10	0.04	(−0.19, −0.02)	−0.10	0.04	(−0.19, −0.01)
β_2	0.17	0.08	(0.01, 0.31)	0.15	0.06	(0.02, 0.28)
β_3	−0.30	0.06	(−0.41, −0.18)	−0.12	0.06	(−0.24, 0.00)
τ_ϕ	1.73	0.53	(0.93, 2.99)	—	—	—
α_ϕ	0.77	0.20	(0.20, 0.99)	—	—	—
τ_η	—	—	—	0.35	0.14	(0.15, 0.71)
α_η	—	—	—	0.73	0.21	(0.18, 0.98)

D Results when using two time covariates

This section presents the results from the same model settings as model 2.3 and model 2.6 in Section 2.2 except for using two time terms (T_1 and T_2), where $T_{1j} = \max(1997 - \text{year}_j, 0)$ and $T_{2j} = \max(0, \text{year}_j - 1997)$.

Model without measurement error adjustment including two time terms:

$$O_{ij}|X_{1ij}, X_{2i}, T_{1j}, T_{2j} \sim \text{Poisson}(m_{ij}),$$

$$\log(m_{ij}) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2i} + \beta_3 T_{1j} + \beta_4 T_{2j} + \log(n_{ij}) + \phi_i.$$

Model with measurement error adjustment including two time terms:

$$O_{ij}|X_{1ij}, X_{2i}, T_{1j}, T_{2j} \sim \text{Poisson}(m_{ij}^*),$$

$$\log(m_{ij}^*) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2i} + \beta_3 T_{1j} + \beta_4 T_{2j} + \log(n_{ij}) + \log(R_{ij}),$$

$$\text{logit}(R_{ij}) = \text{logit}(r_j) + \eta_i.$$

Table D1 Posterior mean, standard deviation (SD) and 95% credible intervals (CI) of parameters for the car crash data with and without measurement error adjustment (under prior 1)

	Without Adjustment			With Adjustment		
	Mean	SD	95%CI	Mean	SD	95%CI
β_0	-8.58	0.08	(-8.75, -8.42)	-7.92	0.08	(-8.08, -7.77)
β_1	-0.16	0.06	(-0.28, -0.05)	-0.11	0.05	(-0.22, 0.00)
β_2	0.18	0.07	(0.05, 0.31)	0.14	0.06	(0.02, 0.26)
β_3	0.03	0.02	(0.00, 0.06)	0.01	0.02	(-0.02, 0.04)
β_4	-0.04	0.02	(-0.07, -0.01)	-0.01	0.02	(-0.05, 0.02)
$\beta_4 - \beta_3$	-0.07	0.02	(-0.10, -0.03)	-0.02	0.02	(-0.06, 0.01)
τ_ϕ	1.42	0.41	(0.78, 2.36)	—	—	—
α_ϕ	0.20	0.04	(0.12, 0.29)	—	—	—
τ_η	—	—	—	0.29	0.10	(0.14, 0.52)
α_η	—	—	—	0.20	0.04	(0.12, 0.29)

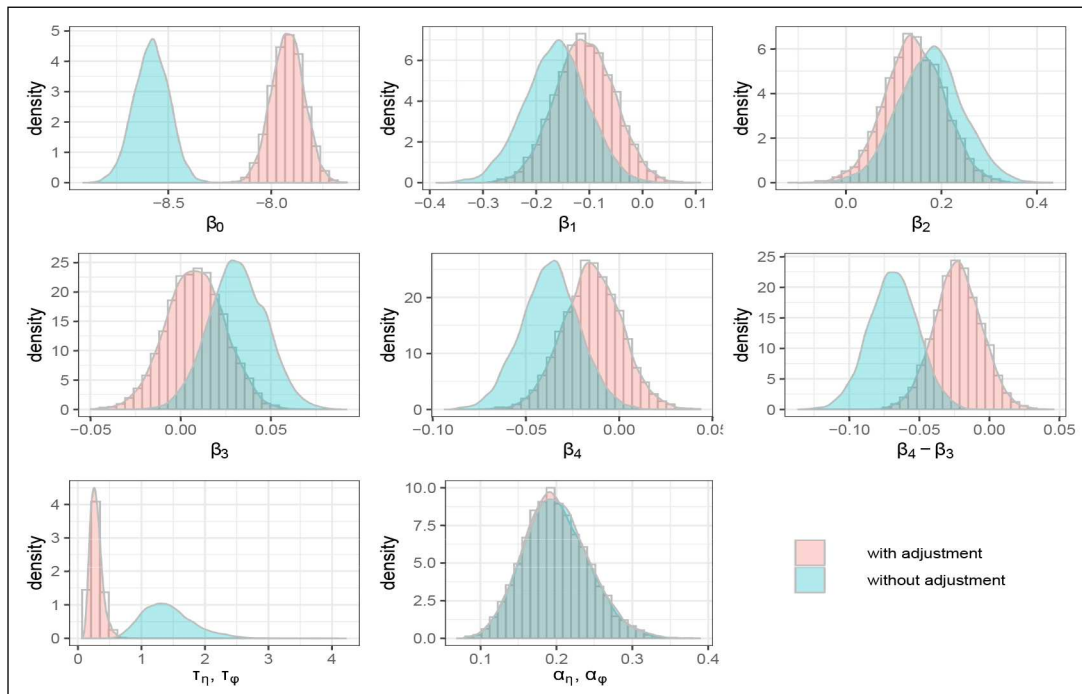


Figure D1 Posterior distributions comparisons with and without measurement adjustment (under prior 1)

References

- Bernadinelli L, Pascutto C, Best NG and Gilks WR (1997) ‘Disease mapping with errors in covariates. *Statistics in Medicine*, **16**, 741–52.
- Brook D (1964) On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika*, **51**, 481–83.
- Carroll RJ, Ruppert D, Stefanski LA and Crainiceanu CM (2006) *Measurement Error in Nonlinear Models: A Modern Perspective, 2nd edition*. Boca Raton, FL: CRC Press.
- Chen Y, Berrocal VJ, Raymond Bingham C and Song PCK (2014) Analysis of spatial variations in the effectiveness of graduated driver’s licensing (GDL) program in the state of Michigan. *Spatial and Spatio-temporal Epidemiology*, **8**, 11–22.
- Cressie N (1993) Regional mapping of incidence rates using spatial Bayesian models. *Medical Care*, **31**, YS60–YS65.
- Dunn PK and Smyth GK (1996) Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**, 236–44.
- Ecology FHT, Regensburg Uo, Regensburg and Germany (2019) *DHARMA: Residual diagnostics for hierarchical (multi-level/mixed) regression models*. URL <https://CRAN.R-project.org/package=DHARMA> (last accessed 8 April 2021).

- Gelman A (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, **1**, 515–34.
- Gelman A, Lee D and Guo J (2015) Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, **40**, 530–43.
- Gustafson P (2004) *Measurement Error and Miss-classification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. London: Chapman & Hall/CRC.
- Huque MH, Bondell HD, Carroll RJ and Ryan LM (2016) Spatial regression with covariate measurement error: A semiparametric approach. *Biometrics*, **72**, 678–86.
- Huque MH, Bondell HD and Ryan L (2014) On the impact of covariate measurement error on spatial regression modelling. *Environmetrics*, **25**, 560–70.
- Lenk PJ (1999) Bayesian inference for semiparametric regression using a Fourier representation. *Journal of the Royal Statistical Society, Series B*, **61**, 863–79.
- Li Y, Tang H and Lin X (2009) Spatial linear mixed models with covariate measurement errors. *Statistica Sinica*, **19**, 1077–93.
- Noland RB and Quddus MA (2004) A spatially disaggregate analysis of road casualties in England. *Accident Analysis & Prevention*, **36**, 973–84.
- Shope JT (2007) Graduated driver licensing: Review of evaluation results since 2002. *Journal of Safety Research*, **38**, 165–75.
- Williams AF (2003) Teenage drivers: Patterns of risk. *Journal of Safety Research*, **34**, 5–15.
- Xia H and Carlin BP (1998) Spatio-temporal models with errors in covariates: Mapping Ohio lung cancer mortality. *Statistics in Medicine*, **17**, 2025–043.
- Yi GY (2017) *Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application*. New York, NY: Springer.
- Zhang K (2019) Measurement error adjustment in the offset variable of a Poisson model. *PhD thesis*, University of Saskatchewan, Saskatchewan, Canada.