4-2016

Bivariate Negative Binomial Hurdle with Random Spatial Effects

Robert McNutt
Western Michigan University, robertwmcnutt@gmail.com

BIVARIATE NEGATIVE BINOMIAL HURDLE WITH RANDOM SPATIAL EFFECTS

by

Robert McNutt

A dissertation submitted to the Graduate College

in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

Statistics

Western Michigan University

April 2106

Doctoral Committee:

Rajib Paul, Ph.D., Chair

Magdalena Niewiadomska-Bugaj, Ph.D.

Joseph W. McKean, Ph.D.

Kathleen M. Baker, Ph.D.

BIVARIATE NEGATIVE BINOMIAL HURDLE MODEL WITH RANDOM SPATIAL EFFECTS

Robert McNutt, Ph.D.

Western Michigan University, 2016

Count data with excess zeros widely occur in ecology, epidemiology, marketing, and many other disciplines. Mixture distributions consisting of a point mass at zero and a separate discrete distribution are often employed in regression models to account for excessive zero observations in the data. While Poisson models are very popular for count data, Negative Binomial models provide greater flexibility due to their ability to account for overdispersion.

This research focuses on developing a method for analyzing bivariate count data with excess zeros collected over a lattice. A bivariate Zero-Inflated Negative Binomial Hurdle (ZINBH) regression model with spatial random effects is developed. The proposed model characterizes spatial and cross-spatial dependencies. Inferences on model parameters and predictions are done using samples from a Markov Chain Monte Carlo algorithm. We applied our proposed model on Michigan county level crime incidence data. In addition, a method for variable selection through Bayesian penalized regression is developed using a LASSO-type method and elastic net.

Acknowledgements

The author would like to acknowledge and thank his advisor, Dr. Rajib Paul, for his unflagging patience, tenacity and helpful advice through the completion of this thesis, without which this would not have come to fruition.

Also, thanks to the other members of the thesis committee, Dr. Magdalena Niewiadomka Bugaj, Dr. Joseph McKean and Dr. Kathleen Baker, for taking the time to read and make suggestions on this thesis.

Finally, the author would like to thank friends and family for their continuous encouragement while enduring (mostly willingly) the author's near continuous 'commentary' of the thesis process.

Robert McNutt

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Count data, responses comprising the value of zero or positive integers, arise in many situations where the quantity of objects or number of incidences of some phenomena is measured. Certain discrete probability distributions, such as Poisson, Binomial, and Negative Binomial, have support comprising these values and are often used to model this type of data. Sometimes the log of the response (or the response plus an offset, to deal with '0' observations) is modeled by regression as a linear function of the covariates. Another approach is to model the log of the expectation of the response, i.e. the distribution mean, as a linear function of the covariates. These 'log-linear' models are similar to many others which have modeled discrete or categorical data, see Agresti [2007].

The discrete distributions that can model these datasets incorporate a non-zero probability for a count of '0.' However, when there are too many zero observations, a

standard discrete distribution cannot account for the variation and overdispersion adequately. These datasets are referred to as 'zero-inflated' count data; for examples in medical and epidemiological studies, see Lachenbruch [2002], J. Tooze and Jones [2002], Wang and Carrivick [2003], Buu et al. [2011], Ghosh et al. [2012], H. Liu and Chan [2012], Moulton and Halsey [1995], and the references therein. In these cases, the data are modeled as a mixture of two distributions, a probability point mass equal to π at zero and a conventional discrete distribution with weight equal to $1 - \pi$. In this way the probability of observing a zero count response is increased by adding a point mass π at response value zero. When zero observations occur from a single source, a 'hurdle' model is employed, wherein all zeros are accounted for by the point mass probability at zero and the discrete distribution is truncated to account for only the non-zero observations (see M. Ridout and Hinde [1998], Hu et al. [2011], Neelon et al. [2012], Neelon et al. [2014], and the references therein).

Two-part regression models of zero-inflated count data involve modeling one or more of the parameters of each distribution as linear functions of covariates measured for each observation (M. Ridout and Hinde [1998]). One parameterization of zero inflated count data is the Zero Inflated Poisson (ZIP) model, in which the point mass at zero is modeled through a logistic, probit, or log-log compliment regression, and the log-transformed mean of the Poisson portion is modeled by Poisson regression (Ghosh et al. [2006]). In the case where the estimates of the variance of the Poisson modeled datasets significantly exceed the estimates of the mean, this 'over dispersion' is a problem, since the single parameter Poisson model has the assumption of

equal magnitudes for mean and variance.

Another, more common way of addressing this problem is to treat the mean of the Poisson model as a product two components, one of which is regressed, the other a random variable from a Gamma distribution. In other words, the mean itself has a 'prior' distribution. In this way another parameter is introduced into the model which can allow for a wider range of variances than the Poisson model by itself. This Poisson-Gamma model leads to the 'posterior' distribution of the response being Negative Binomial (M. Zhou and Carin [2012]). The zero-inflated version of this is well known as the Zero-Inflated Negative Binomial (ZINB) distribution (Greene [1994]).

Spatial count data occurs when count data are obtained over designated spatially defined regions, and when dependence on location is incorporated in the model. Location can be measured continuously, such as latitude and longitude, or with some other referencing system (see Schmidt and Rodriguez [2010] and Cressie [1993]). Alternatively, location data may be aggregated, such as with counties, states, zip codes, or with other boundary divided spatial regions, often referred to as comprising a 'lattice.' For discrete location data, one way to model this spatial dependence is through (intrinsic) conditional autoregressive (ICAR) random effects (J. Besag and Mollie [1991]). In this method, effects within one lattice site are conditionally regressed on the value of the same effects in one or more other, usually neighboring lattice sites. Because the variable is regressed on itself, measured at different lattice

sites, it is said to be 'autoregressive'.

When a bivariate response is measured, for example under multiple condition incidences, the dependence grows more complicated, including cross-spatial and cross-variable (non-spatial) dependencies, involving both the point mass and the non-zero discrete distribution of each response. This requires characterization of a more complex variance-covariance structure and more flexible ways to estimate these parameters.

This research develops a bivariate Zero-inflated Negative Binomial Hurdle (ZINBH) model for analyzing spatial count data on a lattice. By modeling excess zeros through two-part regression and accounting for possible cross-variable and cross-spatial dependencies through multivariate ICAR random effects, bivariate data with many zeros are more accurately modeled than by separate univariate, traditional regression analysis on each response. We employ a Bayesian approach to avail ourselves of increased accuracy for small sample sizes and/or large numbers of zeros. We introduce a multi-level prior structure into the variance-covariance matrix of the multivariate normally distributed ICAR effect, thereby reducing the subjectivity. In addition, we formulate a variable selection method for spatial count data by incorporating penalized regression, via elastic net (Zhou and Hastie [2005]), into a univariate form of our ZINBH model. The rest of this paper will describe and develop the research described above in four chapters. We review, in chapter 2, the current literature surrounding zero-inflated models, ZINBH spatial models, and elastic net variable

selection method, and explain how this research differs. Chapter 3 presents the proposed model, starting with the parameterization of a bivariate ZINBH model with ICAR effects, as well as the weakly informative prior for the variance-covariance matrices associated with the ICAR effects. Model performance comparisons are first assessed using simulated data, through regression estimates and posterior predictive plots. Next, the models are used to analyze bivariate, Michigan county level, crime data. Chapter 3 closes with a discussion on the challenges associated with implementing these models in WinBUGS. We develop a variable selection method for univariate ZINBH regression in chapter 4 through penalized regression. A penalty function, with the regression estimates among the arguments, is introduced into the likelihood through the prior distribution on the regression parameters, and then simulated data are used to examine the performance. This paper closes with chapter 5, where conclusions are drawn from the research and recommendations are made about further research.

Chapter 2

Literature Review

Zero-inflated spatial count data models have been employed for some years. VerHoef and Jansen [2007] used a space-time ZIP model for describing seal discovery and capture patterns in Alaska. D. K. Agarwal and Citron-Pousty [2002] used ZIP regression to model the number of nests of isopods across a region of the Negev desert in Israel. Nadiroh [2009] compared Negative Binomial Regression (NBR), also known as the Poisson-Gamma model, with random area specific effects, to model simulated count data with excess zeros. Estimated MSE's showed that the NBR outperformed ZINB when there was a modest probabilities of zero, while ZINB outperformed NBR when the probability of zero exceeds 0.6. However, their data did not account for spatial correlation. Ismail and Zamani [2013] proposed model selection methods for Poisson, generalized Poisson, and Negative binomial regression models, for both 'regular' and zero-inflated models. These univariate models were implemented for Malaysian auto damage claims (over-dispersed) and German health care visits (zero-inflated

and over-dispersed). Various tests, including likelihood ratio tests (LRT), Wald and Vuong tests, as well as information criteria, including Aikake Information Criteria (AIC) and Bayesian Schwartz Information Criteria (BIC), were applied to determine and compare performances of these models. K. E. Staub and Adnan [2011] showed the advantages of Poisson Quasi-Likelihood (PQL) methods as an alternative to ZIP and ZINB regression models, in cases where the model is misspecified with respect to the actual distribution of count data, i.e. excess zeros (zero-inflated). This was demonstrated via MCMC results on both simulated data and Austrian health care survey data, including physician/specialist visitation count data. Osei [2010] employed ZINB regression on incidences of lip cancer in the 56 districts of Scotland. Neelon et al. [2010] and Neelon et al. [2012] use spatial ZIP hurdle models to describe emergency department visits in Durham County, NC, employing spatial random effects associated with both the occurrence and abundance of visits. However, none of the above work analyzed bivariate responses. The present work extends these models by including both cross-spatial and cross-variable dependencies for the case of bivariate data using an intrinsic conditional auto regressive (ICAR) prior on the spatial effects.

There has been some previous consideration of bivariate models. Wang [2003] employed a bivariate ZINB regression model to analyze two sets of overdispersed, dependent count data, the number of doctor visits and the number of non-doctor health related visits, as reported in an Australian health survey. Wang and Carrivick [2003] modeled a bivariate ZIP regression on the number of muscular and non-muscular loss of time injuries (MLTI, NMLTI) to hospital cleaning staff for pre-

and post ergonomic team intervention. Both Wang [2003] and Wang and Carrivick [2003] used an EM algorithm to obtain maximum likelihood estimates and treat a 'zero observation' as consisting of zeros in both responses simultaneously. This was limited by fitting a point mass at zero only for the probability of simultaneously occurring zeros. Our work differs in that it uses a Bayesian approach to allow for a separate point mass at zero for each response. Also, neither Wang [2003] nor Wang and Carrivick [2003] consider random spatial effects. We fit a separate point mass at zero for each response while including random spatial effects on a lattice. More recently, Majumdar and Gries [2010] developed a regression model for bivariate count data with excess zeros in both responses separately. Their model incorporates data augmentation to account for a mixing probabilities of bivariate responses $[Y_1, Y_2]'$ under the four states $(Y_1 = Y_2 = 0), (Y_1 = 0, Y_2 > 0), (Y_1 > 0, Y_2 = 0), (Y_1 > 0, Y_2 > 0)$, among two independent ZIP regressions. In their application, the responses were the occurrences of two distinct plant species. But in this case, Majumdar and Gries [2010] did not consider random spatial effects.

Variable selection via penalized regression has developed over the years in various forms. Hoerl and Kennard [1970] introduced Ridge Regression which minimized the regression sum of squares (RSS) subject to the constraint $\Sigma_{k=1}^{p}|\beta_k|^2$ <constant, i.e. the L_2 norm of regression coefficients. Frank and Friedman [1993] further refined this constraint to $\Sigma_{k=1}^{p}|\beta_k|^\gamma$ <constant, suggesting that an optimal choice of γ yielded the best subset of predictors. Tibshirani [1996] introduced the Least Absolute Shrinkage and Selection Operator (LASSO) estimation of regression parameters. In this

method, the likelihood function, rather than RSS, is modified by subtracting a norm (proportional to L1 for LASSO) of the regression parameters, similar to the method of Lagrange multipliers. Unimportant predictors have their corresponding regression coefficients effectively 'shrunk' to zero by this method. Extensions of LASSO include adaptive LASSO, Group LASSO, and elastic net (see Casella et al. [2010], C. Leng and Nott [2012], Zhou and Hastie [2005]). Adaptive LASSO was more flexible in that the shrinkage imposed is different for different coefficients. This allows for more flexibility in the selection process. In Group LASSO, a combination of L1 and L2 norms is used in the penalty function, and so categorical variable levels are treated separately. Unlike regular LASSO, Group LASSO can select just certain levels of a categorical variable. In elastic net, a convex combination of norms, L1 and L2, is used to implement the regression penalty, see Zou and Hastie [2005]. Through studies and real data analysis, it was shown that elastic net outperforms LASSO. None of these prior efforts attempt variable selection on zero-inflated data.

Buu et al. [2011] develop variable selection methods for ZIP regression models including LASSO and Smoothly Clipped Absolute Deviation penalty (SCAD). Variable selection for the semi-parametric model, among those covariates that are initially included in the nonparametric portion, would be desirable, as well. Buu et al. [2011] does not apply to ZINBH models, nor does it use a Bayesian approach in its estimation process.

Penalized regression is formulated through a Bayesian approach by using symmetric,

leptokurtic distributions as prior distributions for the regression coefficients (Trevor and Casella [2008] , Yaun and Lin [2005a] and Bae and Mallick [2004]). Bayesian credible intervals from MCMC samples are used to make decisions about predictors. Yaun and Lin [2005b] and Hans [2010] used a mixture distribution, 'zero inflated', consisting at a point mass at zero and Laplace distribution, for priors on coefficients. By this method they were able to arrive at more 'definitive' decisions about coefficients than by Laplace alone. Kuo and Mallick [1998] and Lykou and Ntzoufras [2013] incorporate Bernoullli indicator variables within the priors of the regression coefficients. In this way, objective decisions of variable selection are based upon the posterior distribution of these Bernoulli variables. Again, none of the above literature addresses zero-inflated data.

The present work incorporates several of the above model components into a Bayesian approach in which a Zero-Inflated Negative Binomial Hurdle (ZINBH) regression model is developed for bivariate data with spatial random effects measured on a lattice. Further, both cross-spatial and cross-variable dependencies are included in the covariance structure of our model. There is reason to believe that Bayesian approaches may perform better in data sets that are pertinent to this research. Ghosh et al. [2006] used simulations to show an increased accuracy of Bayesian methods over traditional MLE estimates in distribution parameter estimates for small sample sizes and higher proportions of zero observations. Also, Bayesian methods relax the requirement of non-singularity of the Hessian matrix $X_+^{-1}X_+$, where X_+ represents the rows of the design matrix X which have nonzero response variables. This provides

more flexibility for experimental design, especially in the presence of a sparse Hessian matrix. In order to model the presence of lattice dependence with a Bayesian approach, we first implement a prior distribution of imposed variance-covariance structure on the random effects in the form of an intrinsic conditional auto-regressive (ICAR) Gaussian prior. We perform simulations studies to assess the performance of these approaches. Furthermore, we implement a Two-Level Hierarchical, Inverse Wishart (TLHIW) prior in the scale matrix for the ICAR variances to promote more accurate variance-covariance estimation (see Huang and Wand [2013]). And lastly, we implement the the technique of Kuo and Mallick [1998] and Lykou and Ntzoufras [2013] for variable selection, via elastic net, in a univariate form of our ZINBH model.

Chapter 3

The Modeling Approach

Parameterization

The object of interest in this research is a method of modeling zero-inflated bivariate spatial count data with overdispersion. Two-part regression is applied to each response, consisting of a point mass at zero and the discrete portion modeled as Negative Binomial via Poisson-Gamma hierarchical model (Greene [1994]). The spatial random effects are modeled through a Gaussian Intrinsic Conditional Auto Regressive (ICAR) model (J. Besag and Mollie [1991]).

Multivariate spatial data on a lattice can exhibit a number of dependencies among the responses which must be considered when developing a model:

• Spatial dependence within the same variable from one lattice site to another.

• Cross-variable dependence between different variables within the same lattice site.

• Cross-variable cross-spatial dependence between different variables on different lattice sites.

To develop our model, let $Y_{i,j}$ $(i = 1, 2 : j = 1, \ldots, n)$ be the i^{th} response of a bivariate observation on the j^{th} lattice site, and we assume that there are n sites on the spatial domain under analysis as described below in Figure 3.1.

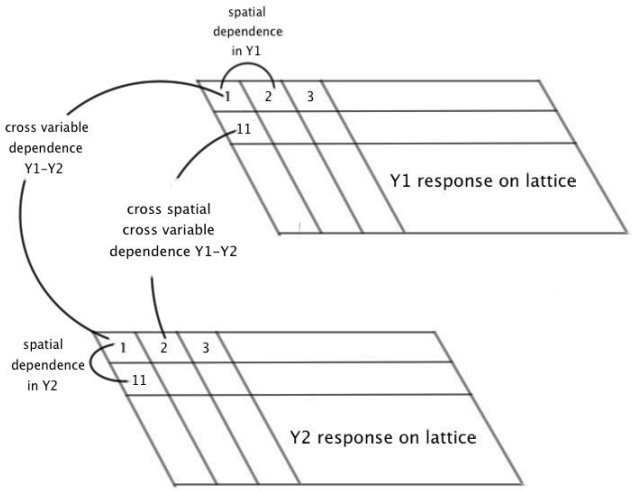


Figure 3.1: Schematic diagram for spatial and cross spatial dependencies of a set of bivariate response on a 2-dimensional lattice

In Figure 3.1, two copies of the same rectangular lattice represent the collections of observed values of the two responses, Y_1 and Y_2, at each lattice site. The curved lines represent the various cross-variable, cross-spatial and cross-spatial cross-variable dependencies described above. The distribution of the zero-inflated response, 3.1

below, is a mixture of two distributions as described in Chapter 1. The first part, the zero, or 'occurrence' portion, consists of the point mass at zero, with probability $\pi_{i,j}$. The second part, the non-zero, or 'abundance' portion, consists of a truncated Negative Binomial distribution ($\geq$1) with mean $\mu_{i,j}$ and dispersion parameter α_{nb}, with mixing probability 1-$\pi_{i,j}$:

$$[Y_{i,j}] = \pi_{i,j} \times [0] + (1 - \pi_{i,j}) \times [V_{i,j}|\mu_{i,j}, \alpha_{nb}]; i = 1, 2; j = 1, \ldots, n \tag{3.1}$$

where $\pi_{i,j}$ is the probability point mass at '0' and $[V_{i,j}|\mu_{i,j}, \alpha_{nb}]$ refers to a truncated Negative Binomial distributed variable $V_{i,j}$, truncated above '1' ($V_{i,j} \geq 1$) and conditioned on the mean $\mu_{i,j}$ and dispersion parameter α_{nb}.

The point mass at zero, the probability, $\pi_{i,j}$, is modeled in terms of explanatory variables and spatial random effects using complementary log-log link function:

$$(\pi_{i,j})_{\text{Transform}} = \log(-\log(1 - \pi_{i,j})) \tag{3.2}$$

The complimentary log-log transformation, in comparison to more traditional probability transforms such as logit or probit, has the advantage of being asymmetric with respect to the q=0.5 axis. Figure 3.2 below, which graphs the value of proportion q (horizontal axis) versus its transformed value (vertical axis), illustrates this asymmetry. This makes this transform more effective when the occurence of zero is either particularly high or low (Williams [2015]).

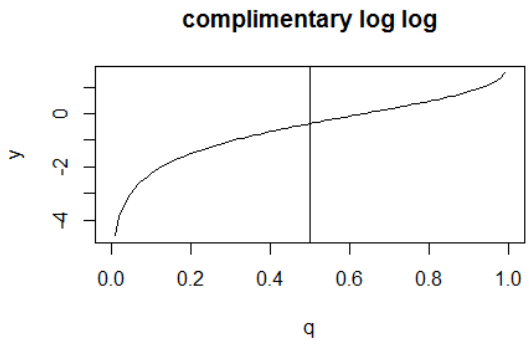


Figure 3.2: Complimentary Log Log Transformation

Using complimentary $\log - \log$ link function, the regression equation for $\pi_{i,j}$ is

$$\log(-\log(1-\pi_{i,j})) = \boldsymbol{x}_j\boldsymbol{\alpha}_i + \psi_{ij}^{(1)}, \tag{3.3}$$

where $\boldsymbol{\alpha}_i$ is the vector of $p \times 1$ regression coefficients, $\boldsymbol{x}_j$ are the set of p covariates for the j^{th} lattice site, and $\boldsymbol{\psi}_i = \{\psi_{i,1}^{(1)}, \psi_{i,2}^{(1)}, \ldots, \psi_{i,n}^{(1)}\}$ are the spatial random effects terms. Spatial and cross spatial dependencies are imposed through these random variables. The regression coefficients $\boldsymbol{\alpha}_{ki}$, k=1,...,p, all have zero mean Gaussian priors with variance σ_α^2. To reduce subjectivity, we make this prior weakly informative by setting σ_α^2 to be 10,000.

To impose a Negative Binomial distribution, we take a Poisson distributed vari-

able and express its mean, $\theta_{i,j}$, as a product of $\mu_{i,j} \times r_{i,j}$. The random variable $r_{i,j}$ is assumed to have a Gamma$(\alpha_{nb}, \alpha_{nb})$) distribution. The log of α_{nb} was uniformly distributed on the interval (a,b) (See section 3.4 for more detail on a,b). Sampling the log of the parameter, and exponentiating that value, ensured that the value of the parameter remained positive during the iteration process. While a uniform distribution on the interval (-∞, ∞) would have been desirable as a non-informative prior of $\log(\alpha_{nb})$, the interval employed was as wide as the WinBUGS software would accept for purposes of sampling. If we integrate out r_{ij}, then the resulting distribution of the response variables become Negative Binomial conditioned on $\mu_{i,j}$ and α_{nb} . Integrating out $r_{i,j}$, we get

$$\begin{aligned}
[V_{i,j}|\mu_{i,j}, \alpha_{nb}] &= \int_{r_{i,j}=0}^{\infty} [V_{i,j}|\mu_{i,j}, r_{i,j}, \alpha_{nb}] dr_{i,j} \\
&= \int_{r_{i,j}=0}^{\infty} \text{Poisson}_{V_{i,j}}(\mu_{i,j} r_{i,j}) \times \text{Gamma}_{r_{i,j}}(\alpha_{nb}, \alpha_{nb}) dr_{i,j} \\
&= \text{Negative Binomial}_{V_{i,j}}(\mu_{i,j}, \alpha_{nb})
\end{aligned} \tag{3.4}$$

This is a Negative Binomial model (Lord and Park [2010]). To implement a hurdle model, we take the truncated version of (3.4) by limiting its support to values ≥ 1.

For the nonzero part, the log of $\mu_{i,j}$ is modeled as a linear function of the covariates and random effects:

$$\log(\mu_{i,j}) = \boldsymbol{x_j \beta_i} + \psi_{ij}^{(2)} \tag{3.5}$$

where $\boldsymbol{\beta_i}$, the regression coefficients, (β_{ki}, k=1,. . . ,p) have zero mean Guassian priors

with variance σ^2_β. Again, we set σ^2_β=10,000 to avoid subjectivity. The random effects terms, $\boldsymbol{\psi}_i^{(2)} = \{\psi_{i,1}^{(2)}, \psi_{i,2}^{(2)}, \ldots, \psi_{i,n}^{(2)}\}$, incorporate both spatial and cross dependencies among other responses.

Gaussian Multivariate intrinsic autoregressive (ICAR) dependence is imposed on the random effects:

$$\boldsymbol{\Psi} = (\boldsymbol{\Psi_1}, \boldsymbol{\Psi_2}, \ldots, \boldsymbol{\Psi_n})' = \begin{pmatrix} \psi_{11}^{(1)} & \psi_{12}^{(1)} & \ldots & \psi_{1n}^{(1)} \\ \psi_{21}^{(1)} & \psi_{22}^{(1)} & \ldots & \psi_{2n}^{(1)} \\ \psi_{11}^{(2)} & \psi_{12}^{(2)} & \ldots & \psi_{1n}^{(2)} \\ \psi_{21}^{(2)} & \psi_{22}^{(2)} & \ldots & \psi_{2n}^{(2)} \end{pmatrix}. \tag{3.6}$$

In this prior, associated with the ICAR process, the random effects at any specific lattice site j, $\boldsymbol{\Psi}_j$, are distributed, conditioned on the neighboring lattice sites, as a multivariate normal with mean equal to the average of the random effects of neighboring lattice sites, $\overline{\boldsymbol{\Psi}}_j$, and variance equal to a diagonal, inverse-Wishart distributed, scaled covariance matrix ,$\frac{\mathbf{Q}}{n_j}$

$$\boldsymbol{\Psi}_j | \boldsymbol{\Psi}_{\omega_j} \sim MVN\left(\overline{\boldsymbol{\Psi}}_j, \frac{\mathbf{Q}}{n_j}\right) \tag{3.7}$$

where

$\overline{\boldsymbol{\Psi}}_j = \left(\overline{\psi}_{1,j}^{(1)}, \overline{\psi}_{1,j}^{(2)}, \overline{\psi}_{2,j}^{(1)}, \overline{\psi}_{2,j}^{(2)}\right)$, $\overline{\psi}_{i,j}^{(m)} = \frac{\sum_{s \in \omega_j} \psi_{i,s}^{(m)}}{n_j}$, ω_j = set of neighbor indices, and

$n_j = \|\omega_j\|$. Furthermore, $\boldsymbol{Q}$ is the 4×4 covariance matrix, with diagonal elements representing the conditional variances of the $\left(\psi_{1,*}^{(1)}, \psi_{1,*}^{(2)}, \psi_{2,*}^{(1)}, \psi_{2,*}^{(2)}\right)$. We impose upon $\boldsymbol{Q}$ a prior distribution of Inverse-Wishart(4,$\boldsymbol{I}_{4\times4}$), as is conventional for multivariate normal covariance structure. This is similar to a Proper CAR distribution, but here $\boldsymbol{Q}$ need only be non-negative definite. It is worth noting here that as $\boldsymbol{Q}$ is non-negative definite, ICAR is not a proper prior. However, the posterior distribution is a legitimate distribution.

Usually, Inverse-Wishart prior is imposed on $\boldsymbol{Q}$ with a pre-specified scale matrix. Here we explore a two-level Inverse-Wishart prior described in Huang and Wand [2013]. This is similar to the previous model with the feature that the matrix $\boldsymbol{Q}$ has an Inverse-Wishart distribution with degrees of freedom $\nu + p - 1$ and 4×4 diagonal square scale matrix $2\nu\times\text{diag}(1/a_1, \ldots 1/a_4)$. The hyperprior distributions for a_k's, k=1,. . . ,4, are independent Inverse-Gamma with shape equal to 1/2 and scale $1/A_k^2$, where ν and A_k are positive scalars. This additional constraint reduces potential bias in the model by making the prior marginally noninformative, as will be explained in the next sections. This bias is related to putting high mass on smaller values in the prior (see Gelman [2006]).

Finally, the likelihood for response $Y_{i,j}$, $\mathcal{L}(Y_{i,j})$, is a function of the probabilities and indicator functions of $Y_{i,j}$ being zero and greater than zero:

$$\mathcal{L}(Y_{i,j}) = (P(Y_{i,j} = 0))^{I(Y_{i,j}=0)}(P(Y_{i,j} = y > 0))^{I(Y_{i,j}\geq1)}. \tag{3.8}$$

Substituting model parameters of the Truncated Negative Binomial $(\mu_{i,j}, \alpha_{nb})$ and Bernoulli$(\pi_{i,j})$ distribuions into (3.9), the expression becomes:

$$\mathcal{L}(Y_{i,j}) = [\pi_{i,j}]^{I(Y_{i,j}=0)} \times \left[(1-\pi_{i,j}) \frac{\frac{\Gamma(y_{i,j}+\alpha_{nb})}{\Gamma(y_{i,j}+1)\Gamma(\alpha_{nb})} \left(\frac{\alpha_{nb}}{\mu_{i,j}+\alpha_{nb}}\right)^{\alpha_{nb}} \left(\frac{\mu_{i,j}}{\mu_{i,j}+\alpha_{nb}}\right)^{y_{i,j}}}{1-\left(\frac{\alpha_{nb}}{\mu_{i,j}+\alpha_{nb}}\right)^{\alpha_{nb}}} \right]^{I(Y_{i,j}\geq 1)}. \tag{3.9}$$

Then, for a set of responses $\{Y_{i,j}\}$ i=1,2; j=1,..,n, the joint data likelihood is the product of the individual responses,

$$\prod_{i=1}^{i=2} \prod_{j=1}^{j=n} \mathcal{L}(Y_{i,j}). \tag{3.10}$$

3.1 Weakly Informative Prior for Variance Covariance Matrices

A well known issue in Bayesian analysis is the effect of the presence of information, or bias, in the prior distribution and its associated impact upon the posterior inference. In the scenario where real prior knowledge is present, due to, for example, a previous study, or a known physical limitation, it is advantageous to have this information reflected in the prior distribution by weighting certain possible values heavier than others. This ability to incorporate extra information into the estimator is, in fact, among the advantages of Bayesian analysis. However, in the scenario when there is no real prior knowledge of a parameter, it is of great importance to avoid imparting any bias through the specified prior distribution which might, as a result, impart bias to the posterior distribution. In this case, non-informative prior distributions, those which don't favor any particular possible value, or set of possible values, of the parameter are sought. This pursuit can lead to improper priors or non-conjugate priors. In some cases, the compromise is to settle for a 'weakly informative' prior, which is 'almost' flat within the region of possible values.

In pursuit of prior noninformativity, for example in hierarchical modeling of a normal distribution in Bayesian analysis, Gelman(2006) advised against the conventional conjugate prior distribution of inverse Gamma for standard deviation parameters in favor of other distributions, such as the non-centralized, or Half -t, with scale A and degrees of freedom ν having pdf:

$$p(\sigma_\alpha) \propto \left(1 + \frac{1}{\nu}\left(\frac{\sigma_\alpha}{A}\right)^2\right)^{-\frac{(\nu+1)}{2}} . \tag{3.11}$$

These priors are less sensitive to hyper-parameters and still retain conditional conjugacy.

For q-dimensional multivariate normal distributions, covariance between responses within the same observation can exist, and the overall variance-covariance parameter is then a square matrix, such as

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \dots & \rho_{1q}\sigma_1\sigma_q \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \dots & \vdots \\ \vdots & \dots & \ddots & \vdots \\ \rho_{q1}\sigma_q\sigma_1 & \dots & & \sigma_q^2 \end{bmatrix} . \tag{3.12}$$

In Bayesian analysis of a q-dimensional multivariate normal response with mean θ and covariance matrix Σ, $MVN_q(\theta, \mathbf{\Sigma})$, the extension of an inverse gamma distribution to a multivariate structure, as a prior distribution for variance-covariance matrix $\mathbf{\Sigma}$, is the inverse Wishart distribution with degrees of freedom ν and $q \times q$ scale matrix $\mathbf{\Psi}$, $\mathbf{W_q^{-1}}(\nu, \mathbf{\Psi})$,

$$p(\mathbf{\Sigma}) \propto |\mathbf{\Psi}|^{\frac{\nu}{2}} |\mathbf{\Sigma}|^{-\frac{\nu+\mathbf{p}+\mathbf{1}}{2}} \exp\{-\mathbf{tr}(\mathbf{\Psi}\mathbf{\Sigma}^{-\mathbf{1}}\} \tag{3.13}$$

where $\nu > 0$, $\mathbf{\Sigma}$, $\mathbf{\Psi}$ both are positive definite.

M. P. Wand and Fruhwirth [2011] showed that the desired Half t distribution arises as a scaled mixture of Inverse Gamma distributions, namely,

(i) If σ^2 conditioned on α is Inverse Gamma distributed with shape $\frac{\nu}{2}$ and scale $\frac{\nu}{\alpha}$

$$p(\sigma^2|\alpha) \propto (\sigma^2)^{-\frac{\nu}{2}-1} e^{-\frac{\nu}{\alpha}/\sigma^2}, \text{ and} \tag{3.14}$$

(ii) α is Inverse Gamma distributed with shape $\frac{1}{2}$ and scale $\frac{1}{A^2}$,

then σ is distributed Half-t with scale A and degrees of freedom ν.

Extending this to q×q random matrices Σ, Huang and Wand [2013] proposed the family of prior distributions where the matrix Σ, conditioned on a set of random variables $(a_1, a_2, \ldots, a_q)$, where each a_i has a Gamma distribution with shape of $\frac{1}{2}$ and scale A_i^2, is Inverse-Wishart distributed with degrees of freedom $\nu + q - 1$ and scale matrix 2ν diag $\left(\frac{1}{a_1}, \frac{1}{a_2}, \ldots, \frac{1}{a_q}\right)$,

$$p(\Sigma) \propto |\Sigma|^{-\frac{\nu+2q}{2}} \prod_{k=1}^{q} \left\{ \nu(\Sigma^{-1})_{kk} + \frac{1}{A_k^2} \right\}^{-\frac{\nu+q}{2}}, \Sigma \text{ is postive definite.} \tag{3.15}$$

Huang and Wand [2013] establish that a variance-covariance matrix Σ with prior distribution as above has the following properties:

Property 1: The marginal distribution of any sub-covariance matrix in Σ has the same distributional form as Σ itself.

Property 2: The marginal distribution of any standard deviation term σ_k in Σ is Half-t(ν, A_k).

Property 3: The marginal distribution of any correlation parameter term ρ_{ij} in Σ has density

$$p(\rho_{ij}) \propto (1 - \rho_{ij}^2)^{\frac{\nu}{2}-1}, -1 < \rho_{ij} < 1. \tag{3.16}$$

Property 4: For the particular choice ν=2, the marginal distribution of any correlation parameter term ρ_{ij} in Σ is uniform on (-1,1).

As described in Figure 1 above, bivariate regression on a lattice contains numerous possible dependencies, or correlations, and because of the two-part nature of zero-inflated model , these correlations are considerably complex. By employing a Two-Level Hierarchical Inverse Wishart(TLHIW) matrix, as described above, as a prior in the spatial random effects distribution, we take advantage of property 4 and

protect against systematic bias towards any particular correlation values, making our spatial random effects model more flexible.

3.2 Model Comparison

In this section we will compare the performances between the Negative Binomial Hurdle spatial models stated in the previous section. In addition, the 'standard' versions of the Negative Binomial spatial model, with no zero inflation, are also compared to check the appropriateness of zero-inflation.

To compare the performance of the models, data were simulated as follows:

1. Random spatial effects were simulated for a a rectangular 25×40 (n=1000) lattice using spatial effects determined by conditional auto regressive covariance with "neighbor" sites defined as sharing a lattice edge. The effects are generated as a mean zero 2ni dimensional multivariate random variable: Φ_{2ni} with covariance matrix

$$\mathbf{Q}_{2ni\times 2ni} = \tau^2(\mathbf{I} - \rho\mathbf{A})^{-1}\mathbf{M} \tag{3.17}$$

where

i : dimension of the multivariate response (=2 for bivariate response), and the "2" represents one part for the occurrence portion of the response, one part for the abundance portion of the response.

τ : standard deviation of the random spatial effects (0.01).

ρ : spatial dependence between lattice sites(set to 0.1 to impose low spatial dependence).

$\mathbf{A}$: weighted neighborhood matrix such that sum of each column is equal to unity.

$\mathbf{M}$: diagonal matrix where M_{ii} is the number of neighbors of lattice j.

These effects contained four separate components, one for each of the Bernoulli and truncated Negative Binomial portions of bivariate zero-inflated data.

2. Covariates are generated as:

1) Normally distributed with mean zero and standard deviation 0.2.

2) Uniformly distributed on the interval (-1,1).

3) Beta distributed with shape parameters $\alpha = \beta$=2.

3. Ten regression coefficients are set, one for each of the covariates of the truncated negative binomial regression, and one for the intercept of the binomial regression. This gives a total of five coefficients for each of the two responses. All coefficients were set to generate both a high ($\sim$ 0.45-60) and low ($\sim$ 0.15-0.30) proportion of zero observations for each of the responses.

4. Using random spatial effects, along with the covariates and regression coefficients described above, we generate bivariate zero-inflated Negative Binomial Hurdle responses for each of the 1000 lattice sites.

Due to the complexity of the model, including non-linearity and depth of hyperpriors, the algorithm takes considerable time to converge into a stable orbit around this distribution space of the estimated parameters, and hence the large burn-in is required. This 'slow mixing' can be seen in the trace plots in Appendix. While MacEachern and Berliner [1994] show that keeping the entire Markov Chain always leads to more precise posterior estimates than thinning, the latter was employed in bivariate simulations due to memory issues with WinBUGS. In fact, to make possible the posterior prediction checks below, more than 1000 values had to be stored for each MCMC iteration. For this simulation 20,000 MCMC samples were used with the initial 10000 discarded as burn-in to allow the algorithm to converge.

The simulated data were analyzed using the three regression models described in the previous section. The two zero-inflated models are compared to each other via regression estimates, while both of these are compared to the non zero-inflated models via posterior predictive plots.

From the preliminary results shown below in Table 3.1, the univariate models appear to be performing close to the same for both higher and lower proportions of

zero observations, with the fixed scale matrix model doing slightly better. In both univariate simulations the constant scale matrix model captures 83.3% (5 out of 6) of the regression coefficients versus 66.7% (4 out of 6) for the TLHIW scale matrix. Note that the latter model 'barely' misses capturing the β_{14} true value in the 95% credible interval in both simulation.

For the bivariate models in table 3.2, the constant scale matrix model clearly does better than the TLHIW scale matrix model, in both lower and higher proportion of zero observations. Again, in the majority of instances where the latter model fails to capture the true value, it does by a very small margin. When the magnitude of the spatial dependence ρ is increased, as in table 3.3, it appears that both models do poorly if the signs of the dependencies alter. This altering of signs could represent a scenario when the two responses depend on neighboring values is opposite ways. The combined effect of increasing dependence and altering of the signs of the spatial dependence may be introducing too much noise with the signal, making it difficult for the models to distinguish between the two. Note that when the signs of the dependence are all positive, both models perform similarly, with most 95% credible intervals relatively close in width and location.

Table 3.1: Univariate Simulation

SIMULATION 1 UNIVARIATE 95 % credible interval ("+" = capture of true parameter)			
Y1:% zeros = 28.1			
variable	true	TLHIW	Fixed Scale Matrix
$\alpha(1)$	0.25	$(0.0624, 0.2384)$	$(0.1515, 0.2914)+$
$\beta(1,1)$	0.5	$(0.2488, 0.9356)+$	$(0.1803, 0.6240)+$
$\beta(1,2)$	2.00	$(1.640, 2.023)+$	$(1.821, 2.190)+$
$\beta(1,3)$	-4.00	$(-4.05, -3.83)+$	$(-4.127, -3.894)+$
$\beta(1,4)$	3.00	$(2.494, 2.809)$	$(2.570, 2.921)$
$\alpha_n b$	3.00	$(0.5563, 18.90)+$	$(2.913, 4.088)+$

SIMULATION 2 UNIVARIATE 95 % credible interval ("+" = capture of true parameter)			
Y1:% zeros = 46.1			
variable	true	TLHIW	Fixed Scale Matrix
$\alpha(1)$	-0.35	$(-0.4238, -0.2261)+$	$(-0.3606, -0.1772)+$
$\beta(1,1)$	0.5	$(-0.6172, -0.09554)$	$(0.3169, 0.5855)+$
$\beta(1,2)$	2.00	$(1.832, 2.341)+$	$(2.209, 2.602)$
$\beta(1,3)$	-4.00	$(-4.038, -3.679)+$	$(-4.147, -3.859)+$
$\beta(1,4)$	3.00	$(2.601, 2.983)$	$(2.912, 3.287)+$
$\alpha_n b$	3.00	$(0.481, 7.022)+$	$(2.783, 4.022)+$

3.3 Posterior Predictive Check

There has been some discussion as to whether the Zero-inflated Negative Binomial model is really ever needed (see Allison [2012a]). Because of the ability to model high probability of zero as well as large non-zero values, via overdispersion, in the Negative Binomial model, the zero-inflated case is thought to be unnecessary, as 'suitable' fit should be achievable without a separate occurrence regression. Obviously, ZINBH models evolve naturally from situations where occurrence is influenced

Table 3.2: Multivariate Simulation

SIMULATION 1 MULTIVARIATE 95 % credible interval: "+" = capture of true parameter)			
Y1:% zeros = 28.6			
variable	true	TLHIW	Fixed Scale Matrix
$\alpha(1)$	0.25	$(0.1054, 0.2856)+$	$(0.1364, 0.3067)+$
$\beta(1,1)$	0.5	$(0.2796, 0.6462)+$	$(0.3585, 0.613)+$
$\beta(1,2)$	2.00	$(1.624, 2.251)+$	$(1.547, 2.204)+$
$\beta(1,3)$	-4.00	$(-4.169, -3.953)+$	$(-4.190, -3.952)+$
$\beta(1,4)$	3.00	$(2.662, 3.216)+$	$(2.815, 3.196)+$
$rnb1$	3.00	$(3.480, 5.149)$	$(2.978, 4.211)+$
Y_2: % zeros = 28.1			
$\alpha(2)$	0.15	$(0.1197, 0.2947)+$	$(.1445, 0.3118)+$
$\beta(2,1)$	0.27	$(-0.1745, 0.2379)$	$(0.03831, 0.31590)+$
$\beta(2,2)$	-3.00	$(-3.508, -2.673)+$	$(-3.012, -2.459)+$
$\beta(2,3)$	4.00	$(4.028, 4.463)$	$(3.792, 4.180)+$
$\beta(2,4)$	-2.00	$(-2.379, -1.913)+$	$(-2.078, -1.591)+$
$rnb2$	5.00	$(5.517, 18.890)$	$(4.584, 9.728)+$

SIMULATION 2 MULTIVARIATE 95 % credible interval:"+" = capture of true parameter)			
Y1:% zeros = 46.1			
variable	true	TLHIW	Fixed Scale Matrix
$\alpha(1)$	-0.35	$(-0.4843, -0.2989)+$	$(-0.3533, -0.1755)+$
$\beta(1,1)$	0.5	$(0.2746, 0.4959)$	$(0.4321, 0.7642)+$
$\beta(1,2)$	2.00	$(2.043, 2.622)$	$(2.075, 2.573)$
$\beta(1,3)$	-4.00	$(-4.056, -3.826)+$	$(-4.054, -3.757)+$
$\beta(1,4)$	3.00	$(2.827, 3.133)+$	$(2.615, 3.062)+$
$rnb1$	3.00	$(3.691, 5.855)$	$(2.824, 4.064)+$
Y_2: % zeros = 52.3			
$\alpha(2)$	-0.50	$(-0.6816 - 0.4831)+$	$(-0.5324, -0.3579)+$
$\beta(2,1)$	0.27	$(-0.39030, 0.08882)$	$(0.0852, 0.4976)+$
$\beta(2,2)$	-3.00	$(-3.414, -2.760)+$	$(-3.101, -2.471)+$
$\beta(2,3)$	4.00	$(4.169, 4.792)$	$(3.803, 4.260)+$
$\beta(2,4)$	-2.00	$(-2.484, -1.470)+$	$(-2.454, -1.812)+$
$rnb2$	5.00	$(4.398, 14.600)+$	$(3.204, 7.021)+$

Table 3.3: Multivariate Simulation with Altered Spatial Dependence

SIMULATION 3 MULTIVARIATE (new spatial correlation) $\rho_{11} = 0.5; \rho_{12} = -0.5; \rho_{21} = 0.5; \rho_{21} = -0.4$ 95 % credible interval: "+" = capture of true parameter)			
Y1:% zeros = 48.9			
variable	true	TLHIW	Fixed Scale Matrix
$\alpha(1)$	-0.35	$(-0.4913, -0.3256)+$	$(-0.4489, -0.2678)+$
$\beta(1,1)$	0.5	$(1.070, 1.296)$	$(0.9763, 1.3350)$
$\beta(1,2)$	2.00	$(2.061, 3.115)$	$(1.953, 2.805)+$
$\beta(1,3)$	-4.00	$(-4.342, -4.025)+$	$(-4.082, -3.826)+$
$\beta(1,4)$	3.00	$(2.282, 2.673)$	$(2.336, 3.010)+$
$rnb1$	3.00	$(0.8152, 1.1290)$	$(2.117, 2.162)$
Y_2: % zeros = 57.0			
$\alpha(2)$	-0.50	$(-0.7830, -0.5559)$	$(-0.7078, -0.5029)$
$\beta(2,1)$	0.27	$(0.2439, 0.9335)+$	$(0.6342, 1.1630)$
$\beta(2,2)$	-3.00	$(-2.787, -1.908)$	$(-3.137, -2.189)+$
$\beta(2,3)$	4.00	$(3.363, 4.169)+$	$(3.371, 3.787)$
$\beta(2,4)$	-2.00	$(-1.731, -1.154)$	$(-2.325, -1.454)+$
$rnb2$	5.00	$(0.5505, 0.9350)$	$(2.118, 2.240)$

SIMULATION 4 MULTIVARIATE (new spatial correlatoin) $\rho_{11} = 0.5; \rho_{12} = 0.5; \rho_{21} = 0.4; \rho_{21} = 0.4$			
Y1:% zeros = 43			
variable	true	TLHIW	Fixed Scale Matrix
$\alpha(1)$	-0.35	$(-0.3054, -0.0894)$	$(-0.2603, -0.1075)$
$\beta(1,1)$	0.5	$(0.2949, 0.5067)+$	$(0.2921, 0.6307)+$
$\beta(1,2)$	2.00	$(1.454, 2.164)+$	$(1.592, 2.118)+$
$\beta(1,3)$	-4.00	$(-4.027, -3.771)+$	$(-4.010, -3.759)+$
$\beta(1,4)$	3.00	$(2.831, 3.190)+$	$(2.753, 3.304)+$
$rnb1$	3.00	$(2.957, 4.374)+$	$(2.410, 3.417)+$
Y_2: % zeros = 46.8			
$\alpha(2)$	-0.50	$(-0.4127, -0.2502)$	$(-0.3609, -0.2018)$
$\beta(2,1)$	0.27	$(-0.1864, 0.1433)$	$(-0.0740, 0.2572)$
$\beta(2,2)$	-3.00	$(-3.390, -2.663)+$	$(-3.553, -2.813)+$
$\beta(2,3)$	4.00	$(3.816, 4.268)+$	$(3.732, 4.231)+$
$\beta(2,4)$	-2.00	$(-2.170, -1.411)+$	$(-1.815, -1.372)$
$rnb2$	5.00	$(8.452, 19.750)$	$(3.862, 8.420)+$

by factors which preclude abundance a priori. For example, a survey of people about how often they attended a fitness center the previous month. Responses would predictably include zero's from subjects that did not belong to a fitness club, being precluded (theoretically) a priori from a non-zero response, as well as those that had access but did not avail themselves that particular month. In this section, we will compare the performance of the ZINBH models described previously with the traditional NB model with no zero-inflation. For comparing the zero-inflated models and nonzero-inflated model, performance will be assessed through posterior predictive distributions of the responses. That is, based upon the estimates of regression coefficients and the random spatial effects, we will generate random samples from the modeled distributions of the responses and compare those random samples against the actual data. Because of the large sample size, it is not prudent to compare predictive distributions for every data point. Therefore, we randomly sample 20 responses from each of the two (bivariate) collection of responses, making sure to select 10 zero and 10 non-zero data points from each. In each plot, the actual data point to which the predictive distribution is associated is marked for comparison. In some cases, the data point does not sit on the software chosen horizontal axis.

In general, it is important to understand the complexity of the problem we attempt to address. Like any regression, outliers can possibly 'mislead' the model, as two, what appear to be similar, sets of covariates lead to vastly different responses. In zero-inflated regression, this issue is compounded by the presence of the extra weight at zero. When the mean of the abundance portion of the distribution is far away

from zero, the variation of the observed responses can be vast. In our model, we assume a different probability of zero as well as a different abundance mean at every lattice site. One observation per lattice site is equivalent to a single realization of an unknown distribution. This approach lends itself well to MCMC methods, where many different possible realizations of each unknown distribution are generated with the exploration of the sample space tending to stay away from the less 'likely' realms of values.

All three models do a fair job with zero observations, as a rule. A 'fair' job for zero-inflated model is when a significant probability of zero accompanies an actual zero data point, regardless of the mean of the abundance distribution. A very small mean, and hence many posteriorly predicted zero observations, is judged a 'fair' job for the traditional Negative Binomial model. Both of these are seen in most of randomly chosen zero observations from the actual data. See plots for variable 1, Y_1, at lattice sites 14 below and 54, 73, 94, 97, 203, 330, 420, 744 and 890 in Appendix. Similarly, see plots for variable 2, Y_2 at lattice sites 796 below and 240, 305, 477, 688, 712, 773, 877 and 992 in Appendix.

For non-zero observations, a 'fair' job for the zero-inflated models is when the actual data point is located somewhere near the mean of the abundance histogram, regardless of the weight at zero. For the traditional Negative Binomial, a larger positive tailed histogram, i.e. a more right skewed posterior plot, which overlaps the actual data point, is a sign of a 'fair' job. All three models do a fair job for Y_1 at lattice

sites 156 and 528 (Y_1=1) and Y_2 at site 100 below and 72, 141, 312, 530, 625, and 971 in Appendix (Y_2=1). For non-zero observations greater than 1, the advantage of the zero-inflated distribution becomes more apparent. For example, see for Y_1 at sites 242 and 590 below. See also, for Y_2 at site 954 below and 593 in Appendix. For some sites, the observation is in the tail of the distributions and none of the models do well (see Y_1 at sites 38, 60 and 735 in Appendix).

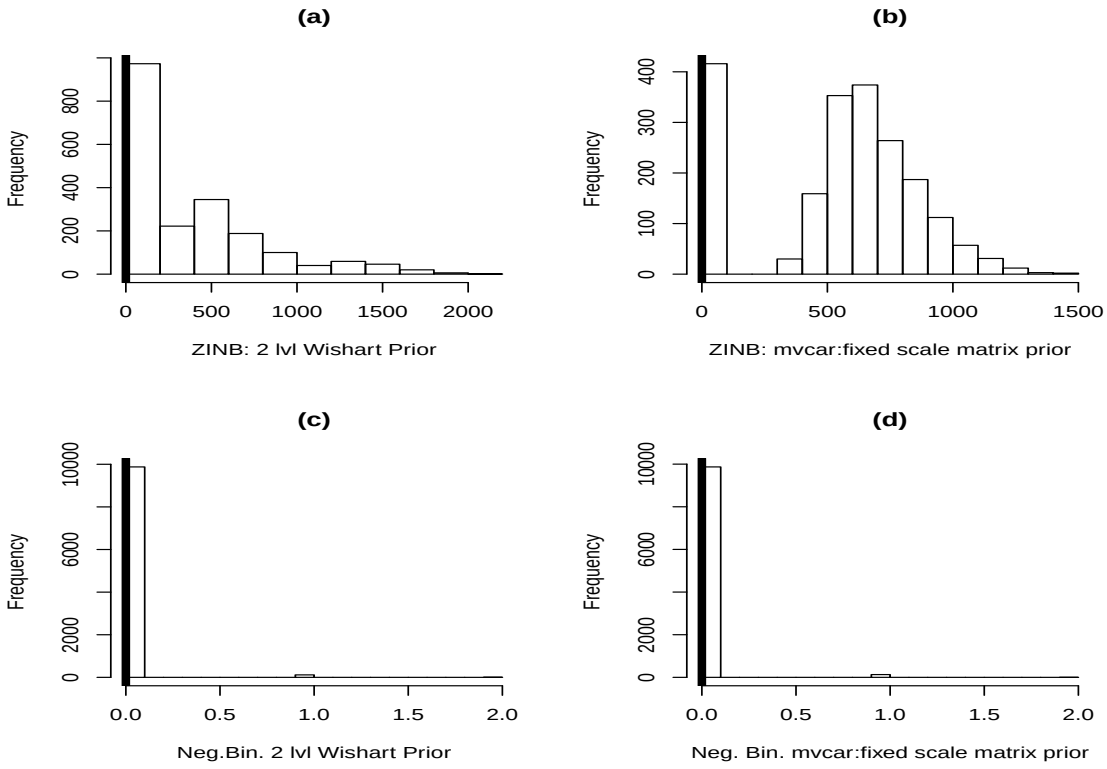


Figure 3.3: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 14, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 14, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.

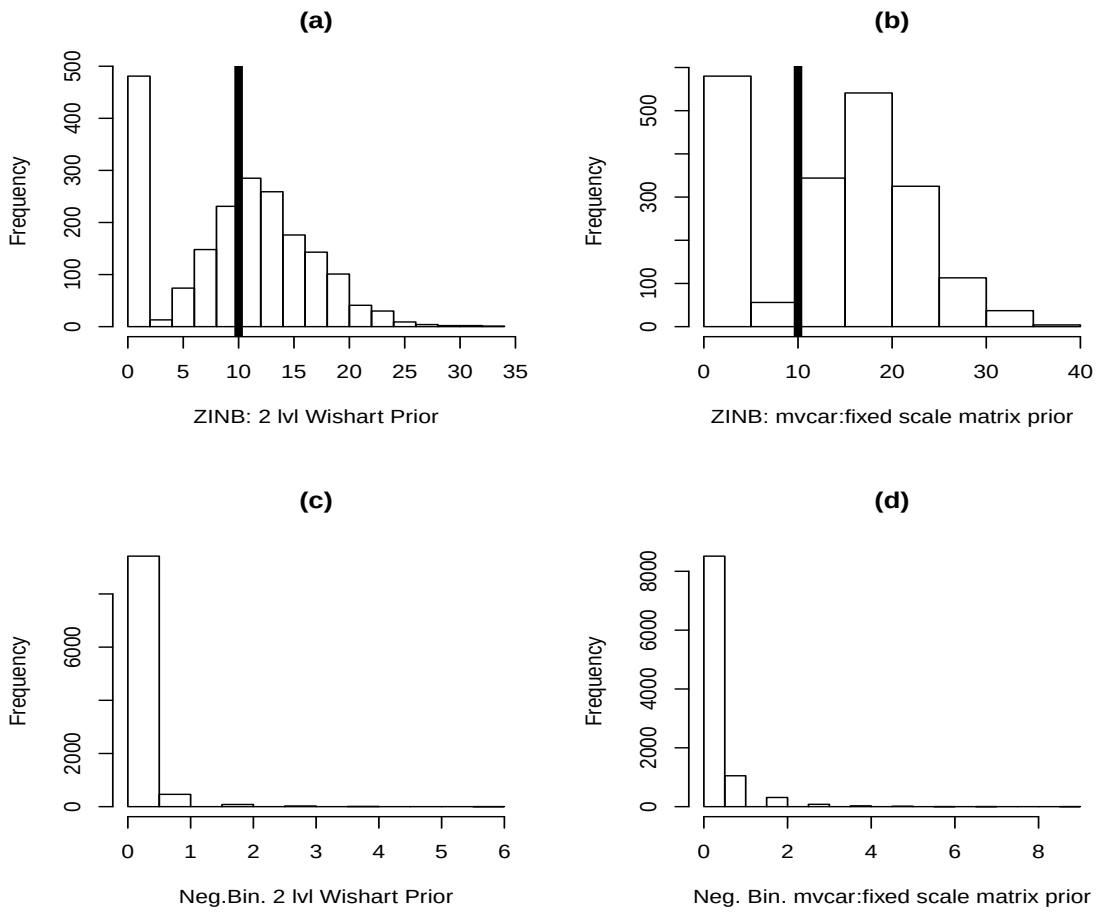


Figure 3.4: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 242, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y1 for ZINBH and Negative Binomial at site 14, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.

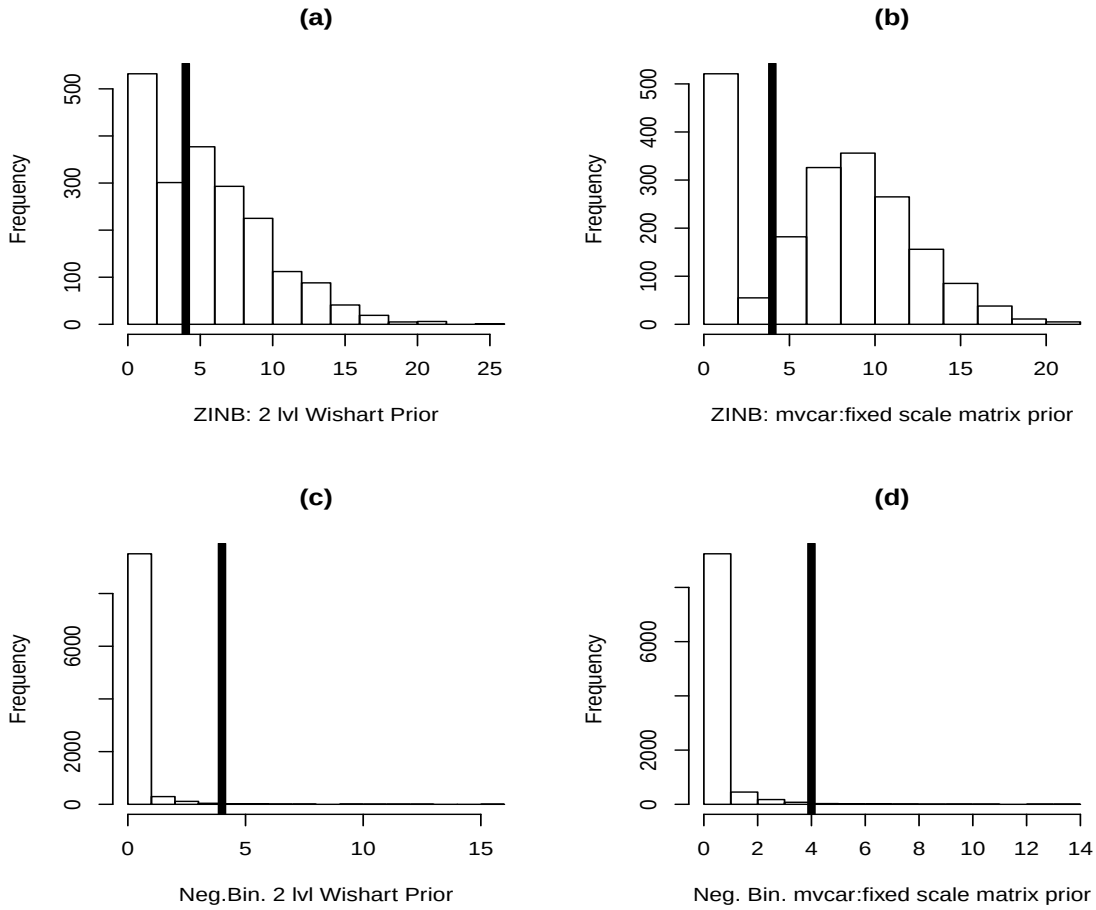


Figure 3.5: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 590, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 14, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.

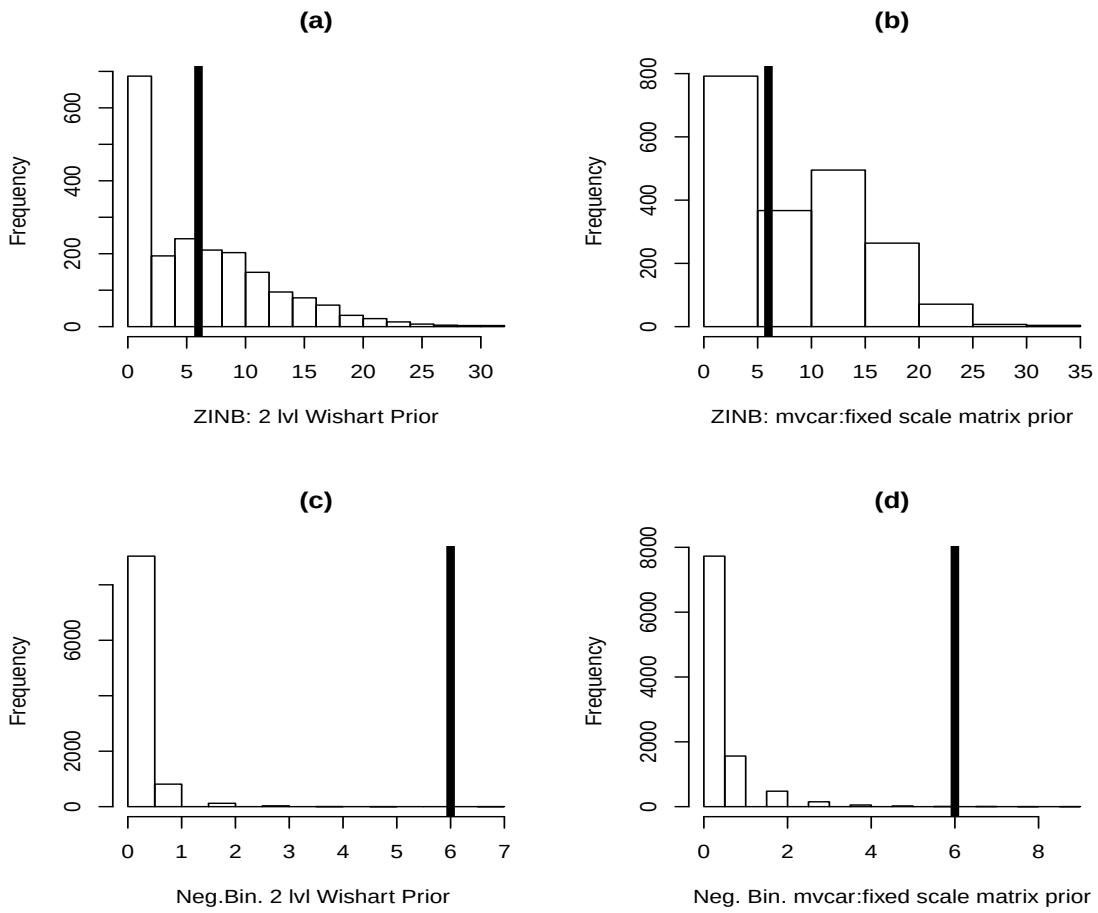


Figure 3.6: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 954, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 954, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.

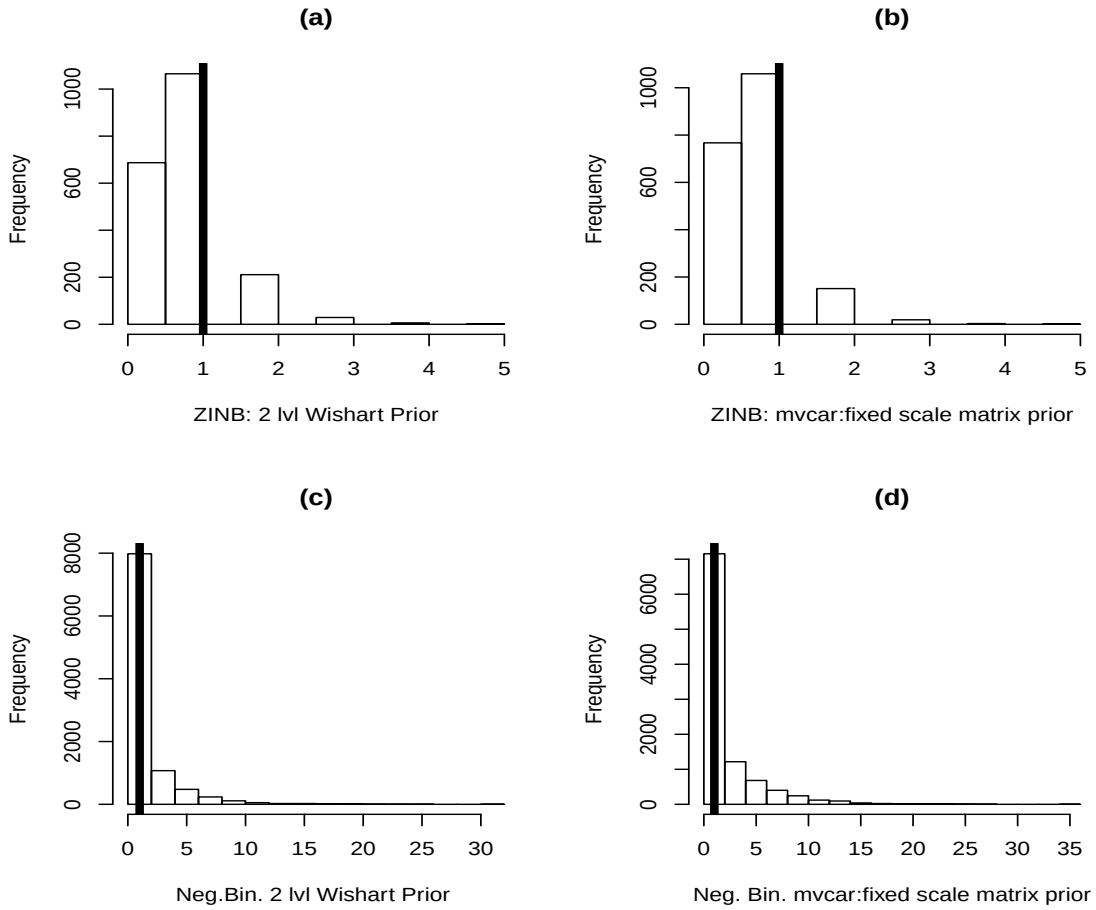


Figure 3.7: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 100, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 100, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.

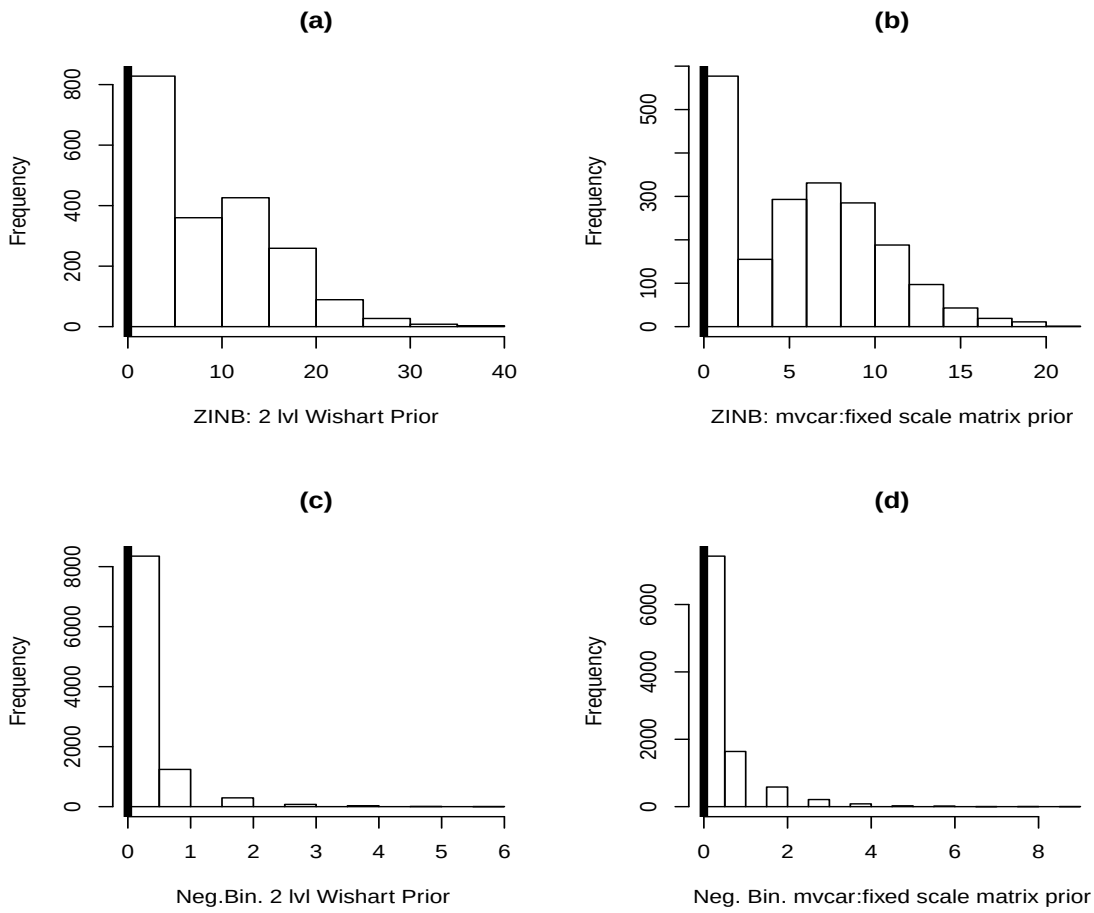


Figure 3.8: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 796, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 796, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.

3.4 Real Data Example: Michigan Crime Counts by County 2011

To test the TLHIW prior and fixed scale matrix models, we apply them both to Michigan crime data (number of incidences of a particular crime) at the county level (source:http://www.michigan.gov/msp). The bivariate response variables are county level counts on **murder or non-negligent manslaughter** and **burglary forced entry**. Table 3.4 and Figure 3.9, which describe these two responses, reflect moderate zero-inflation (38.6 % and 37.3%, respectively) and presence of overdispersion.

Table 3.4: Summary of MI Crime Responses by County

Crime	**Min**	2^{nd} quartile	**Median**	**Mean**	3^{rd} quartile	**Max**
(Y_1)Murder or Non-negligent Manslaughter	0.00	0.00	1.00	6.988	2.00	345.00
(Y_2)Burglary with Forced Entry	0.00	0.00	1.00	12.95	3.00	524.00

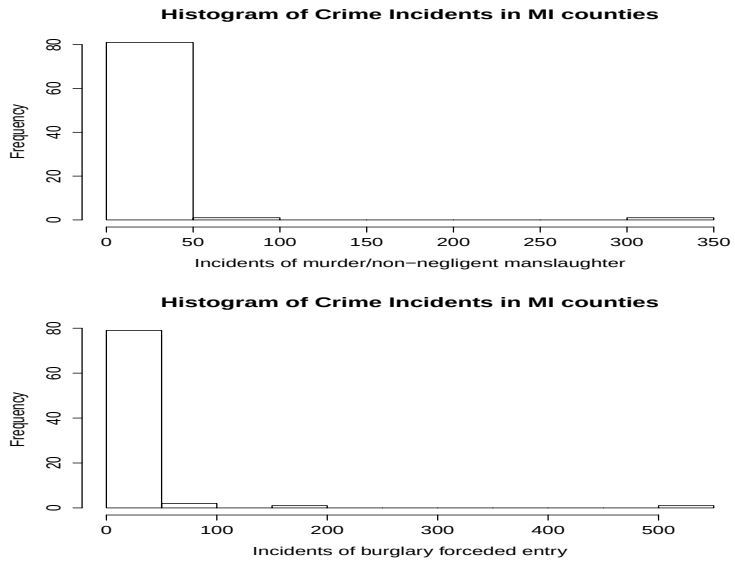


Figure 3.9: Histogram of Crime Incidences in MI Counties

The following covariates were considered for this study:

•X_1: % unemployment rate as of February 2011

•X_2: change in % unemployment over the prior 12 months

(source:http://data.bls.gov/map/MapToolServlet)

•X_3: % High School Graduation Rate as of 2011

(source:http://www.countyhealthrankings.org/app/michigan/2011

/measure/factors/21/data)

•X_4: Median Income in $10000's, Averaged over 2010-2013

(source:https://en.wikipedia.org/wiki/Michigan_locations_by_per_capita_income)

•X_5: Log of Estimated population as of 2011

(source:http://www.michigan.gov/cgi/

0,4548,7-158-54534_51713_51716-325583–,00.html)

In addition, an intercept was also added to the model.

The distribution of the covariates are reflected in Fig. 10 below. All are unimodal and most of them are generally symmetric. Skewness exists moderately in % Unemployment and Median Income. The log transformation of Population yielded a symmetry that does not tend to dominate in influence among the covariates.

Another consideration is multicollinearity among these covariates, which could significantly affect the accuracy of the regression estimates. For instance, 'change in % Unemployment over the prior 12 months' and the '% Unemployment in February' might very well track each other and pull the model towards spurious estimates. To check for the presence of multicollinearity, a correlation matrix for the covariates was generated as well as variance inflation factors (VIF) for each of the covariates. Table 4 shows some moderate to significant correlations, such as '% Unemployment in Feb.' with 'Median Income' and 'Log of Population'. The former is somewhat expected, as the latter may reflect higher employment rates in counties with larger cities. The highest is between 'Median Income' and 'Log of Population', which, again, follows conventional wisdom. However, there were no unexpectedly high correlations between 'suspect' pairs. Furthermore, all of the VIF are below 2.5, which is a conventional threshold for suspected multicollinearity (Allison [2012b]).

To assess the fit and confidence in the regression estimates of these two models,

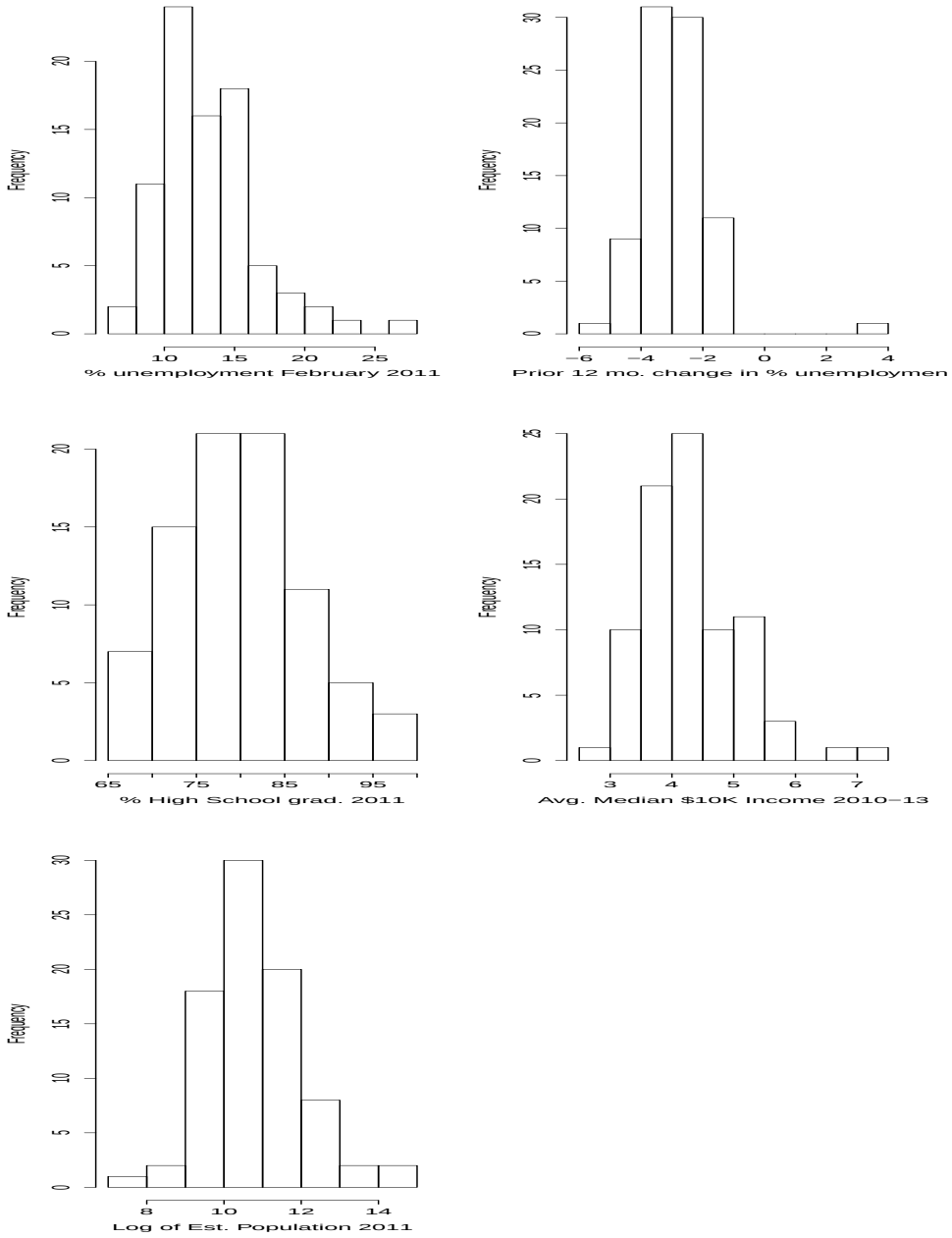


Figure 3.10: Histograms of MI Crime CovariatesX_1, X_2, X_3, X_4, X_5

Table 3.5: Correlation Matrix and VIF for MI Crime Covariates

	% % Unemploy. in Feb.	Change in % Unemploy.	H.S. Grad. Rate %	Median Income	Log of Pop.
% Unemploy. in Feb.	1.00	-0.091	-0.047	-0.507	-0.594
Change in % Unemploy.		1.00	0.220	-0.015	-0.184
H.S. Grad. Rate			1.00	0.310	-0.101
Median Income				1.00	0.606
Log of Pop.					1.00
VIF	1.726	1.143	1.331	2.101	2.322

both of which employ MCMC methods, we will look again at posterior predictive plots of the responses. To do this, we proceed as before and randomly select 3-5 each of zero and non-zero responses from the data and examine where the actual data points locate within the histograms of the collection of predicted responses based on estimates.

For this example the prior distributions are the same as those used for the analysis of the simulated data. $\boldsymbol{\alpha}$, the regression coefficients for the occurrence portion of the two part regression, and $\boldsymbol{\beta}$, the coefficients for the non-zero part, were both normally distributed with mean 0 and precision 0.0001 (variance = 10000). The log of α_{nb} was uniformly distributed on the interval (-0.8,3). Sampling the log of the parameter, and exponentiating that value, ensured that the value of the parameter remained positive during the iteration process. While a uniform distribution on the

interval (-∞, ∞) would have been desirable as a non-informative prior of $\log(\alpha_{nb})$, the interval employed was as wide as the WinBUGS software would accept for purposes of sampling. For the MCMC simulation, 20000 iterations were performed, discarding the first 10000 as burn in, and then thinning by a factor of five (nthin=5).

The posterior predictive plots for some non-zero observations are shown in figure 3.11 and 3.12 below (and for sites 4,8,19,12,81 in Appendix). In the situation of a zero observation, both models do well (see for site 56 in Appendix). This was the case for simulated data, as well. In the situation of non zero observations in actual data, both models perform about the same, with the TLHIW prior model performing slightly better in some cases. For example, at site 33 in Figure 3.11 (a) and (b), for Murder and Non-negligent Manslaughter, the height of the histogram at the observed value is higher for the TLHIW prior model than that of the fixed scale matrix prior model . Similarly, this true for the other histograms in figures 3.11 and 3.12. Based upon the performance at the posterior predictive plots for both responses, below and in the Appendix, confidence in the estimates is warranted.

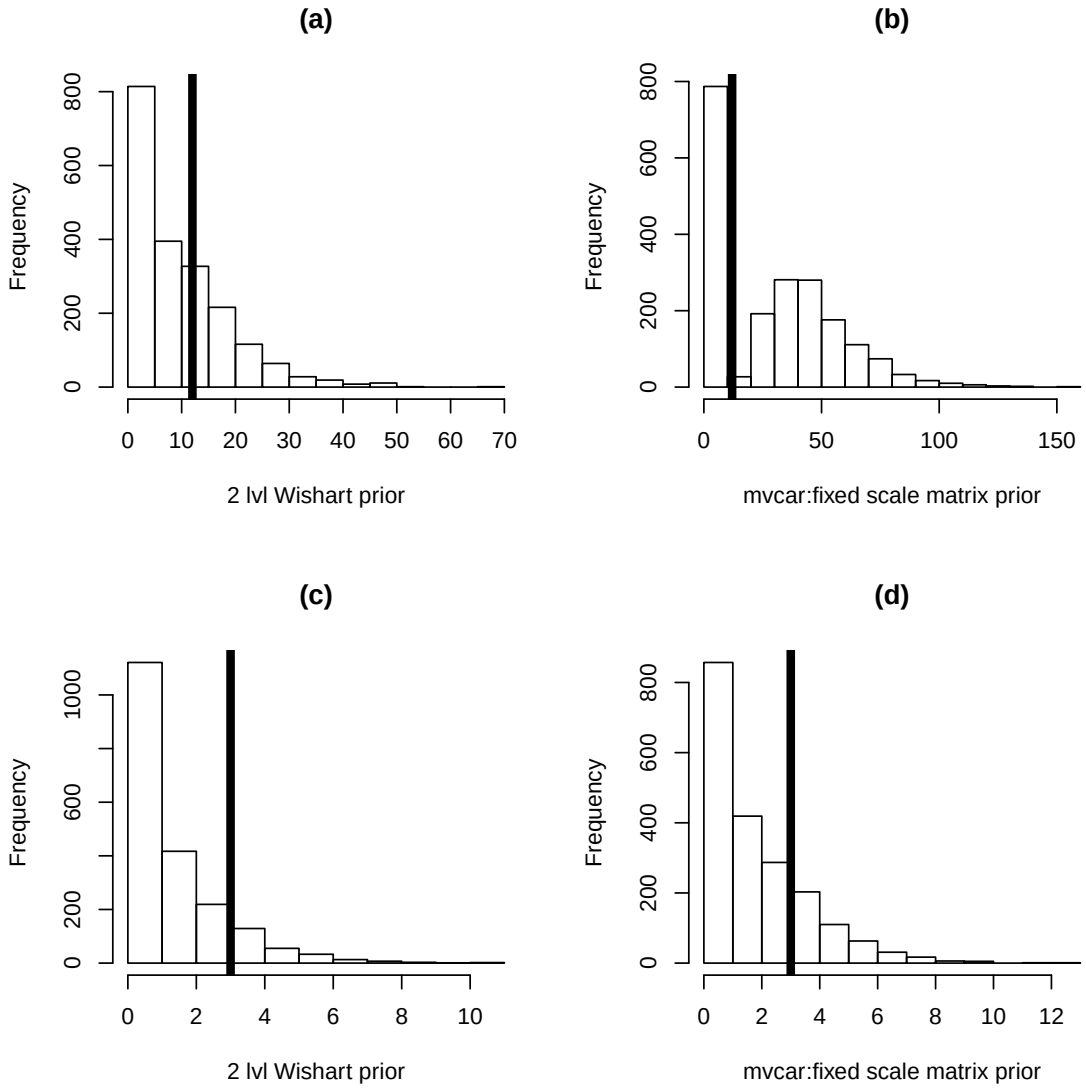


Figure 3.11: (a), (c): The histograms show the posterior predictive distribution of Murder and Non-negligent Manslaughter data for ZINBH at sites 33 and 52, respectively, under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of Murder and Non-negligent Manslaughter data for ZINBH at sites 33 and 52, respectively, under fixed inverse-Wishart prior. The vertical lines represent the observed values.

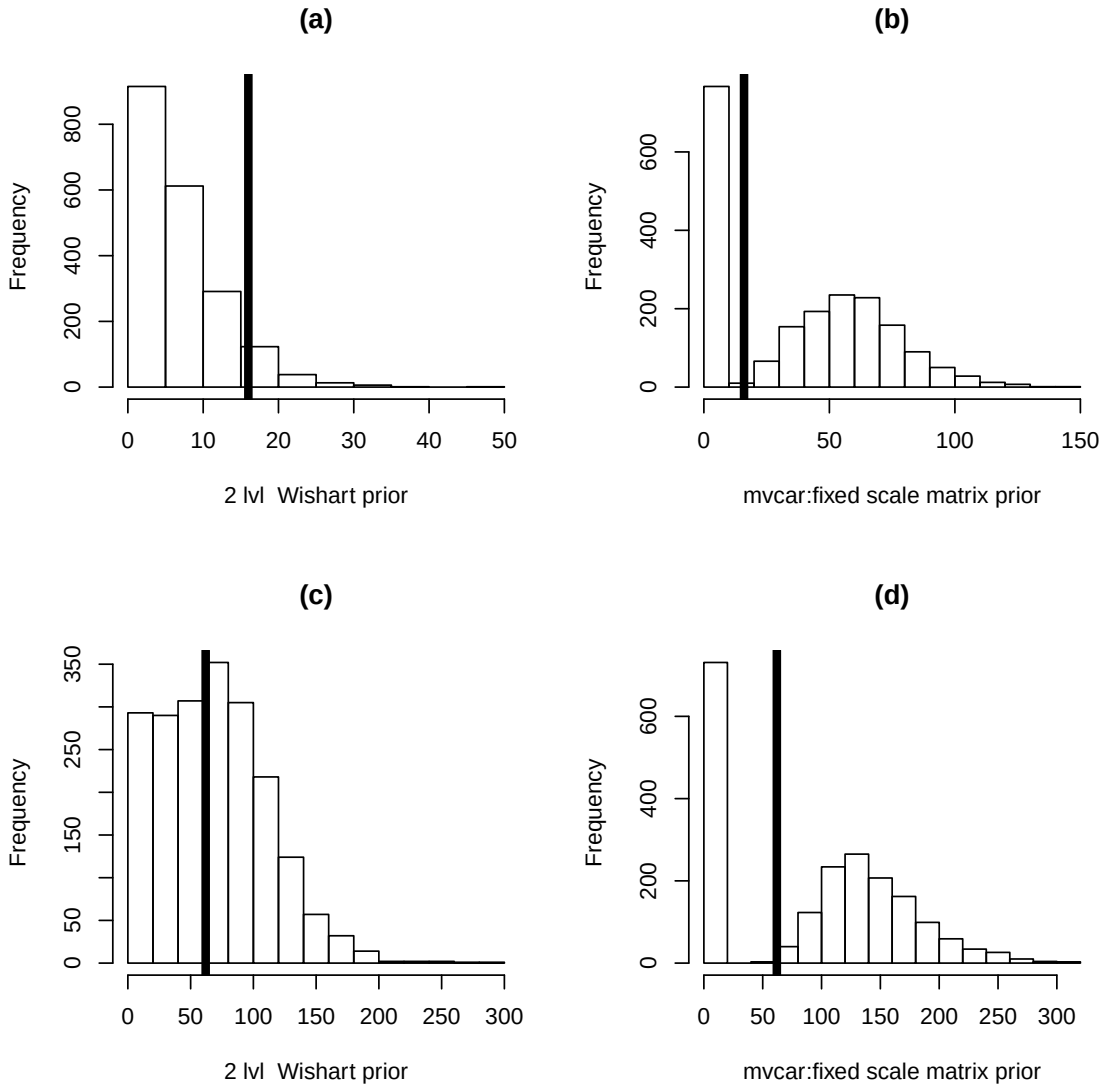


Figure 3.12: (a), (c): The histograms show the posterior predictive distribution of Burglary Forced Entry data for ZINBH at sites 33 and 41, respectively, under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of Burglary Forced Entry data for ZINBH at sites 33 and 41, respectively, under fixed inverse-Wishart prior. The vertical lines represent the observed values.

Table 3.6 shows the 95% credible intervals for α converted to probabilities of zero observations under bivariate model for TLHIW and fixed scale matrix priors, as well as for a univariate model for the TLHIW. These probability intervals (as proportions) capture the actual proportion of zero observations, 38.6% and 37.3% respectively, in the data.

Table 3.6: 95% Credible Intervals for α Converted to Probabilities of Zero Observations

crime	**actual % zeros**	**2 level Wishart**	**fixed scale matrix**	**Univariate TLHIW**
(Y_1)Murder or Non-negligent Manslaughter	38.6%	(31.1%, 55.8%)	(32.8%, 52.1%)	(25.93%, 50.4%)
(Y_2)Burglary with Forced Entry	37.3%	(29.1%,51.5%)	(30.7%,51.2%)	(22.43%, 44.03%)

Table 3.7 shows the regression estimates for bivariate regression estimates under bivariate model for TLHIW and fixed scale matrix priors, as well as for a univariate model for the TLHIW. For the abundance portion of the regression, % Change in Unemployment and Median income (for Multivariate ICAR only) have 95 % credible intervals which contain zero. Based on these intervals, the rest of the covariates are influential and we will interpret their signs and magnitudes. Also, all three models agree on the sign of each coefficient save for Median Income, and intercept.

Conforming to conventional wisdom, High School Graduation Rate has a negative

effect on both crime measurements. Also, Median Income has a negative effect on Murder and Non-negligent Manslaughter. A positive effect of Median Income on Burglary Forced Entry would seem logical due the fact that higher income neighborhoods will tend to have more valuable items within a residence. Only the bivariate TLHIW model predicts this. This may be evidence of a better performance from this model. Population has a positive effect, and in this case plays the roll of an offset, except that we do not fix the coefficient at '1'. Surprising is the negative effect of % Unemployment, indicating that less employment would lower incidence. This is in contrast with the theory that unemployment would accompany higher tendency of crimes, due to stress, unease or desperation. Regression analysis quantifies these effects, but cannot establish causation. A more in depth look at the data, and more measurements of social, economic and psychological nature may illuminate the reasons for these findings. The difference in the sign of the intercept for Burglary Forced Entry in the univariate TLHIW prior model is also not readily explainable. This appears to be at least partially offset with the more negative estimates of % Unemployment and % Change in Unemployment over the last 12 months and maybe just a model predilection for this particular data set rather than an overarching difference in the performance of the three models. Finally, the estimates of the dispersion rate, α_{nb}, are quite different between the TLHIW prior and Fixed Scale Matrix prior models, the former indicating significantly more overdispersion than the latter. From the simulated data in the previous section, it appears that the Fixed Scale Matrix tends towards a more accurate prediction of this parameter, and so we should lean towards concluding that the over dispersion is smaller than that what the TLHIW

prior model is indicating.

It is worth noting that the 95 % credible intervals of the Univariate TLHIW prior model are, on average, wider than those of the bivariate models. While it is theoretically predicted, this is empirical evidence of the benefits of a bivariate approach which accounts for the interdependencies of the two measured responses.

Figures 3.13 through 3.24 represent the countywise medians, the lower bounds of 95% credible intervals, and the upper bound of 95% credible intervals of spatial random effects terms for "occurrence" and "abundance" parts of the regressions. Larger values indicate that for that particular region the geographical influence increases the effect once adjusted for the effects of covariates. For example, in Figure 3.13, the chance of "zero" crime is increased by the spatial influence in Kalkaska and Roscommon counties.

Table 3.7: MI Crime Data Regression Estimates

Regression Estimates						
Murder and Non-negligent Manslaughter:% zeros = 38.6						
	TLHIW		Fixed Scale Matrix		Univariate TLHIW	
variable	Median	95% Credible Int.	Median	95% Credible Int.	Median	95% Credible
α	-0.2362	(-0.5413, 0.1454)	-0.14125	(-0.4264 , 0.1082)	0.007	(-0.378, 0.300)
Intercept	-0.8023	(-1.715, 0.430)	-0.2684	(-0.8029, -0.0026)	0.942	(-0.446, 1.278)
% Uemployment	-0.2304	(-0.4527, -0.0197)	-0.2214	(-0.3339, -0.026)	-0.402	(-0.524, -0.118)
Change in Unemployment	0.3498	(-0.3627, 0.9883)	0.1698	(-0.1308 ,0.4299)	-0.003	(-0.423, 0.269)
H.S. Grad. Rate	-0.065	()-0.0814, -0.0297)	-0.1044	(-0.111, -0.0869)	-0.042	(-0.071 ,-0.032)
Median Income	-0.4542	(-0.7591, -0.0772)	-0.1775	(-0.3892, 0.0612)	-0.918	(-1.200, -0.373)
log(Population)	1.07	(0.9235, 1.1690)	1.098	(0.9703, 1.277)	1.051	(0.740,1.277)
α_{nb1}	0.8532	(0.478, 1.863)	3.0545	(2.153, 6.258)	1.11	(0.535, 3.083)
Burglary Forced Entry : % zeros = 37.3						
α	-0.1816	(-0.4087, 0.2115)	-0.0331	(-0.4019, 0.1659)	0.098	(-0.198, 0.402)
Intercept	-0.5859	(-2.456 ,-0.1117)	-1.847	(-2.361, -1.512)	1.274	(0.936, 1.752)
% Uemployment	-0.193	(-0.2497, -0.092)	-0.2663	(-0.3716, -0.1896)	-0.465	(-0.592, -0.301)
Change in Unemployment	-0.6235	(-0.7914, -0.2308)	-0.3301	(-0.6156, -0.003)	-0.865	(-1.185 ,-0.323)
H.S. Grad. Rate	-0.05511	(-0.0805, -0.0324)	-0.0342	(-0.049 -0.0247)	-0.025	(-0.035, -0.017)
Median Income	0.5627	(0.3288, 1.098)	-0.2483	(-0.4665, -0.0076)	-0.304	(-0.696,-0.025)
log(Population)	0.4249	(0.2800, 0.6432)	0.8604	(0.7328, 0.9840)	0.664	(0.390, 0.751)
α_{nb2}	0.6511	(0.4653, 1.048)	2.408	(2.127, 3.532)	0.857	(0.517, 1.556)

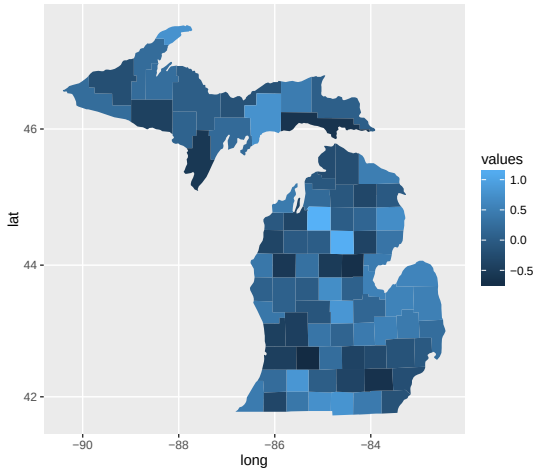


Figure 3.13: Countywise medians of spatial random effects terms for the abundance part of the regression for the variable Murder and Non-negligent Manslaughter

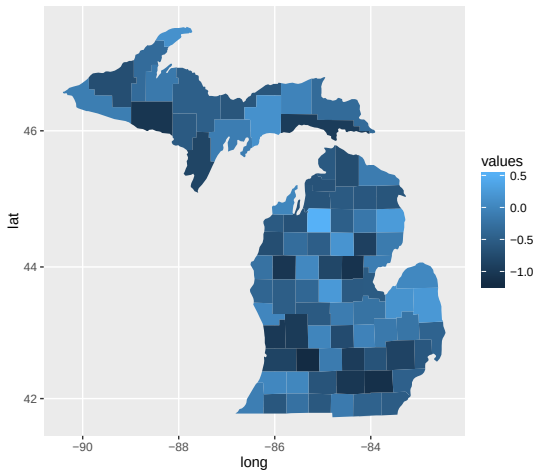


Figure 3.14: Countywise lower bounds of 95% credible intervals for spatial random effects terms for the abundance part of the regression for the variable Murder and Non-negligent Manslaughter

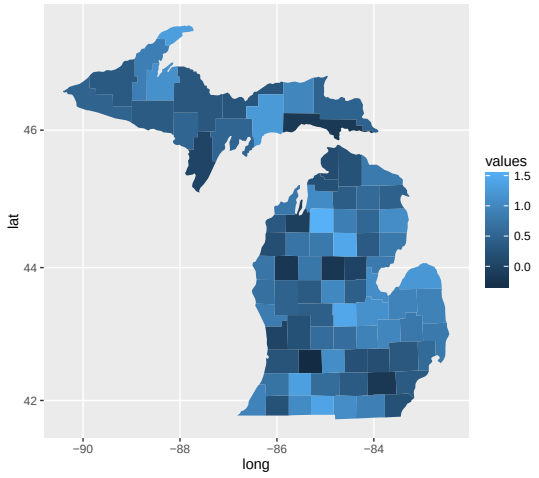


Figure 3.15: Countywise upper bounds of 95% credible intervals for spatial random effects terms for the abundance part of the regression for the variable Murder and Non-negligent Manslaughter

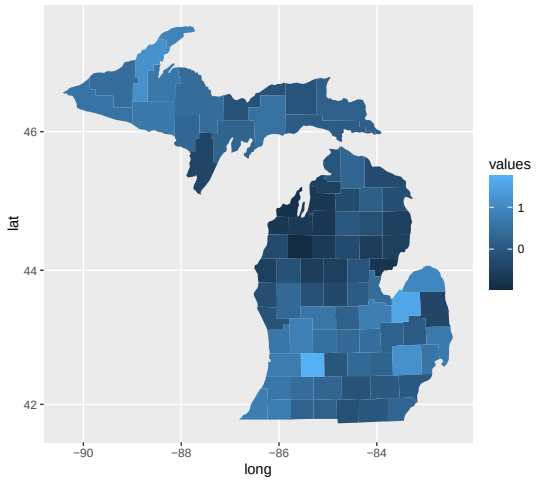


Figure 3.16: Countywise medians of spatial random effects terms for the occurrence part of the regression for the variable Murder and Non-negligent Manslaughter

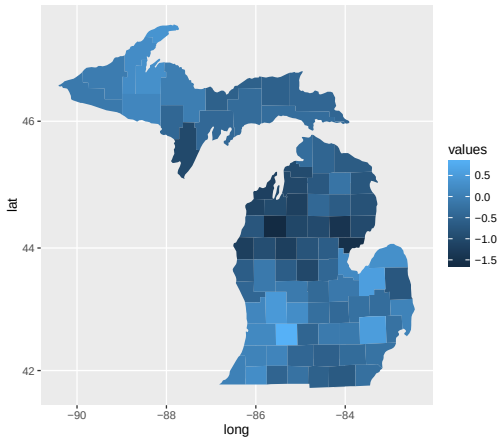


Figure 3.17: Countywise lower bounds of 95% credible intervals for spatial random effects terms for the occurrence part of the regression for the variable Murder and Non-negligent Manslaughter

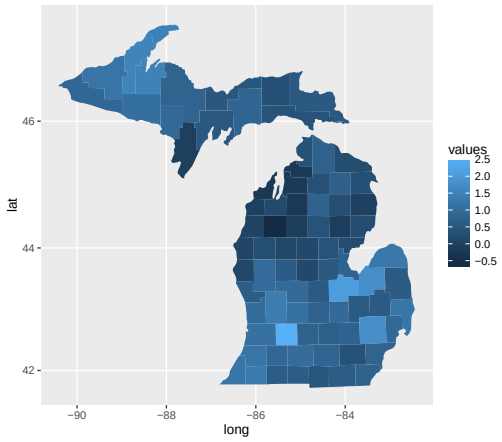


Figure 3.18: Countywise upper bounds of 95% credible intervals for spatial random effects terms for the occurrence part of the regression for the variable Murder and Non-negligent Manslaughter

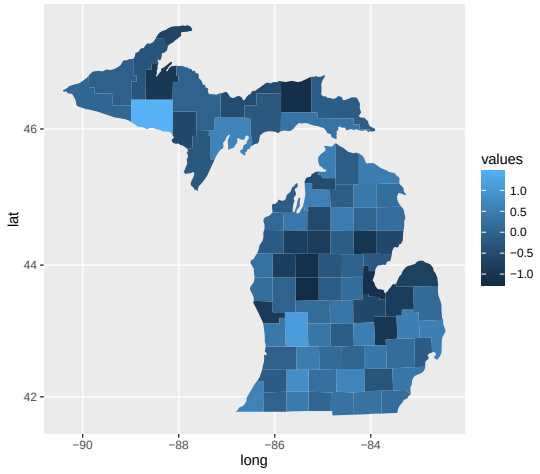


Figure 3.19: Countywise medians of spatial random effects terms for the abundance part of the regression for the variable Burglary Forced Entry

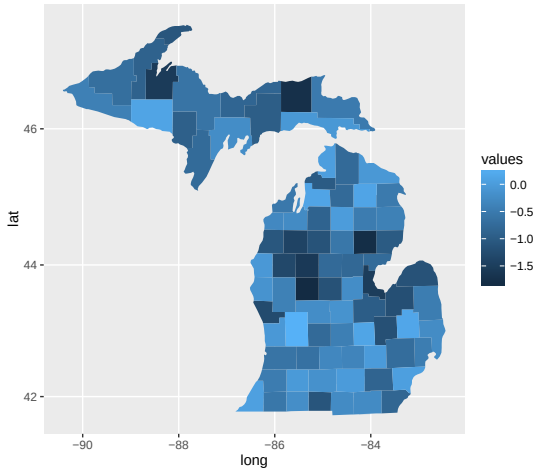


Figure 3.20: Countywise lower bounds of 95% credible intervals for spatial random effects terms for the abundance part of the regression for the variable Burglary Forced Entry

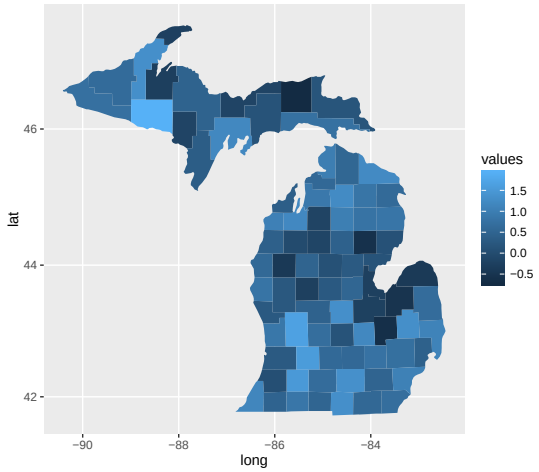


Figure 3.21: Countywise upper bounds of 95% credible intervals for spatial random effects terms for the abundance part of the regression for the variable Burglary Forced Entry

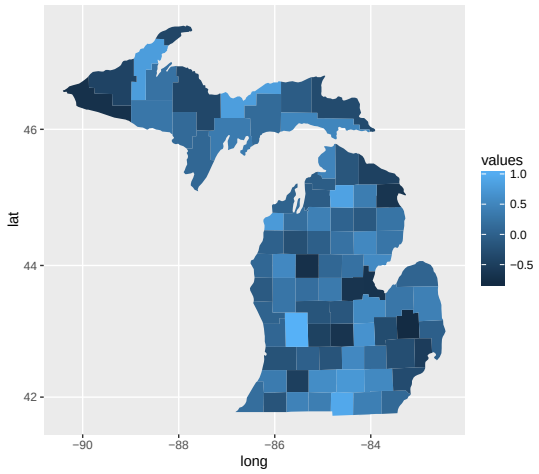


Figure 3.22: Countywise medians of spatial random effects terms for the occurrence part of the regression for the variable Burglary Forced Entry

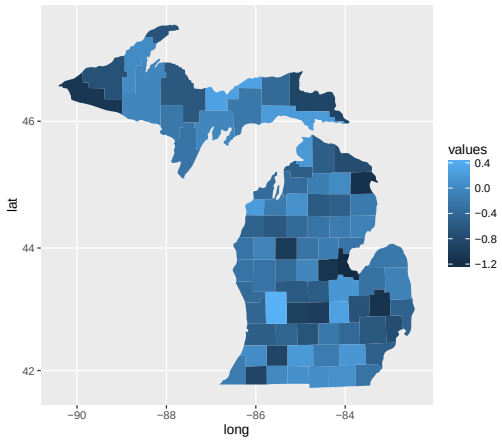


Figure 3.23: Countywise lower bounds of 95% credible intervals for spatial random effects terms for the occurrence part of the regression for the variable Burglary Forced Entry

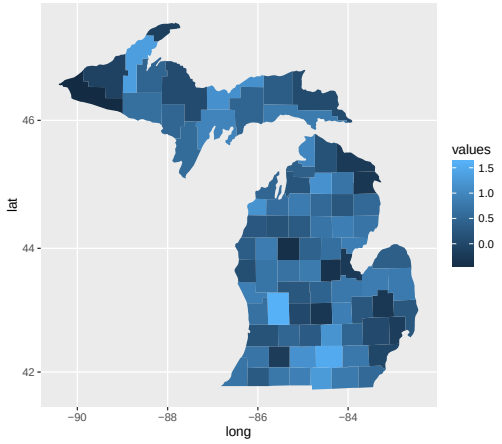


Figure 3.24: Countywise upper bounds of 95% credible intervals for spatial random effects terms for the occurrence part of the regression for the variable Burglary Forced Entry

3.5 WinBUGS

Bayesian analysis has always faced the issue of large computational demands with respect to MCMC methods. With the recent and continuing advances in processor and random access memory technology, even personal computers have become very fast, and these demands have become less of an obstacle. Even so, calculating posterior distributions often lead to integrals with no closed-form solutions, or expressions which are not proportional to known distributions, making sampling from MCMC very challenging.

Then BUGS (Bayesian inference Using Gibbs Sampling; see source: http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/), an open source software package, was released which allowed users to do Bayesian analysis from knowing only the model equations, sampling and prior distributions of the model to which the data are fitted. Later an interactive GUI Windows version of this software, WinBUGS, was released. R2WinBUGS, a package developed for open source statistical software R, facilitated a 'bugs' function to be called from within R which allowed the sampling tools of WinBUGS to be accessed.

While there are packages that exist for R which can be used for spatial analysis, e.g spBayes and CARBayes, and zero-inflated distribution analysis, e.g. pscl, these are not conveniently combined for multivariate data. In particular, where CAR effects are related between multiple responses within a single observation as in the situation of the research (See Figure 1 above) presented here.

Neelon et al. [2012] employed WinBUGS for ZIP and ZINB models with ICAR effects. Namely, the function mv.car(), from the open source software GEOBUGS designed for spatial analysis, was implemented to sample from mulitvariate Guassian ICAR random variables. The function mv.car() requires, as an input argument, along with lattice structure information, a precision matrix (the inverse of a variance-covariance matrix) with a Wishart prior distribution. The function is so structured that it requires this Wishart prior to be specified by the WinBUGS built-in distribution function dwish(). The input arguments for dwish() are degrees of freedom and scale matrix. WinBUGS makes a restriction that both the degrees of freedom and the scale matrix are both constants, not subject to updating. This restriction precludes implementation of the TLHIW prior for the precision matrix used in the ICAR effects distribution as described previously.

Specification of multivariate Gaussian ICAR effects with prior and hyper priors on the precision matrix, so as to ensure a flat posterior distribution for the correlations between effect components as per Huang and Wand [2013], requires that the both the ICAR effects and the multilevel Wishart prior be manually coded. While the conditional prior distribution for ICAR effects on a lattice (see section 3 above) is a relatively straight forward construction, that of the Wishart prior is not. The WinBUGS software is somewhat limited in its capacity to manipulate and sample vectors and matrices, making the implementation of the Wishart distribution function very challenging. This obstacle was surmounted by employing the following method of

generating random Wishart matrices of given parameters.

Suppose n p-variate observations $\boldsymbol{x}_i$ are identically normally distributed with mean 0 and variance-covariance matrix Σ, with x_i perpendicular x_j for $i \neq j$, and

$$\boldsymbol{S} = \sum_{i=1}^{n} \boldsymbol{x_i x_i'}, \tag{3.18}$$

then $\boldsymbol{S}$ conditioned on $\boldsymbol{\Sigma}$ is distributed, up to a constant, as a p×p random Wishart matrix with degrees of freedom n and scale matrix $\boldsymbol{\Sigma}$.

Choosing ν=2, n= ν + p -1 = 5, Σ to be a diagonal matrix with entries $\frac{1}{2\nu a_i}$, and a_i each having a hyper prior Gamma distribution with shape $\frac{1}{2}$ and scale $\frac{1}{A^2}$, the conditions of Huang and Wand [2013] above are met. This implementation of ICAR with hyperpriors, involving a large number of random samples from normally distributed random variables, greatly increases the run time for each iteration of this parameterization.

Furthermore, WinBUGS interprets use of (i.e. formula involving) a particular component of a multidimensional, block updated random variable, e.g. a matrix, as requiring a separate updating scheme for that particular component. This requires the use of marginal and conditional distributions for components in order to get WinBUGS to update.

Chapter 4

Variable Selection

The primary goal of regression is to model the response variable as a function of the covariates as accurately as possible, while still allowing reasonable interpretation of the results: the simpler the model, i.e. the fewer covariates, the more transparent the interpretation. Thus it is of great interest to the investigator to find the "best" subset of covariates for the regression model. Conventional techniques, such as forward, backward and stepwise selection, based upon criteria including R^2 and AIC, are available but often suffer from "over fitting" of regression parameters, especially in the presence of strongly competing parameters, i.e. co-linearity Casella et al. [2010], Bae and Mallick [2004]. One method of addressing this drawback is by penalizing the likelihood function for over-estimation of parameters. A popular method implementing this penalization is the Least Absolute Shrinkage and Selection Operator (LASSO) estimation of regression parameters. In this method, introduced by Tibshirani [1996], the likelihood function is modified by subtracting a norm (pro-

portional to L1 for LASSO) of the regression parameters, similar to the method of Lagrange multipliers. In this way large numbers of regression coefficients, as well as large estimates of regression coefficients, reduce the likelihood function. Buu et al. [2011] develop variable selection methods for ZIP regression models including LASSO and Smoothly Clipped Absolute Deviation penalty (SCAD). Variable selection for the semi-parametric model, among those covariates that are initially included in the nonparametric portion, would be desirable, as well. Li and Lian [2008] first estimate regression coefficients and smoothing parameters through penalized maximum likelihood and quasi- likelihood methods, and then employ a series of F-tests on nested subgroups of nonparametric oriented covariates, similar to backwards, forward, and stepwise regression.

This research employs a Bayesian approach to facilitate the the penalization function of the penalized regression technique above. The imposition of symmetric, leptokurtic distributions as prior distributions for the regression coefficient carries out the 'shrinkage' effect seen in LASSO. Trevor and Casella [2008], Yaun and Lin [2005a] and Bae and Mallick [2004] used double exponential distributions to this end. The sharper peak and fast tail decay of Laplace like distribution cause unimportant variable coefficients to shrink to zero faster than Gaussian distributions. In a Bayesian approach, Bayesian credible intervals are constructed by MCMC samples from posterior distributions, and decisions are made based upon these intervals. Yaun and Lin [2005b] and Hans [2010] used a mixture distribution, 'zero-inflated,' consisting of a point mass at zero and Laplace distribution, for priors on coefficients. By

this method they were able to arrive at more 'definitive' decisions about coefficients than by Laplace alone. This research will implement the the technique of Kuo and Mallick [1998] and Lykou and Ntzoufras [2013] and incorporate Bernoullli indicator variables within the priors of the regression coefficients. In this way, objective decisions of variable selection are based upon the posterior distribution of these Bernoulli variables.

Extensions of LASSO include adaptive LASSO, Group LASSO, and elastic net Casella et al. [2010] C. Leng and Nott [2012]. In adaptive LASSO, the degree of shrinkage imposed is different for different coefficients. This allows for more flexibility in the selection process. In Group LASSO, a combination of L1 and L2 norms is used in the penalty function, categorical variable levels are treated separately. Unlike regular LASSO, Group LASSO can select just certain levels of a categorical variable. In elastic net, a convex combination of norms, L1 and L2, is used to implement the regression penalty, see Zou and Hastie [2005]. Through studies and real data analysis, it was shown that Elastic Net outperforms LASSO.

In this research, we will restrict ourselves to the elastic net method of penalized regression and apply it to univariate Negative Binomial Hurdle data with spatial random effects on a lattice. In the two part regression for each of the zero-inflated responses, multivariate ICAR spatial random effects are used to account for the random spatial effects.

4.1 Parameterization

Let Y_j (j=1,...,n) be a response, each assumed to have a Negative Binomial Hurdle distribution for this research, with likelihood, conditioned on parameter collection $\boldsymbol{\Theta}$, $L(Y_j|\boldsymbol{\Theta})$. $\boldsymbol{\Theta}$ includes parameters upon which Y_j depends, including the set of regression coefficients $\boldsymbol{\beta}$. Then the joint likelihood for the set of observations $\mathbf{Y}=\{Y_1, Y_2, \ldots, Y_n\}$ is the product of their individual likelihoods:

$$L(\mathbf{Y}|\boldsymbol{\Theta}) = \prod_{j=1}^{n} L(Y_j|\boldsymbol{\Theta}) \tag{4.1}$$

Penalized regression, which allows for modifying which parameter sets are favored, is realized by minimizing the sum of the negative expectation of the log likelihood and the penalty function:

$$-\mathbf{E}_{\boldsymbol{\Theta}}[\log L(\boldsymbol{Y}|\boldsymbol{\Theta})] + h(\boldsymbol{\theta}, \boldsymbol{\lambda}) \tag{4.2}$$

with respect to $\boldsymbol{\theta}$. As mentioned above, this resembles the use of Lagrange multipliers to optimize with respect to constraints. Here $h(\boldsymbol{\theta}, \boldsymbol{\lambda})$ is the penalty function, with arguments $\boldsymbol{\theta}$, the collection of regression coefficients, and $\boldsymbol{\lambda} \geq 0$, the collection of tuning parameters. $\boldsymbol{\lambda}$ controls the rate at which regression coefficients for less important covariates approach zero, with smaller values of $\boldsymbol{\lambda}$ associated with coarser

models. As $\boldsymbol{\lambda} \to 0$, and the amount of penalization diminishes to 0, the regression coefficients approach their maximum likelihood estimates.

For elastic net (see Zhou and Hastie [2005]), the penalty function is

$$h(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \lambda_1 \Sigma_{k=1}^{p-1} |\theta_k| + \lambda_2 \Sigma_{k=1}^{p-1} \theta_k^2 \tag{4.3}$$

In a Bayesian approach, the implementation of the above penalty scheme is accomplished through careful imposition of a prior distribution on the regression coefficients. For our model we impose on the regression coefficients, $\boldsymbol{\beta}_k$ (k=1...,p), an exponential prior with the elastic net penalty function (4.4) in the the argument:

$$f(\boldsymbol{\beta}, \boldsymbol{\lambda}) \propto \exp\left\{-\lambda_1 \Sigma_{k=1}^{p} |\beta_k| - \lambda_2 \Sigma_{k=1}^{p} \beta_k^2\right\} \tag{4.4}$$

Incorporating this in the likelihood for Negative Binomial Hurdle distributed responses (3.10):

$$\begin{aligned}\mathcal{L}(Y) = \prod_{j=1}^{j=n} \Bigg\{ & [\pi_j]^{I(Y_j=0)} \\ & \times \left[(1-\pi_{i,j}) \frac{\frac{\Gamma(y_{i,j}+\alpha_{nb})}{\Gamma(y_{i,j}+1)\Gamma(\alpha_{nb})} \left(\frac{\alpha_{nb}}{\mu_{i,j}+\alpha_{nb}}\right)^{\alpha_{nb}} \left(\frac{\mu_{i,j}}{\mu_{i,j}+\alpha_{nb}}\right)^{y_{i,j}}}{1 - \left(\frac{\alpha_{nb}}{\mu_{i,j}+\alpha_{nb}}\right)^{\alpha_{nb}}} \right]^{I(Y_{i,j}\geq 1)} \Bigg\} \\ & \times \exp\left\{-\lambda_1 \Sigma_{k=1}^{p} |\beta_k| - \lambda_2 \Sigma_{k=1}^{p} \beta_k^2\right\}\end{aligned} \tag{4.5}$$

Taking the log of (4.5), and then taking negative expectation, we obtain an expression in the desired form of (4.2)

The expectation of the log of (4.5) with respect to the estimated parameters $\{\boldsymbol{\pi}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\lambda}\}$, for Y_i's distributed as Negative Binomial Hurdle, is not available in closed form. Therefore, some numerical method of estimation must be employed to minimize this expression. The Bayesian approach utilizing MCMC estimates makes it an appealing choice.

In the traditional LASSO form of penalized regression, the penalty function takes the form of only the L_1 measure portion of $h(\boldsymbol{\theta}, \boldsymbol{\lambda})$ above. This choice of penalty function is ineffective in certain situations, such as when a subset of predictors exhibit high pairwise correlations, or when p$\gg$ n. In the former, LASSO tends to select only one predictor from the group, forgoing the rest, and without a clear preference among them. For the latter, LASSO is restricted to selecting at most "n" predictors. The addition of the L_2 norm of the vector of regression coefficients, as set forth in $h(\boldsymbol{\theta}, \boldsymbol{\lambda})$ for elastic net, addresses both of these situations, as described by Zhou and Hastie [2005]. First, optimization of under the elastic net penalty for data $\mathbf{X}$, with sample size = n, can be transformed to an equivalent optimization with the LASSO penalty on augmented data $\mathbf{X}$, with a sample size = n + p. In this transformed optimization problem, $\mathbf{X}$ has rank p, and thus all predictor can potentially be selected. Further more, the elastic net penalty function is strictly convex, and, as a result, assigns very

similar coefficients to highly correlated predictors, including the same coefficient to identical predictors in the extreme case.

As previously mentioned, the penalty functions for penalized regression can be incorporated into the likelihood via careful choice of prior distribution for the regression coefficients. For Bayesian elastic net, we follow the formulation of Casella et al. [2010]. Namely, we impose upon $\boldsymbol{\theta}$ a prior distribution of a zero-mean, multivariate Gaussian distribution with diagonal variance covariance matrix having entries of the form $\tau^2((s_1^{-2}+\lambda_2)^{-1}, (s_2^{-2}+\lambda_2)^{-1}, \ldots, (s_p^{-2}+\lambda_2)^{-1})$, where $(s_1^2, s_2^2, \ldots, s_p^2)$ are independently and identically distributed exponentially distributed random variables with mean $2\lambda_1^{-2}$. Integrating out the scale terms $(s_1^2, s_2^2, \ldots, s_p^2)$, the marginal distributions of the regression coefficients, conditioned on τ, are as follows:

$$f(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\tau}) \propto \exp\left\{ -\frac{\lambda_1}{\tau^2} \sum_{k=1}^{p} |\beta_k| - \frac{\lambda_2}{\tau^2} \sum_{k=1}^{p} \beta_k^2 \right\} \tag{4.6}$$

While this leads to a desirable unimodal posterior distribution, the full conditional distribution is not independent of λ_2.

As described in Lykou and Ntzoufras [2013], for more objective variable selection criteria, a set of indicator variables, $(\kappa_1, \kappa_2, \ldots, \kappa_p)$, each identically and independently distributed Bernoulli random variables with prior probability of success = 0.5, has been employed. This forms a new set of regression coefficients: $\theta_k^\star = \kappa_k \theta_k$. Then variable selection is performed by monitoring the posterior medians of $\theta_k^\star$, where

a median value of zero means the variable is not selected.

Because of the primary interest in the novel formulation of ICAR effect using the TLHIW, variable selection will be performed using that model only. Variable selection will be performed only on the non-zero part of the regression as the Bernoulli part of the regression is modeled only through an intercept for simplicity. More complex modeling of probability of zero observations, on bivariate data, is goal for future research.

4.2 Results

The performance of the variable selection method was assessed using simulated univariate Negative Binomial Hurdle distributed data on a lattice, with spatial random effects. The simulated data were generated similarly to that in the method comparison section 3.2 above, except nine covariates, each standard normally distributed, along with an intercept was used. 15000 MCMC iterations were completed with the first 5000 discarded as burn in. For the elastic net penalty function (35), both λ_1 and λ_2 were set to 0.1.

Table 9 shows results from eight separate simulations, with first three having lower (20%-30%) percentages of zero observations and the remaining five with higher (45%

to 65%) percentages of zero observations. Each column displays the median MCMC values of the nine regression coefficients, β_{1-9}. A median of '0' indicates the model did not select that variable as influential. Beside each median is a symbol indicating the performance of the model with respect to the corresponding regression coefficient on that particular simulation. Table 9 reflects good performance of the TLHIW Prior model in variable selection via elastic net. Here we see appropriate selection of an important variable, indicated with '+', 91.6 % of the time (22/24). We also see appropriate exclusion of a unimportant variable, indicated by '*', 94.4% of the time (51/54). Further more, an unimportant variable was selected, indicated by '!', only 5.6% of the time (3/54), and an important variable was excluded, indicated by '-", only 8.3 % of the time (2/24). As the proportion of zeros increases, the model may be more likely to select a non-important variable, but only slightly if at all. One out of three for lower proportions of zeros, and two out of five for higher proportions of zeros, are not significantly different for these sample sizes, and the model still 'deselects' these non-important variables for the majority of time for either level of zeros in these simulations. The proportion of zeros doesn't appear to affect the selection of important variables. Again, one in three for lower proportions and one in four for larger proportions are not significantly different.

Table 4.1: Univariate Variable Selection

UNIVARIATE VARIABLE SELECTION									
A median of '0' indicates variable not selected ("+" = important variable selected) ("*" = unimportant variable not selected) ("!" = unimportant variable selected) ("-" = important variable not selected)									
		TLHIW MCMC median							
variable coefficient	true	$sim1$ $\%zero$ 28.1	$sim2$ $\%zero$ 26	$sim3$ $\%zero$ 28.3	$sim4$ $\%zero$ 52.5	$sim5$ $\%zero$ 60	$sim6$ $\%zero$ 61.6	$sim7$ $\%zero$ 51.2	$sim8$ $\%zero$ 47.2
$\beta(1)$	-2.77	$-2.66+$	$-2.53+$	$-2.66+$	$-2.43+$	$-2.69+$	$-2.91+$	$-2.6+$	$-2.71+$
$\beta(2)$	0	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$	0.192!	$-0.048!$
$\beta(3)$	0	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$
$\beta(4)$	0	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$
$\beta(5)$	0	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$
$\beta(6)$	-4.3	$-4.12+$	$-3.86+$	$-3.94+$	$-3.2+$	$-4.46+$	$-4.42+$	$-4.03+$	$-3.96+$
$\beta(7)$	0	$0*$	0.26!	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$
$\beta(8)$	-1.5	$-1.25+$	$-1.7+$	$0-$	$-1.69+$	$-1.18+$	$0-$	$-1.49+$	$-1.16+$
$\beta(9)$	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$	$0*$

Chapter 5

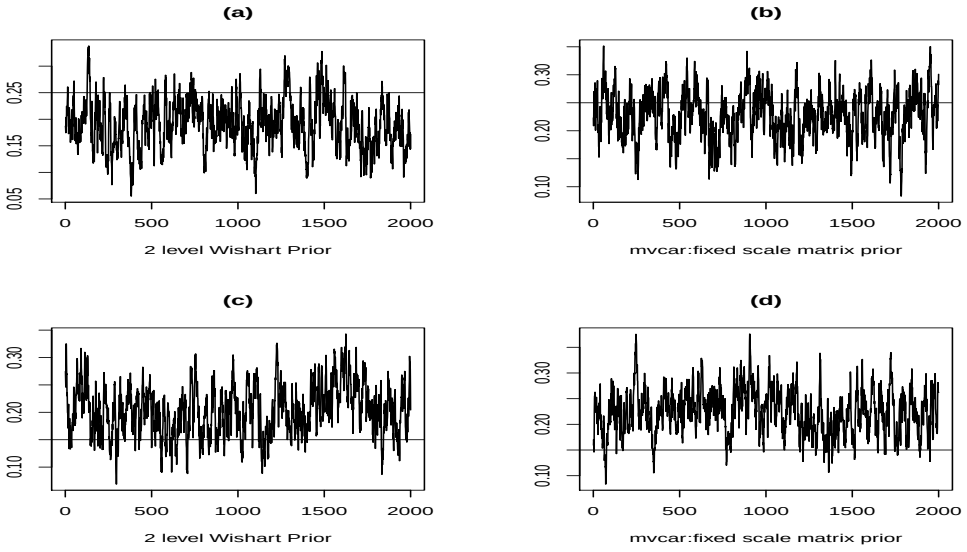
Summary and Conclusions

The effectiveness of the bivariate ZINBH regression model with random spatial effects, as developed in this manuscript, has been demonstrated for both a standard regression model and, univariately, in variable selection. Employing ICAR effects to model spatial dependency can help account for random noise and spatial influences that may accompany data taken over a lattice. Using a Bayesian approach we are able to gain ground on the complex and challenging problem of successfully modeling zero-inflated data with spatial dependencies. Also, the Bayesian approach allows simple and effective implementation of penalized regression, specifically elastic net, as means for objective variable selection. Introducing a multi-level prior structure into the Wishart scale matrix of multivariate normally distributed ICAR effect, we reduce the chance of subjective bias in spatial correlations and ensure a non-informative posterior distribution for these correlations.

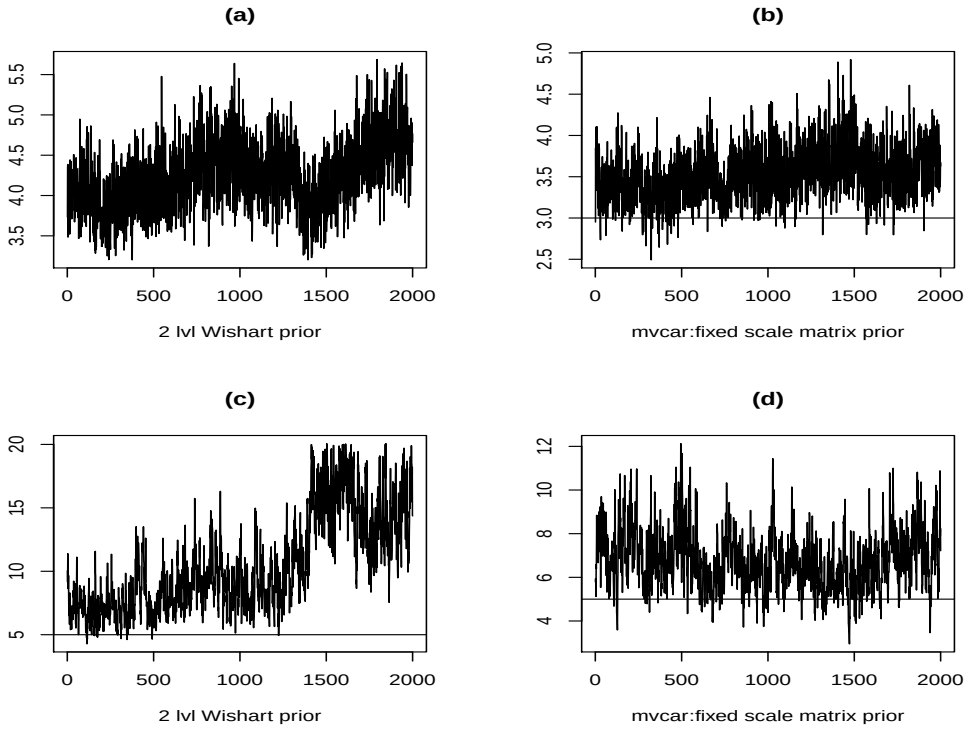
Suggestions for further research include extending the occurrence portion of the regression beyond an intercept and include more covariates. Moreover, a more detailed neighborhood structure, beyond edge sharing, could improve the sensitivity and accountability of the spatial relationships present in the model. Also, extending the model to a non-hurdle model, where zero observations can originate from the abundance portion of the regression, would further expand the applications of this model. New approaches which would decrease the MCMC simulation run time would improve the usefulness of this method. Such approaches might include running the analysis totally within R, avoiding the significant limitations of WinBUGS. In addition, because of the complex nature of the zero-inflated models, it would be very advantageous to have a more objective measure of the fit of the model to the data, perhaps in the use of posterior predictive distributions. Finally, extension to a Spatial-Temporal model would expand the use of this model in many areas including epidemiology and economics.

Appendix

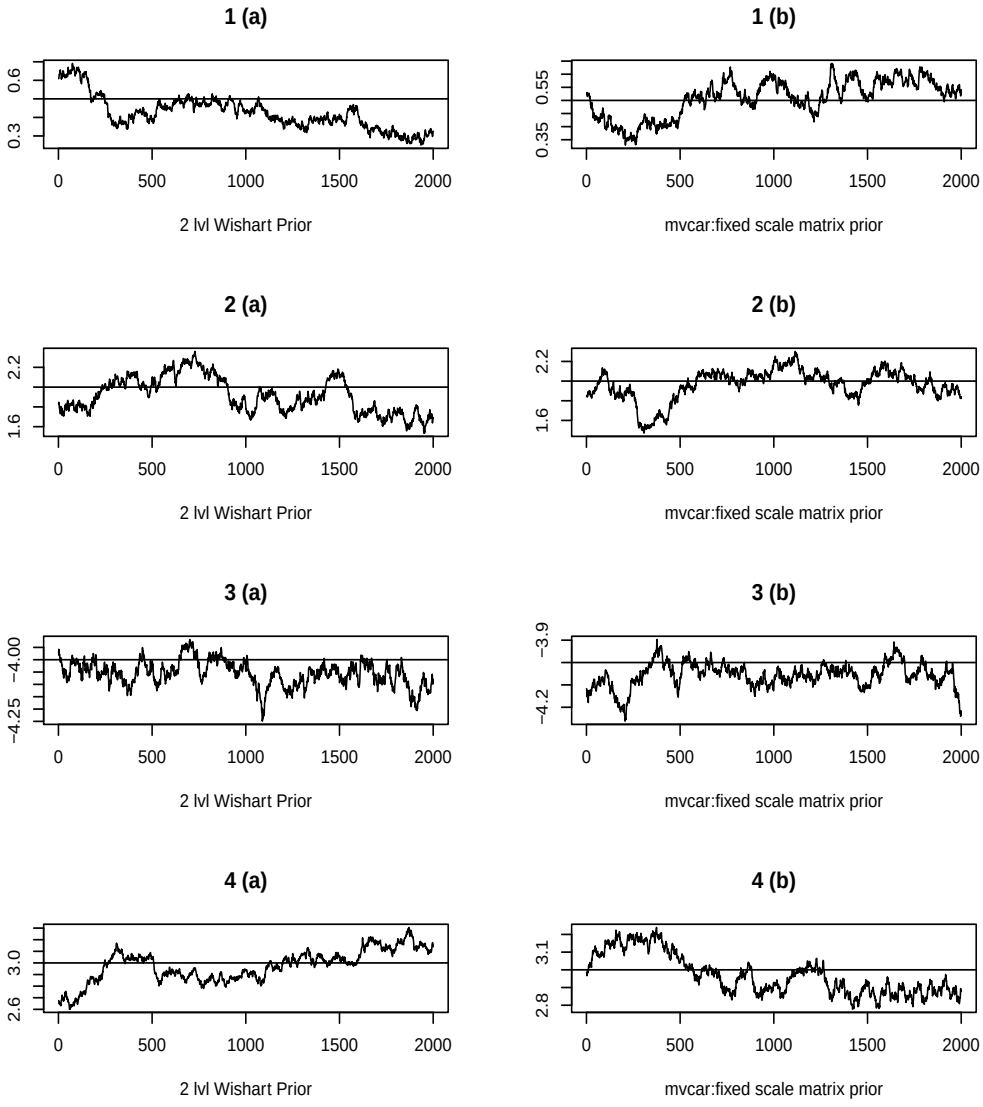
MCMC trace plots for occurrence regression coefficients: (a), (b): the graphs show the trace of iteration values of α for simulated data of Y_1 for ZINBH under TLHIW and fixed scale matrix prior models, respectively. (c), (d): show the trace of iteration values of α for simulated data of Y_2 for ZINBH under TLHIW and fixed scale matrix prior models, respectively. The horizontal lines represent the true values.



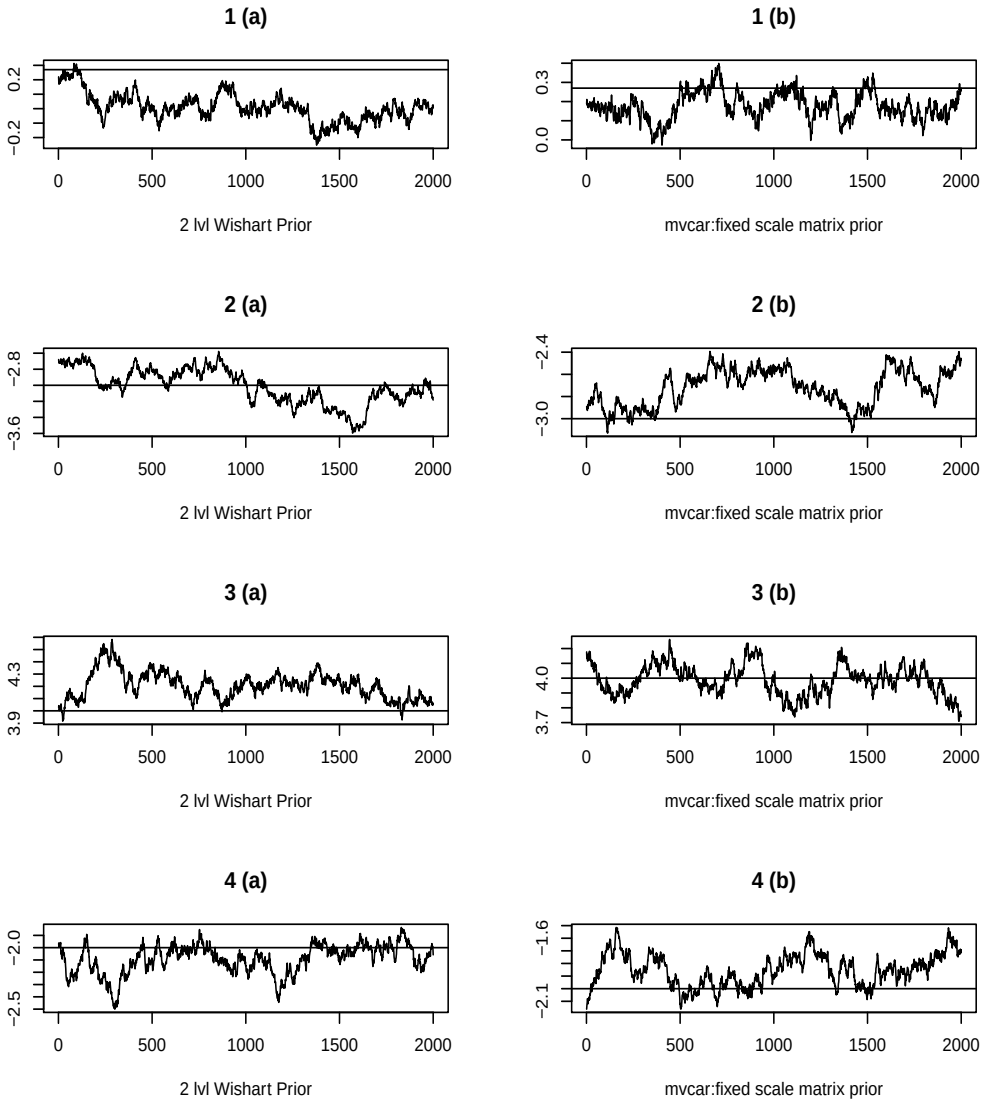
MCMC trace plots for dispersion parameters: (a), (b): the graphs show the trace of iteration values of α_{nb} for simulated data of Y_1 for ZINBH under TLHIW and fixed scale matrix prior models, respectively. (c), (d): show the trace of iteration values of α_{nb} for simulated data of Y_2 for ZINBH under TLHIW and fixed scale matrix prior models, respectively. The horizontal lines represent the true values.



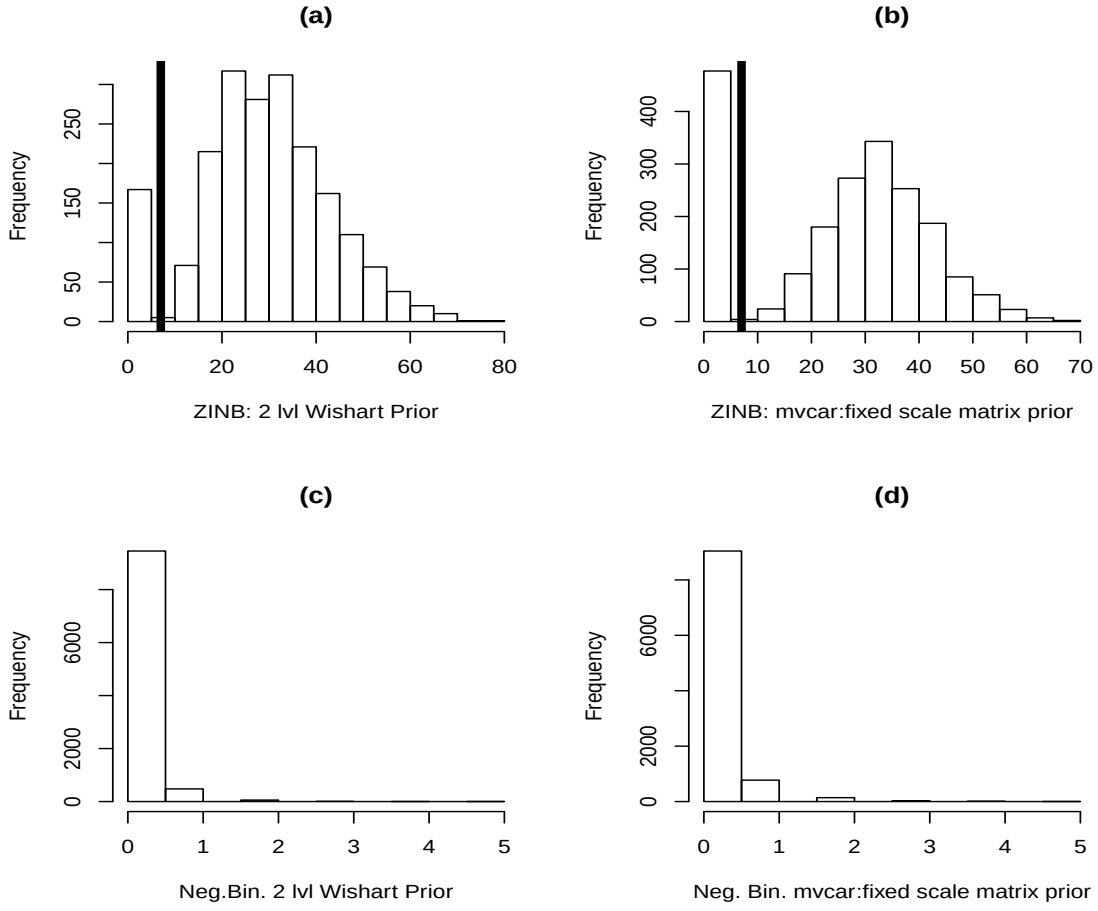
MCMC trace plots for abundance regression coefficients for simulated data of Y_1 for ZINBH: 1(a), 2(a), 3(a), 4(a):the graphs show the trace of iteration values of $\beta_1, \ldots, \beta_4$, respectively, under TLHIW and fixed scale matrix prior models. 1(b), 2(b), 3(b), 4(b):the graphs show the trace of iteration values of $\beta_1, \ldots, \beta_4$, respectively, under fixed scale matrix prior models. The horizontal lines represent the true values.



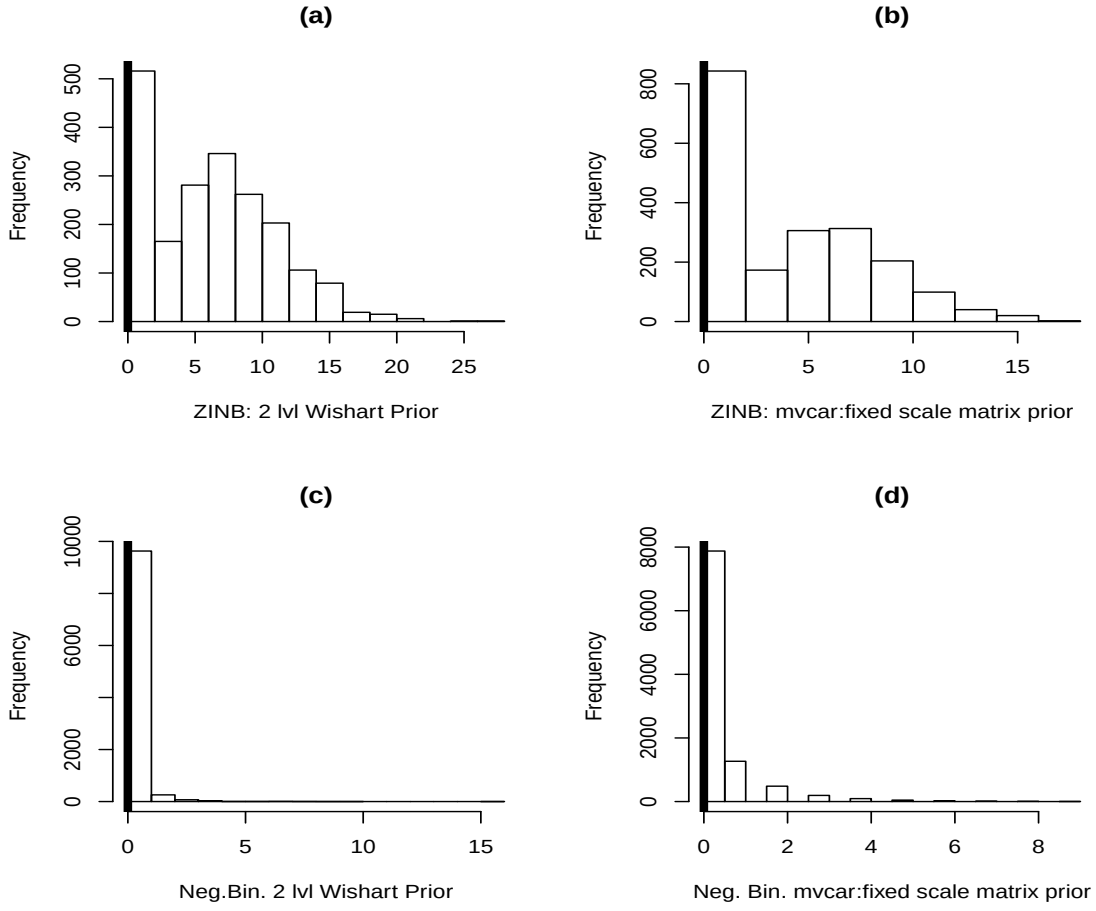
MCMC trace plots for abundance regression coefficients for simulated data of Y_2 for ZINBH: 1(a), 2(a), 3(a), 4(a):the graphs show the trace of iteration values of $\beta_1, \ldots, \beta_4$, respectively, under TLHIW and fixed scale matrix prior models. 1(b), 2(b), 3(b), 4(b):the graphs show the trace of iteration values of $\beta_1, \ldots, \beta_4$, respectively, under fixed scale matrix prior models. The horizontal lines represent the true values.



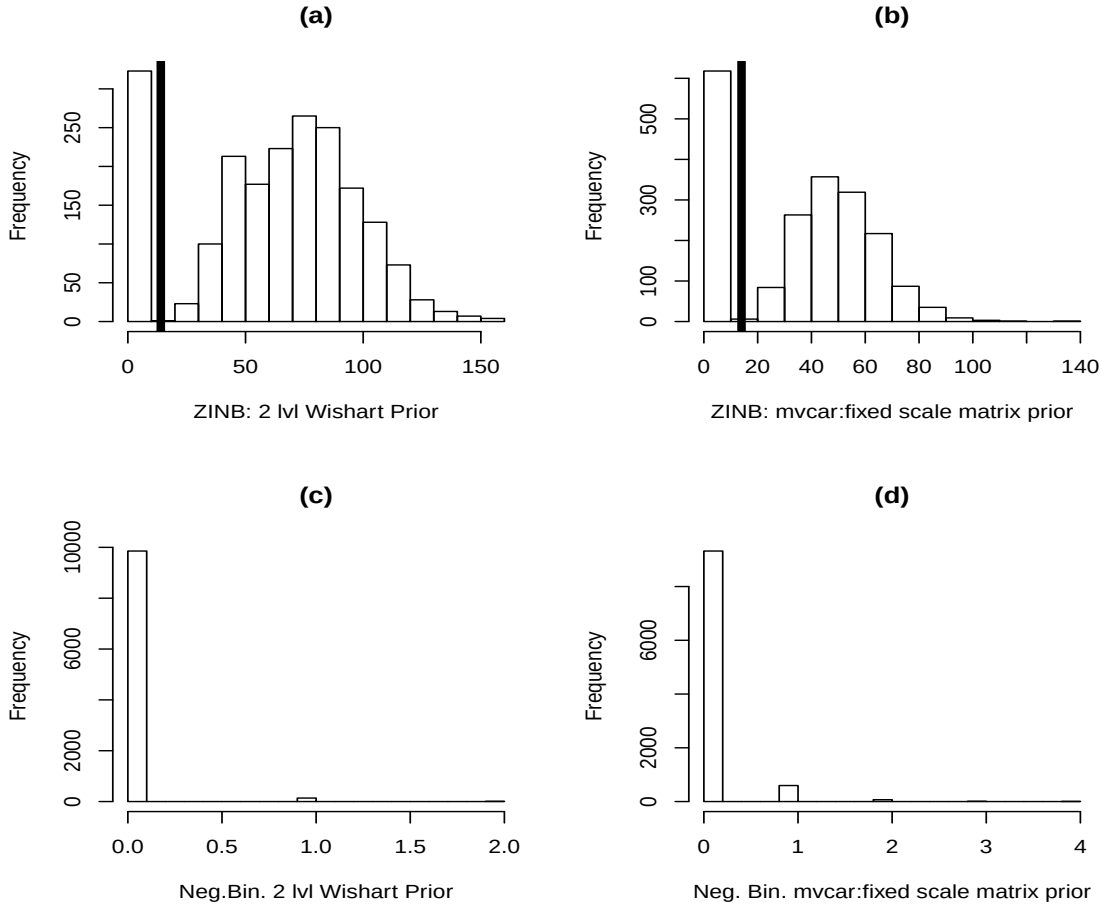
Posterior predictve Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 38, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 38, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



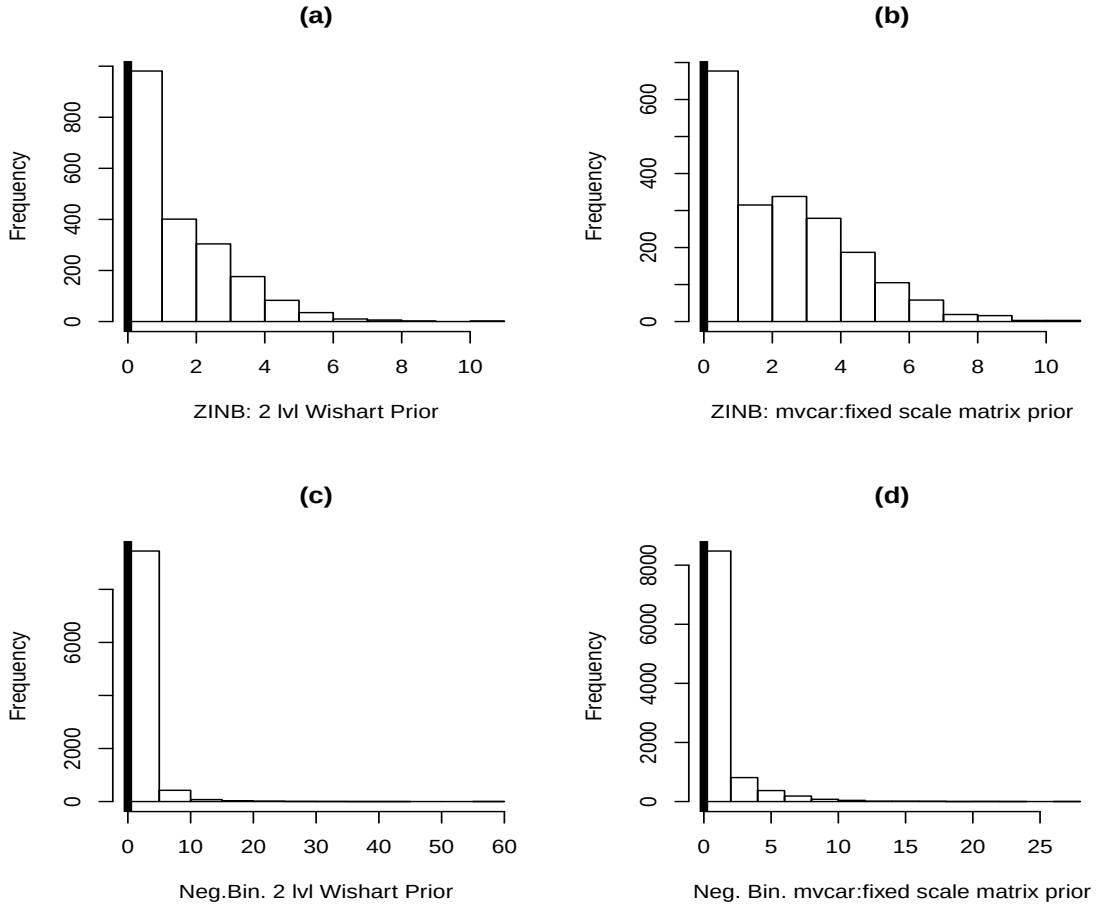
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 54, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 54, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



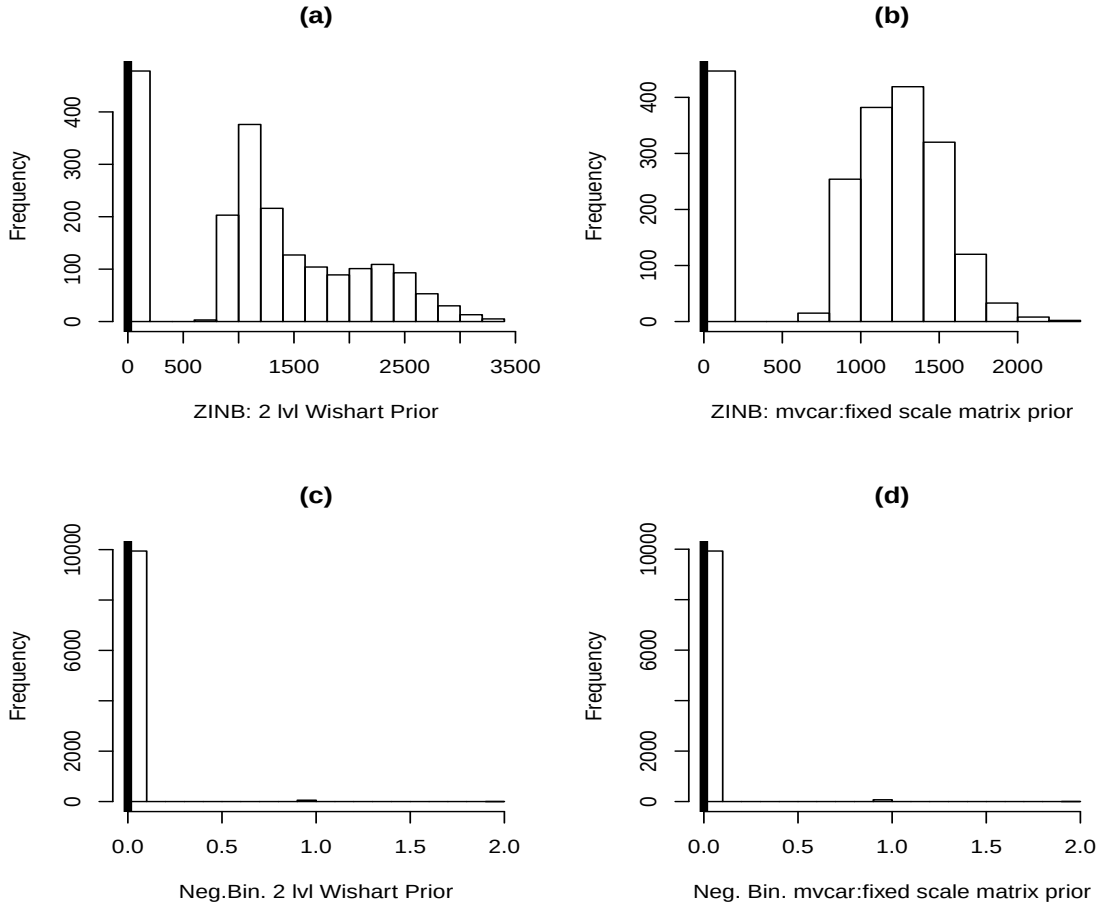
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 60, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 60, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



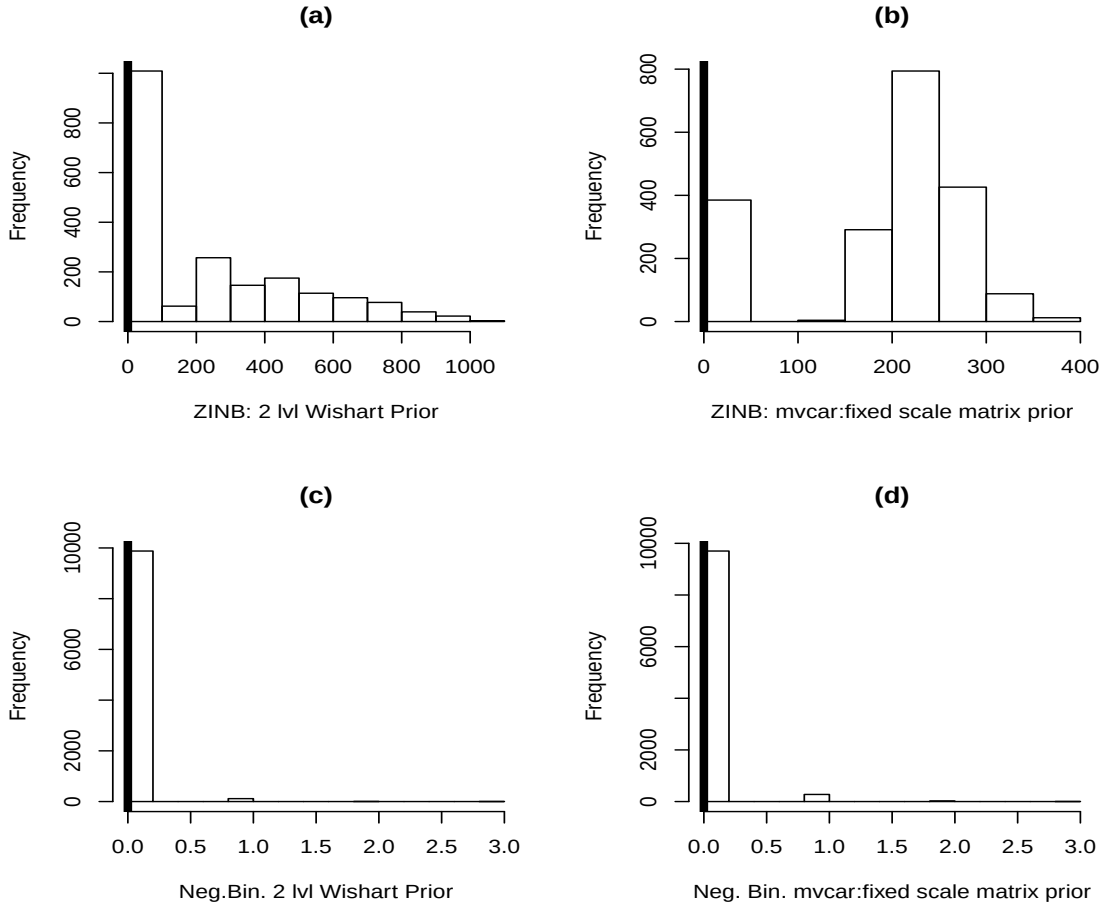
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 73, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 73, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



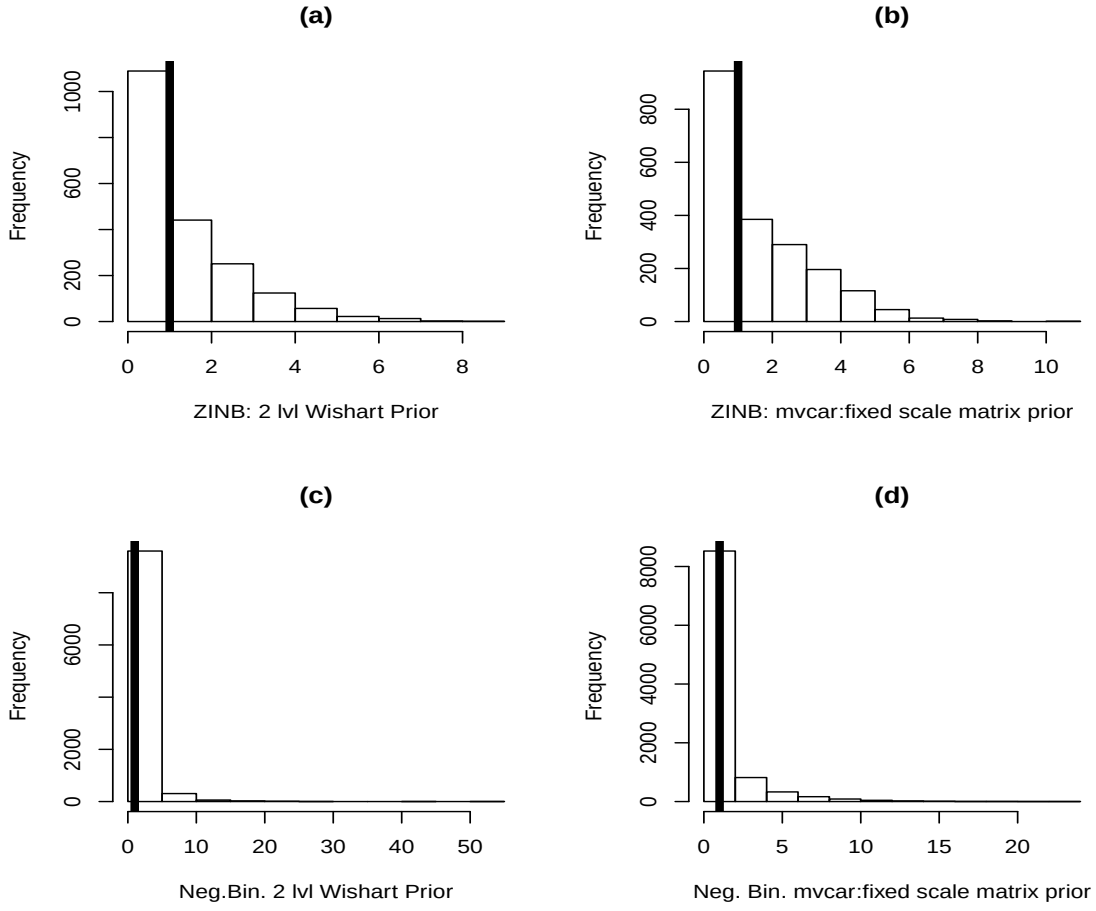
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 94, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 94, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



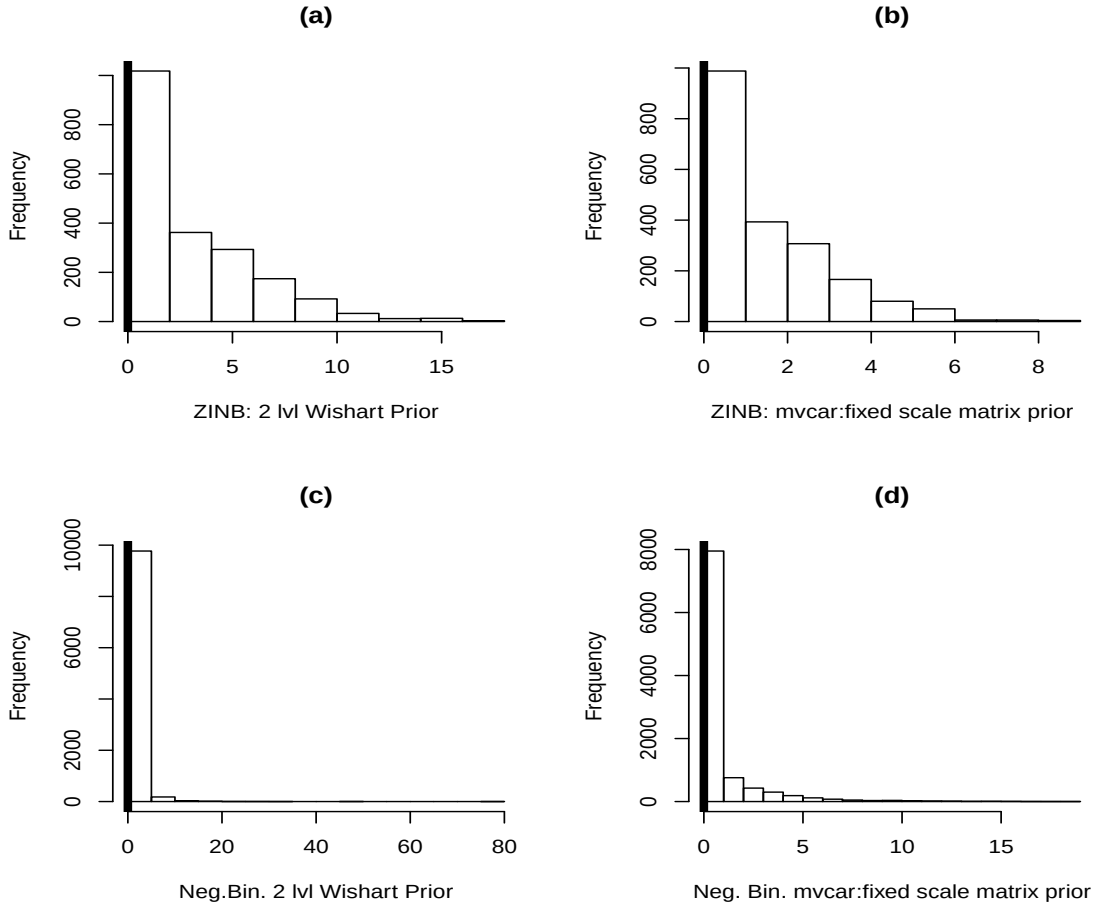
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 97, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 97, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



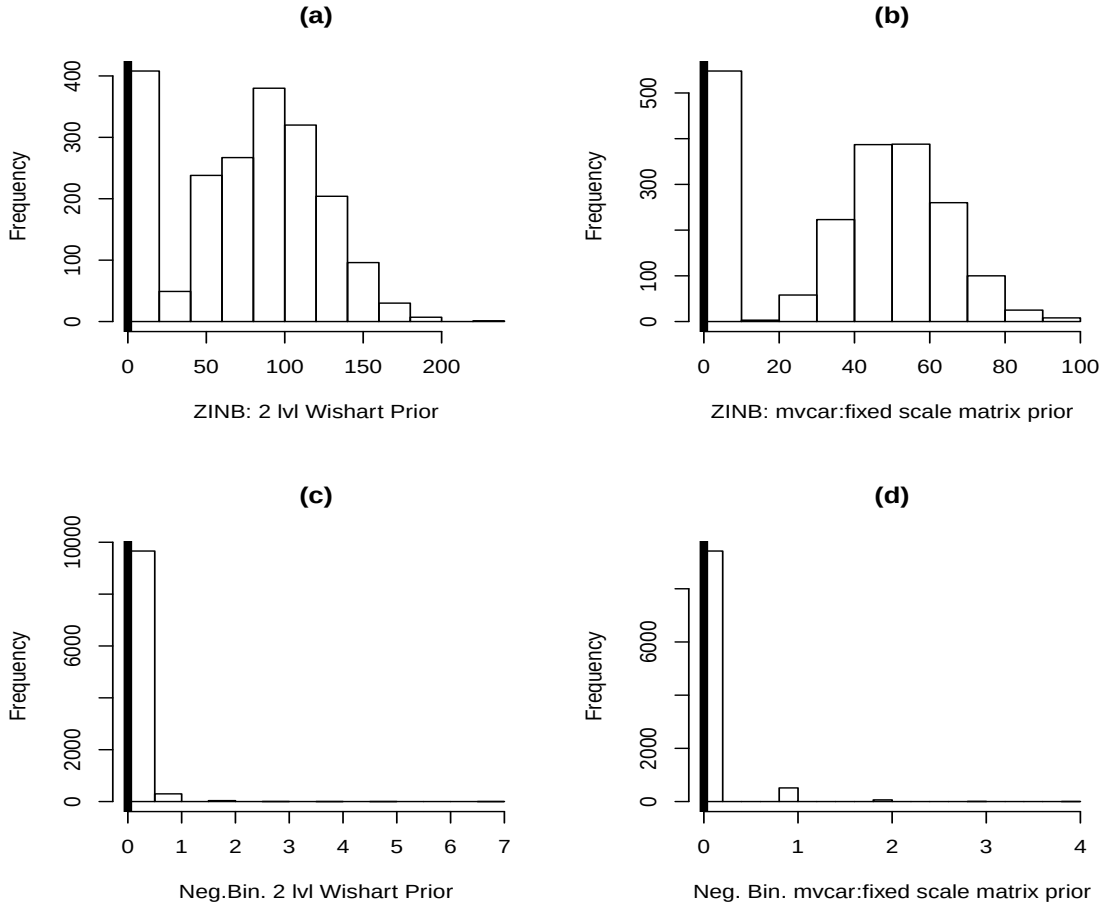
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 156, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 156, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



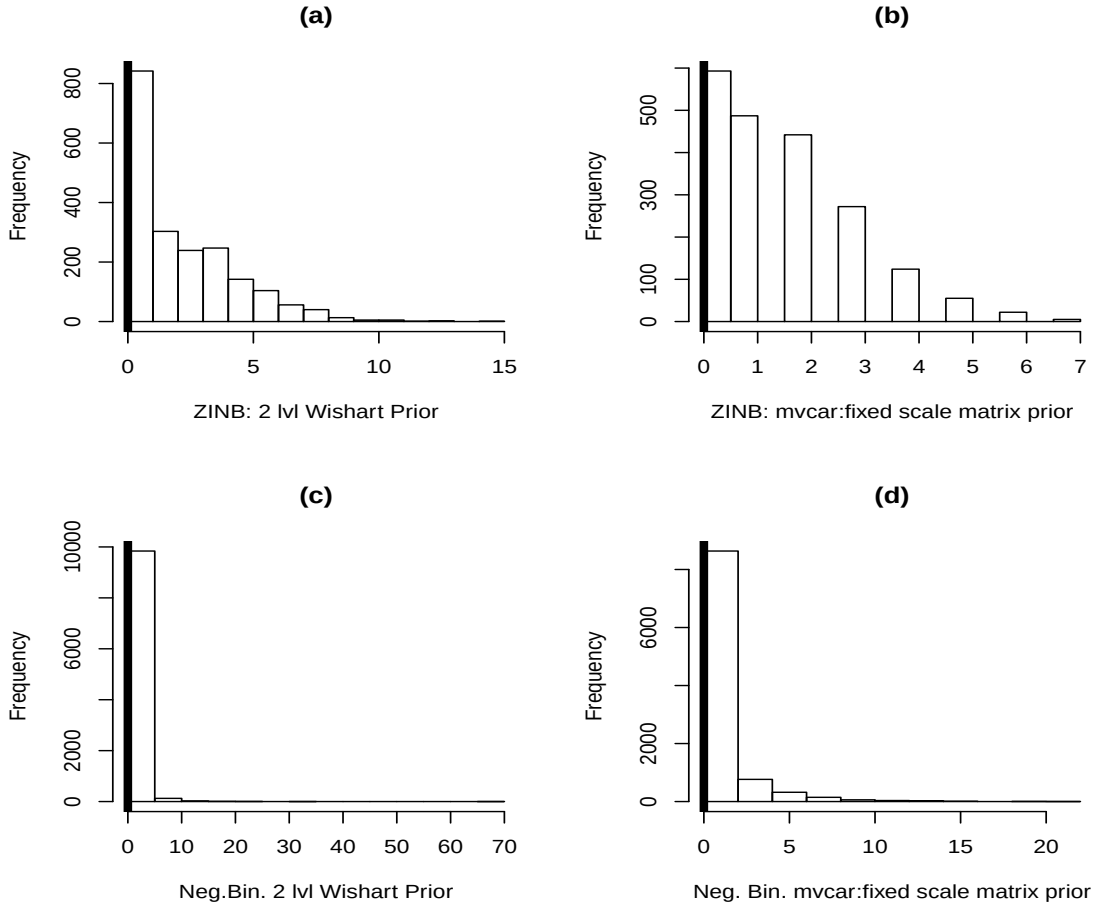
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 203, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 203, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



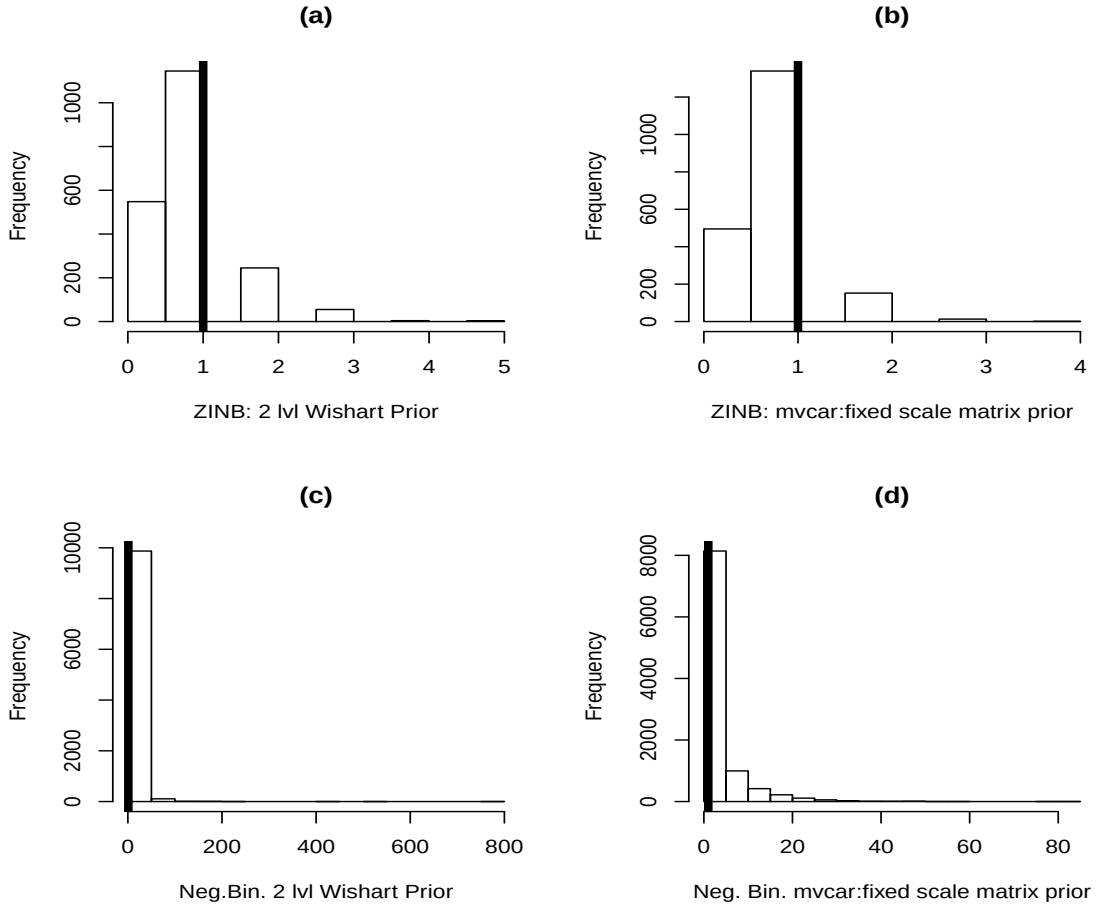
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 330, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 330, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



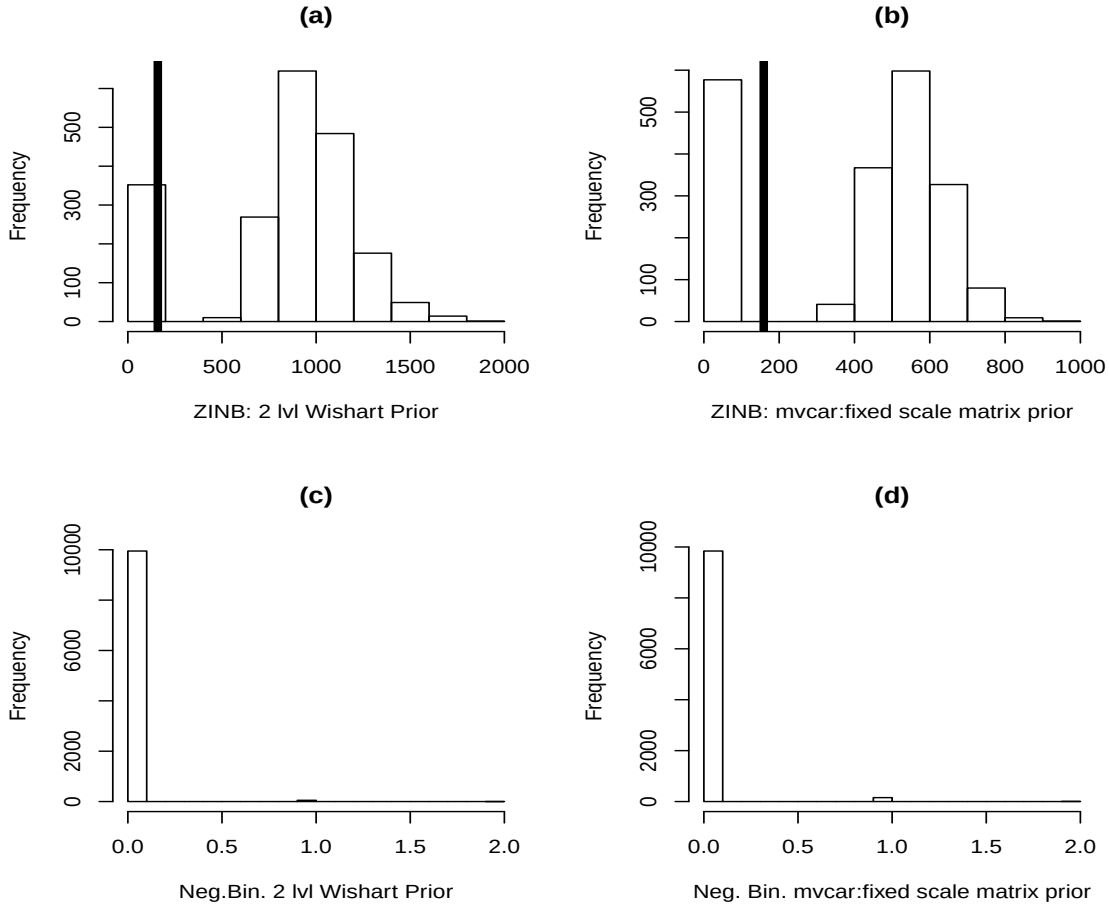
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 420, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 420, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



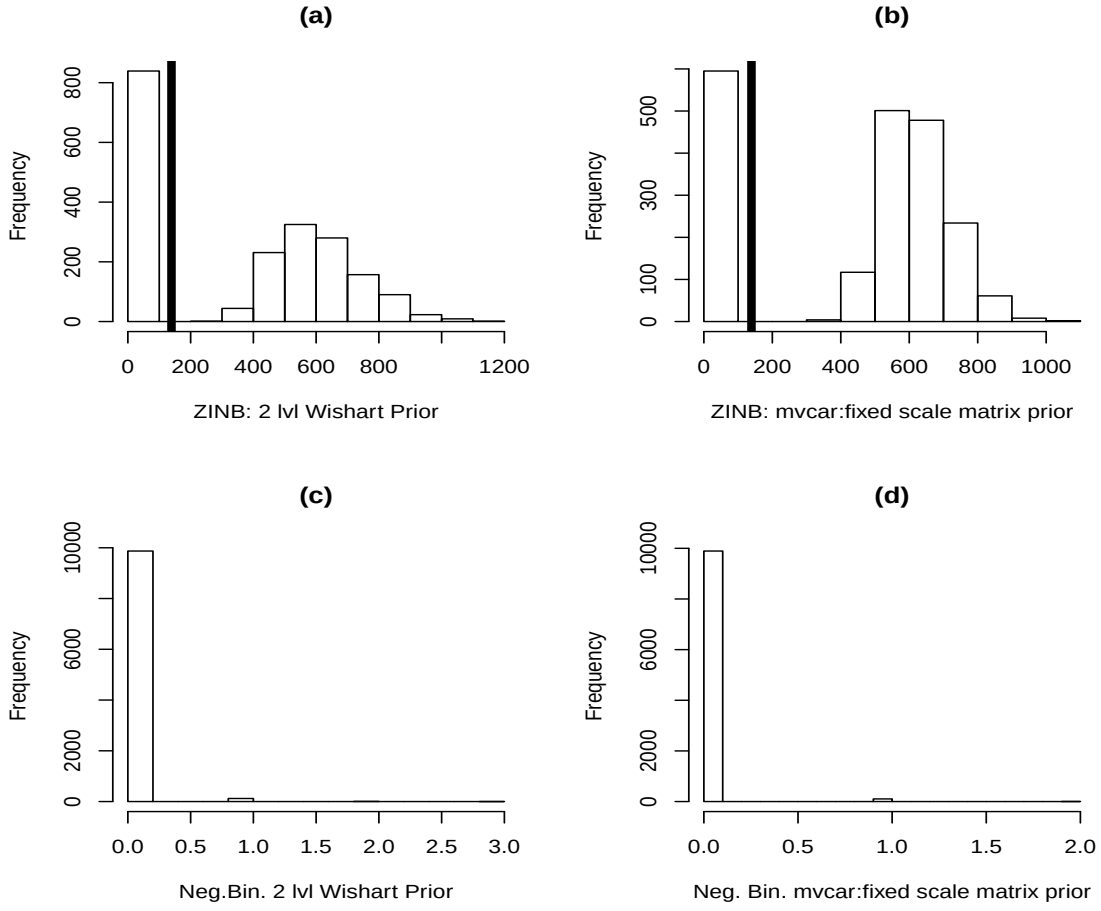
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 528, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 528, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



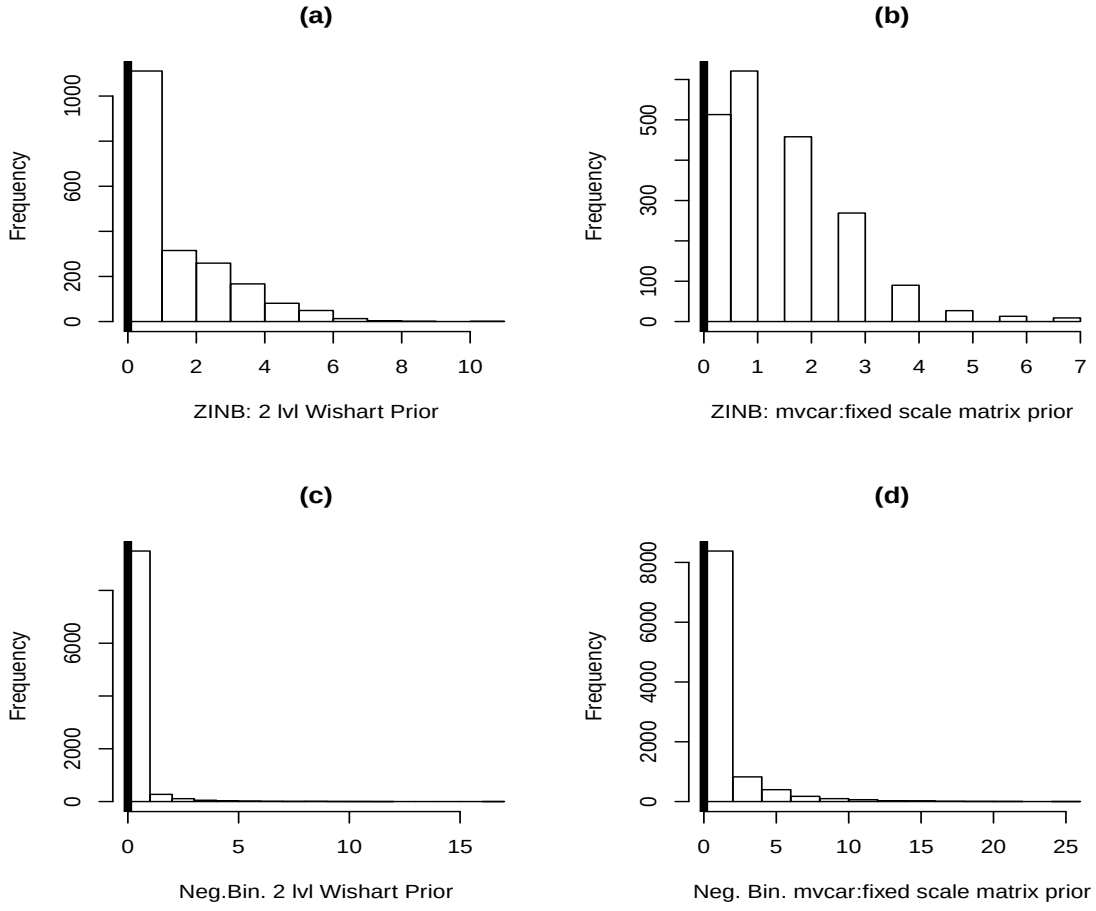
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 161, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 161, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



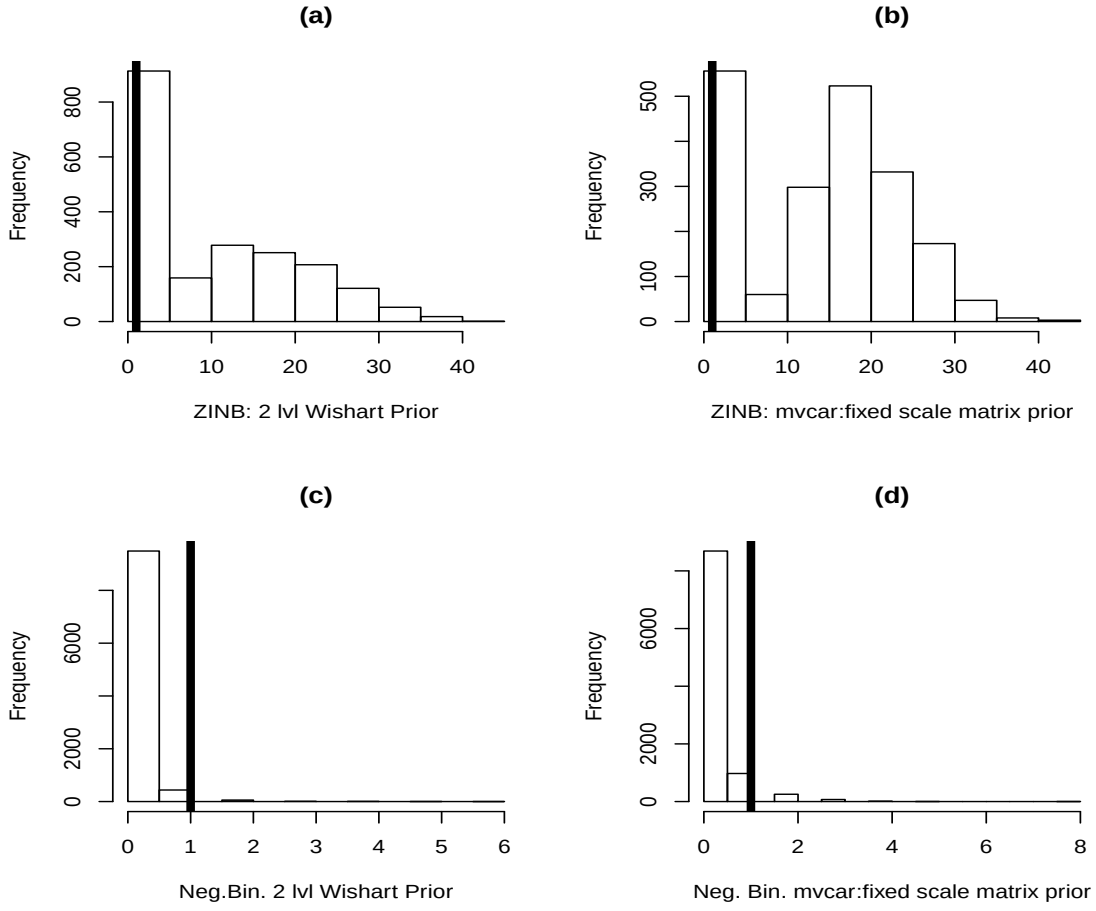
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 735, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 735, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



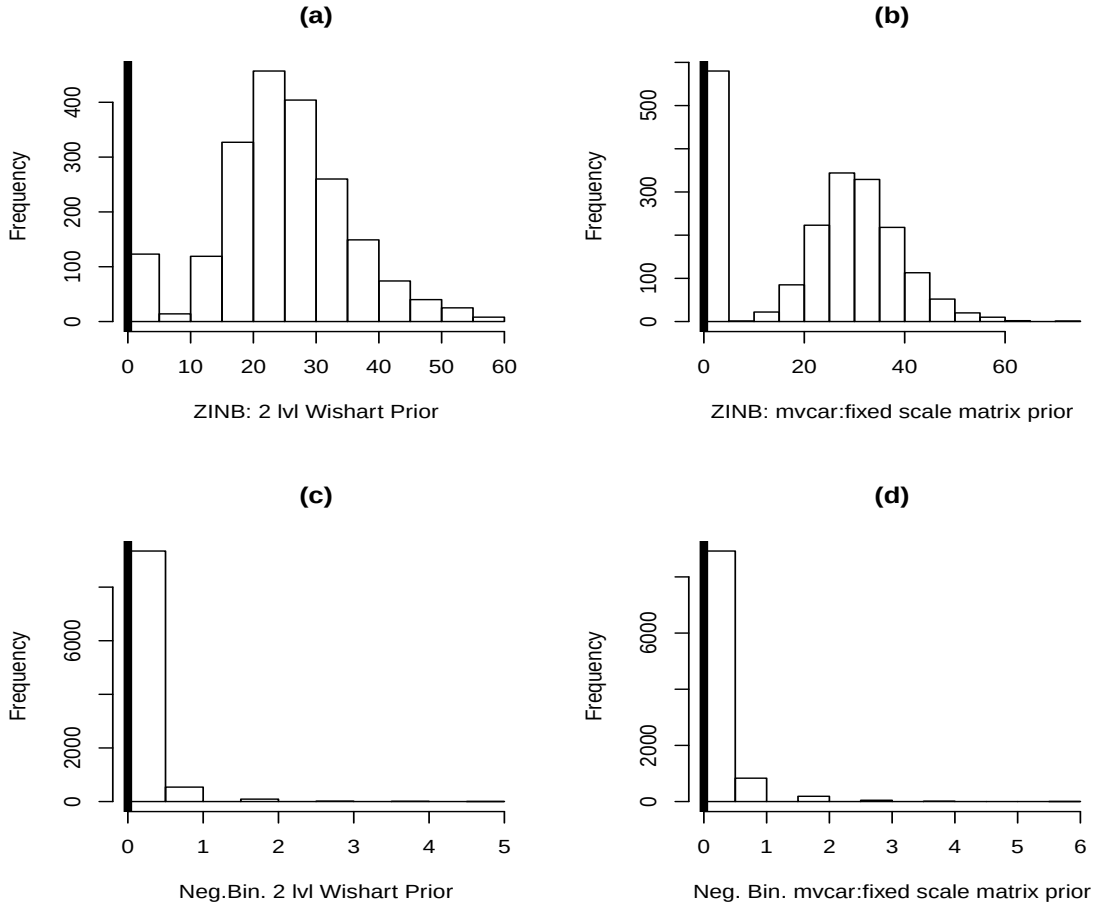
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 744, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 744, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



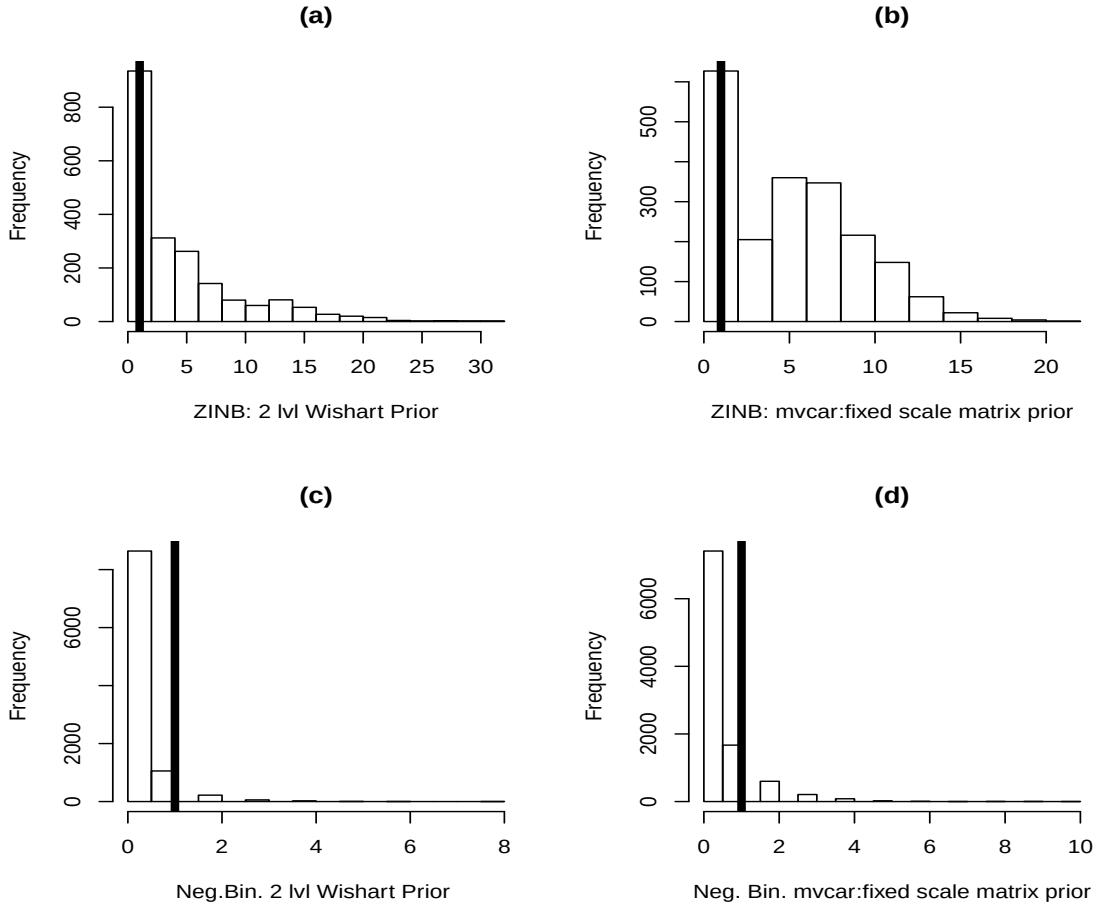
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 795, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 795, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



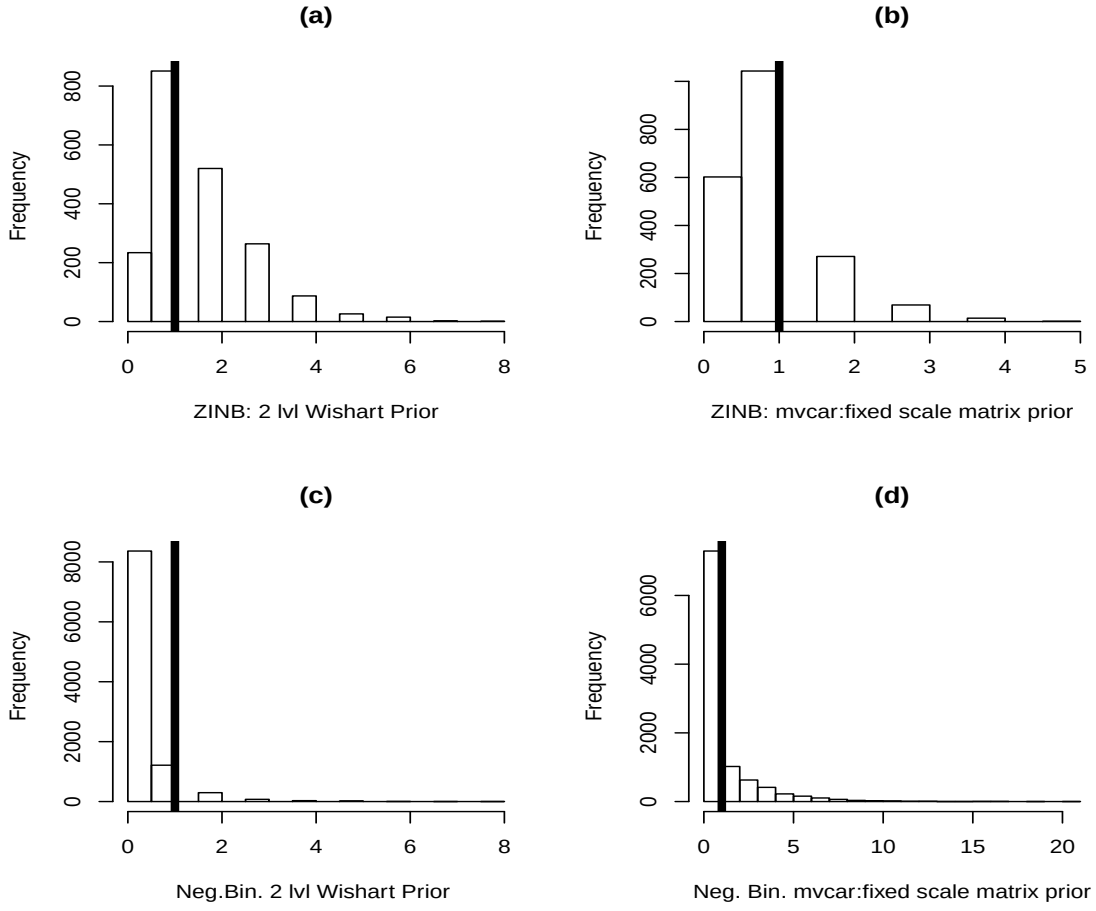
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 890, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_1 for ZINBH and Negative Binomial at site 890, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



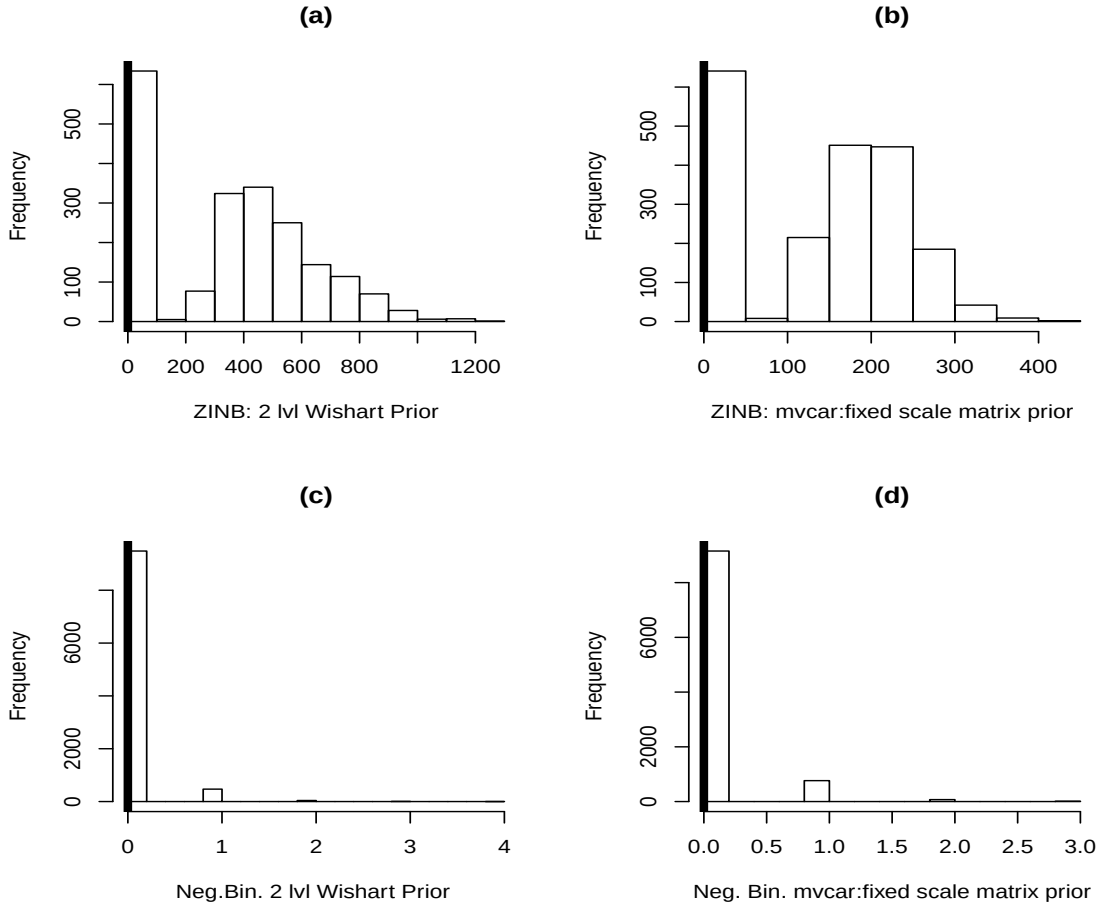
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 72, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 72, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



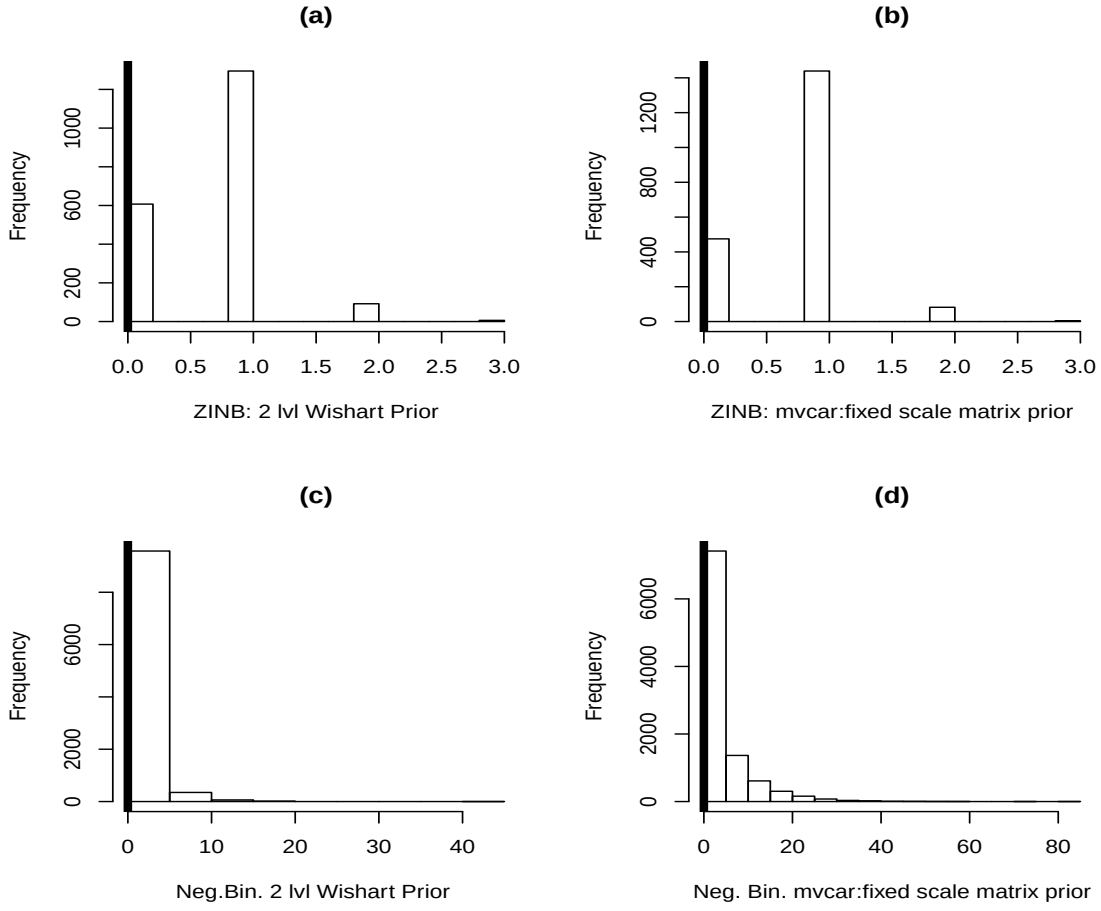
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 141, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 141, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



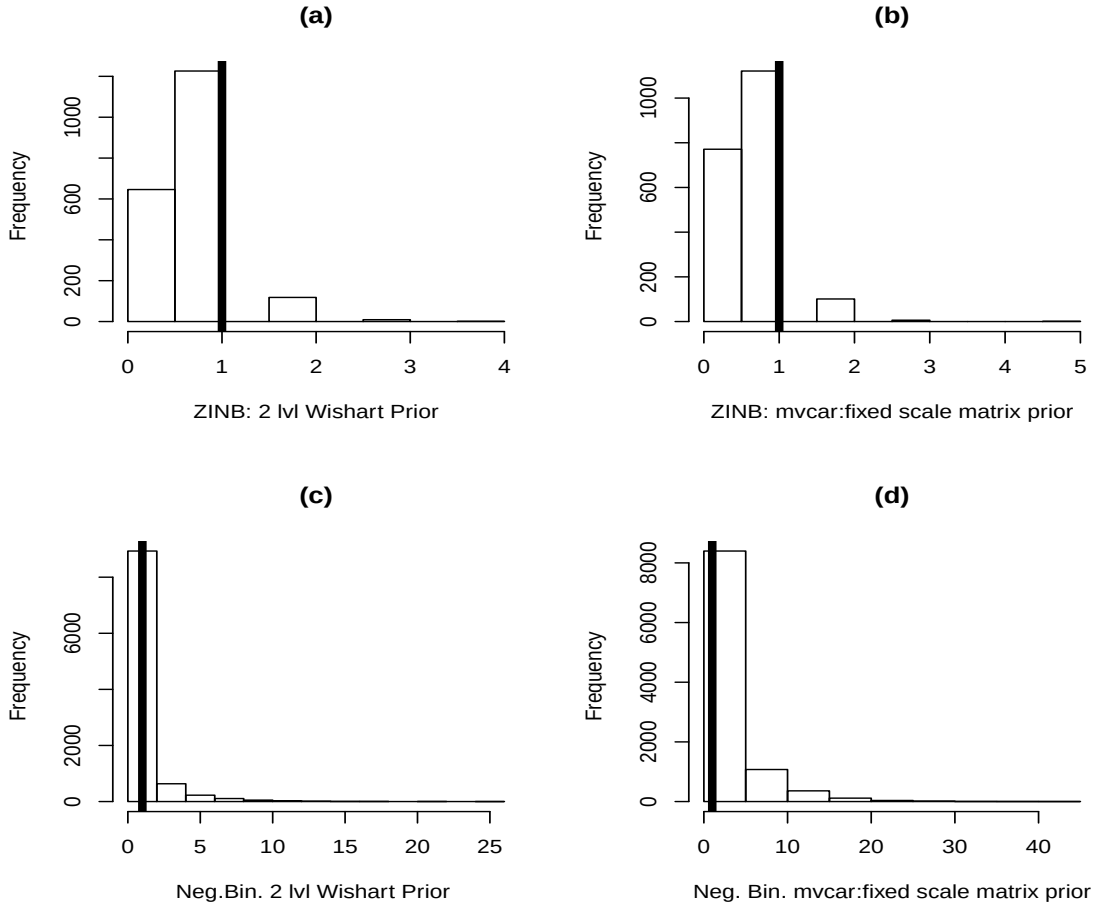
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 240, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 240, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



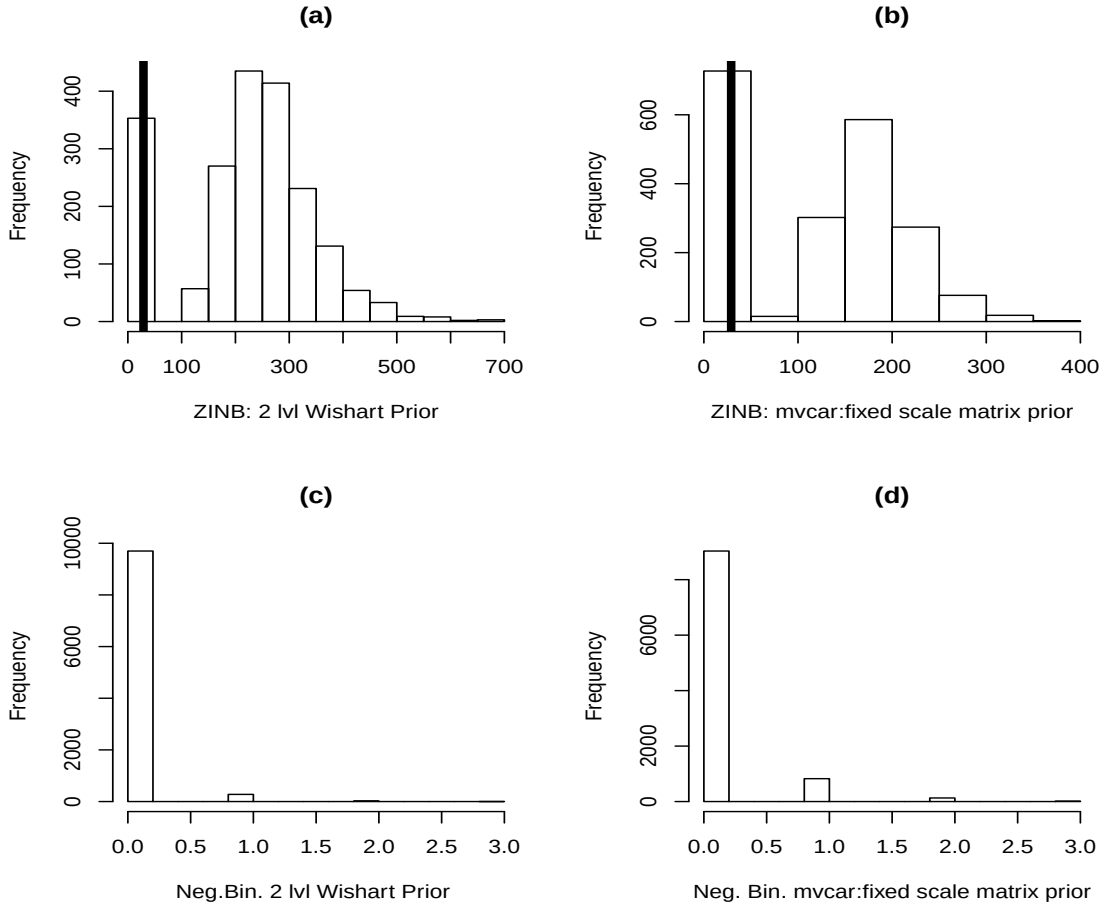
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 305, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 305, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



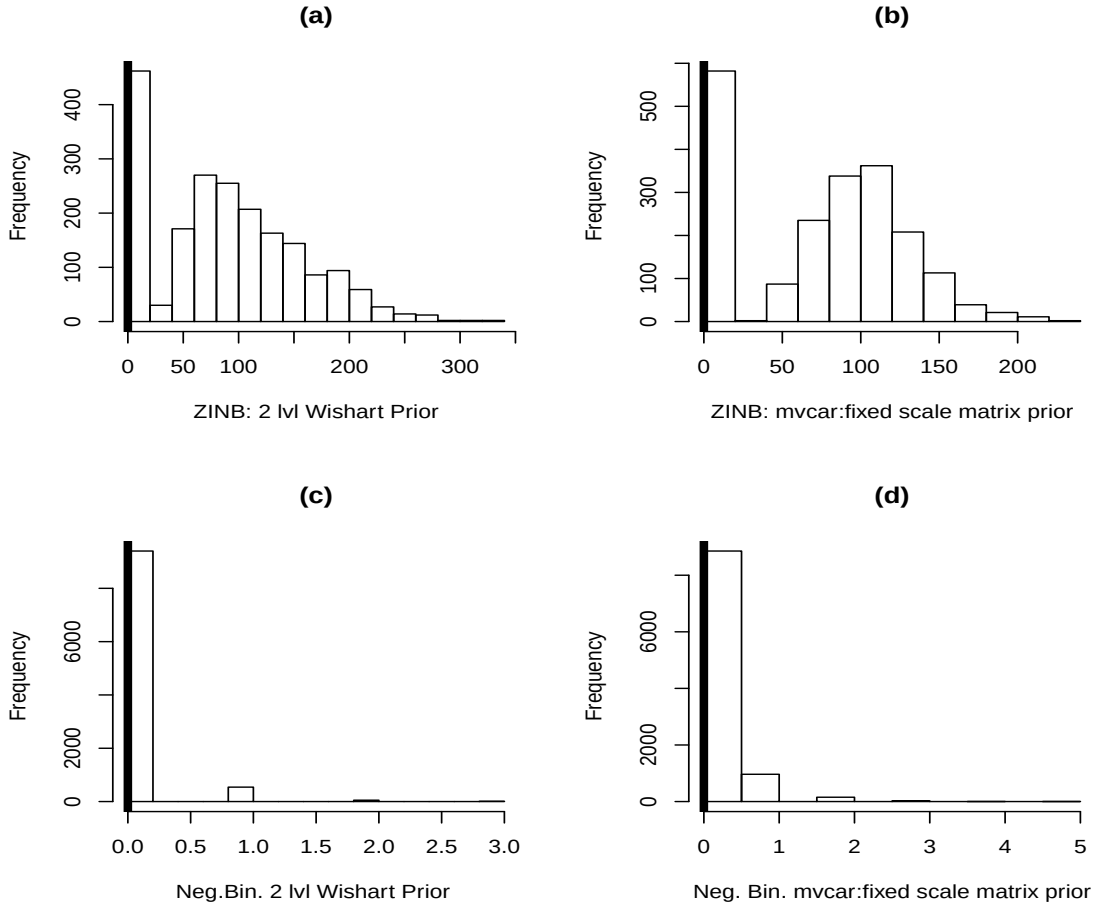
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 312, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 312, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



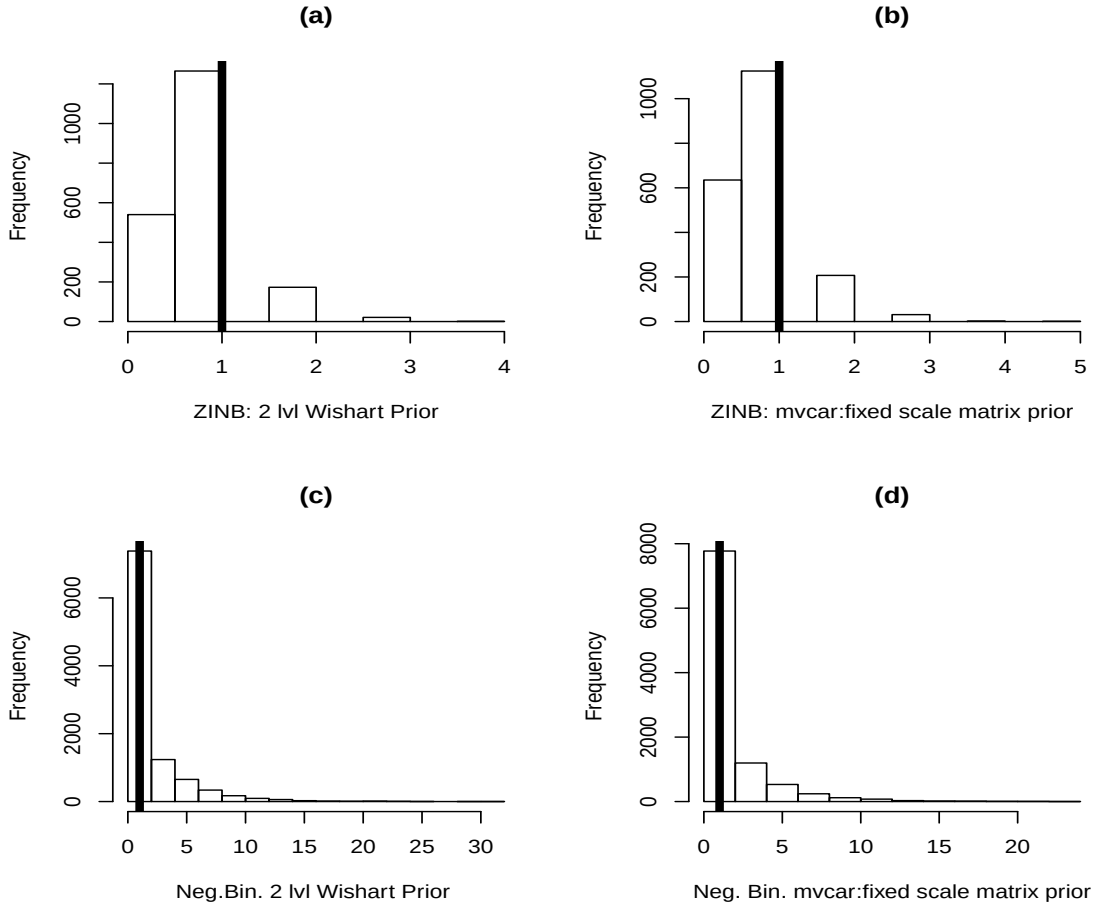
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 335, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 335, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



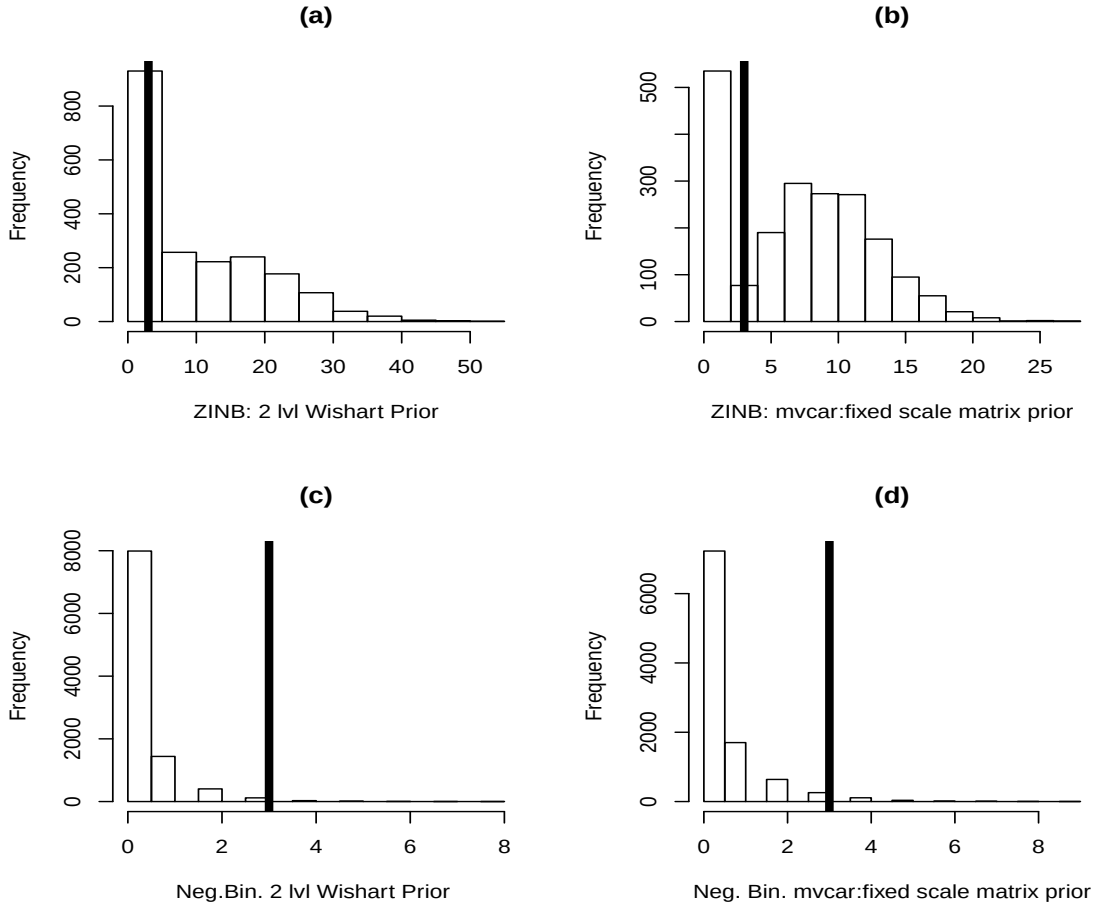
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 447, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 447, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



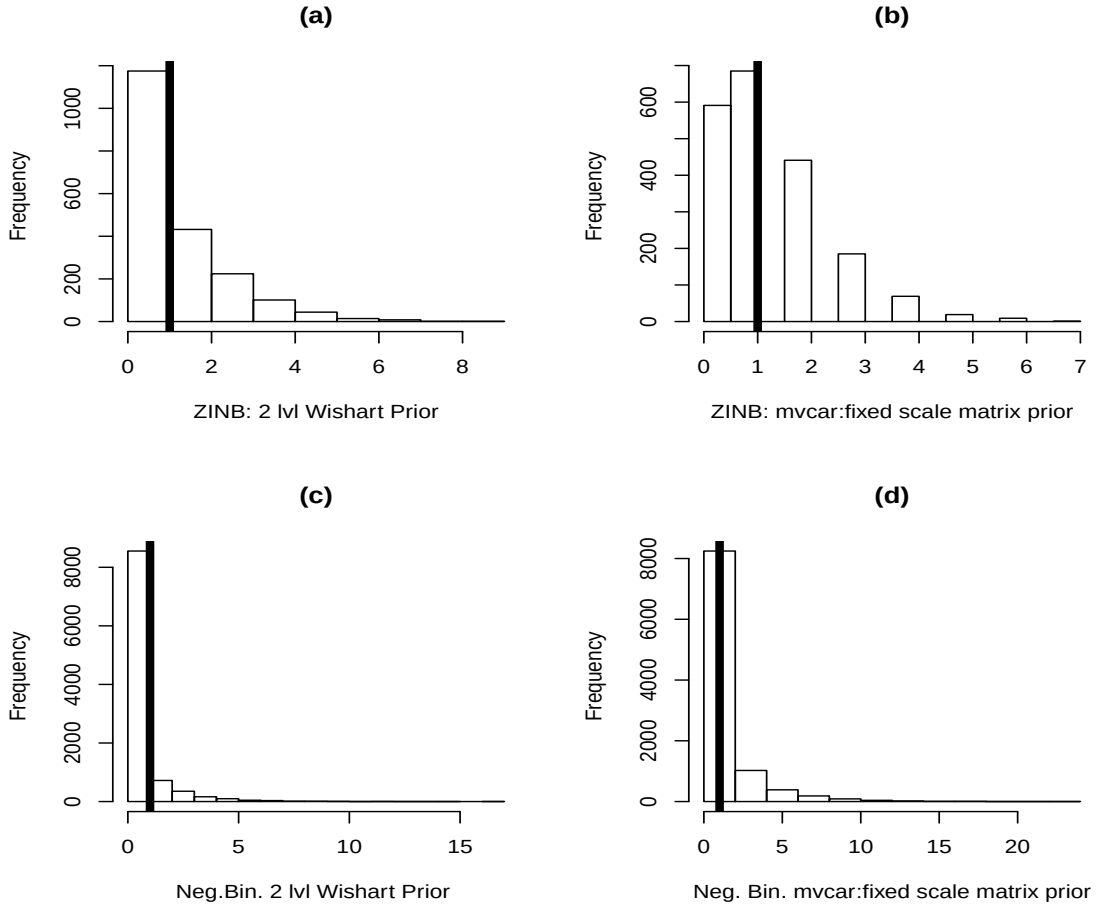
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 530, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 530, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



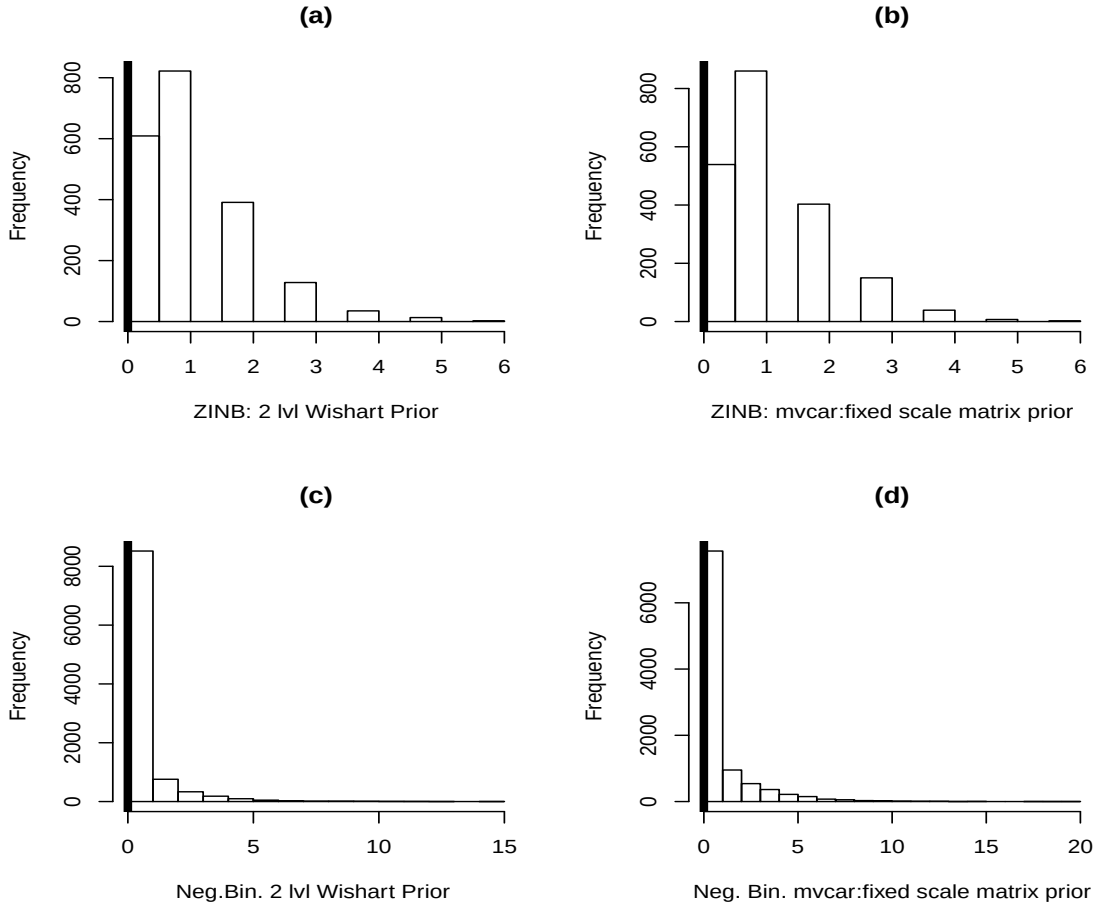
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 593, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 593, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



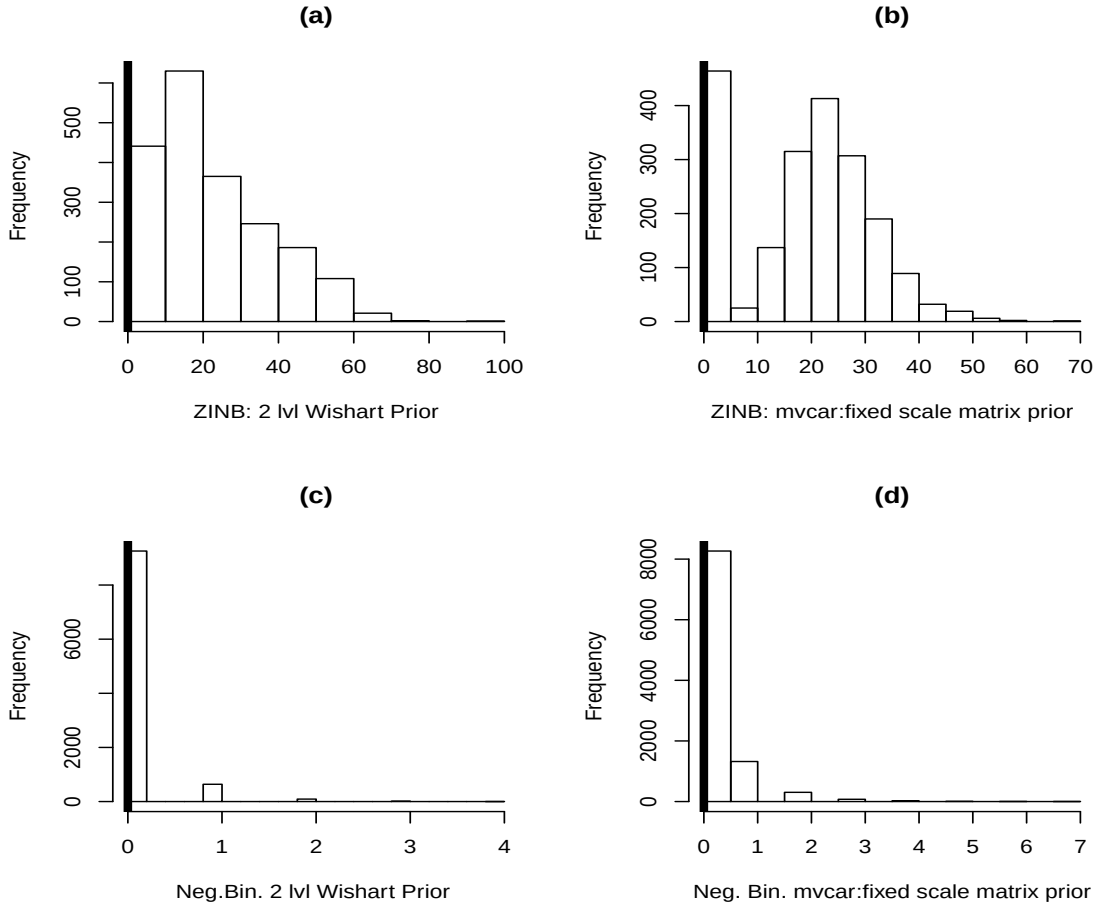
Posterior Predictve Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 625, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 625, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



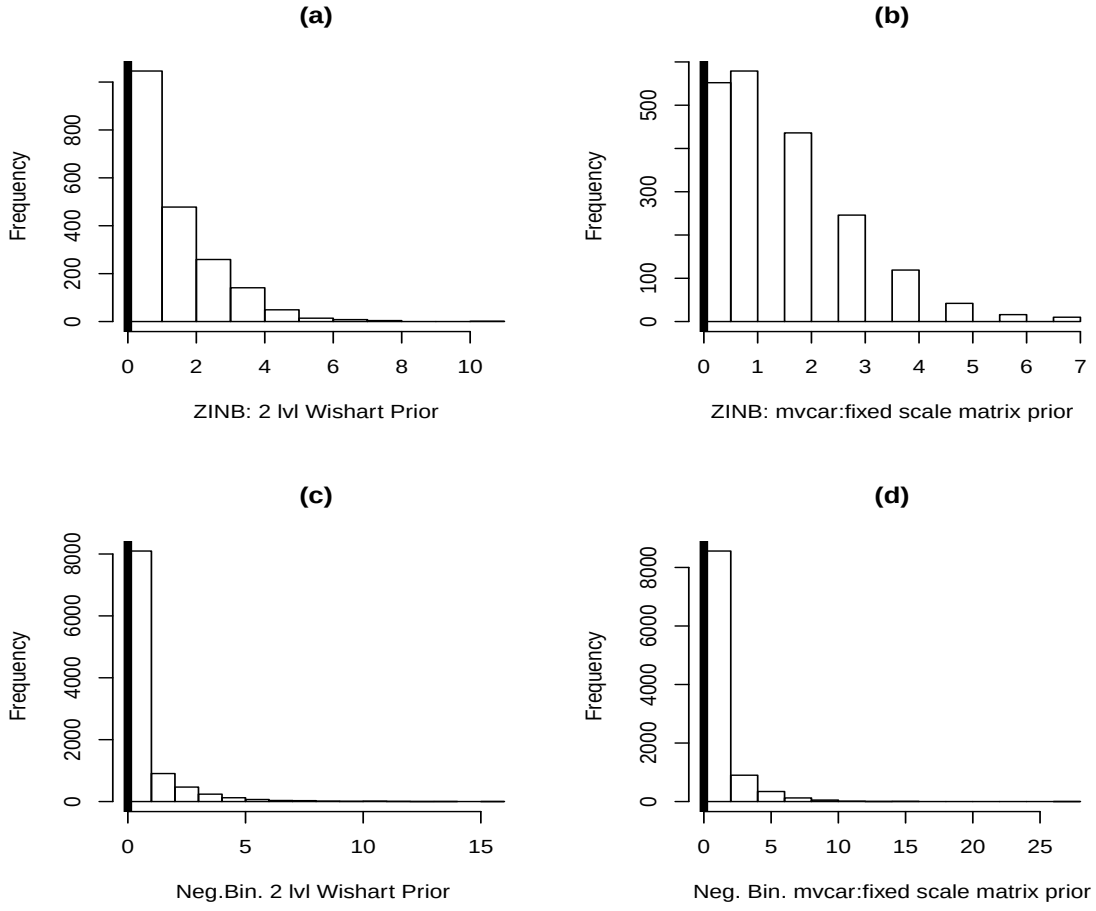
Posterior Predictve Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 688, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 688, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



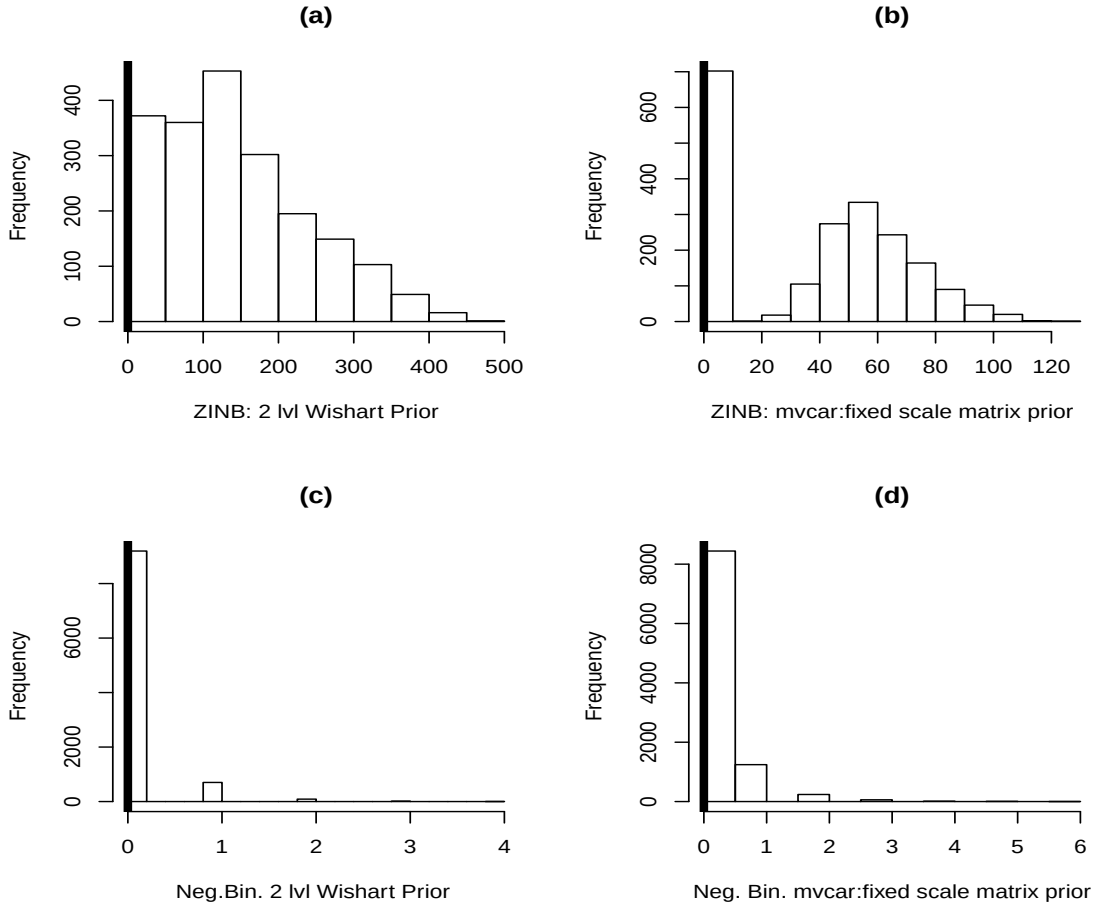
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 712, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 712, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



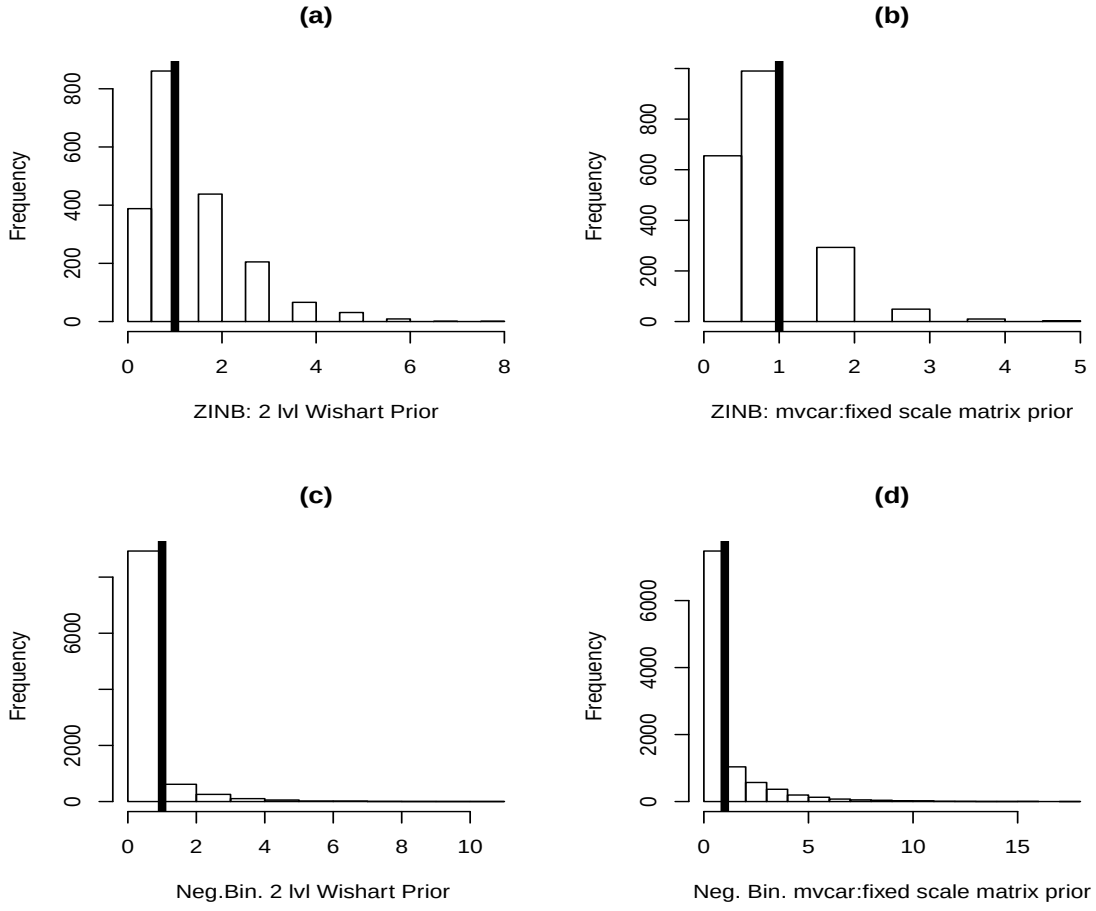
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 773, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 773, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



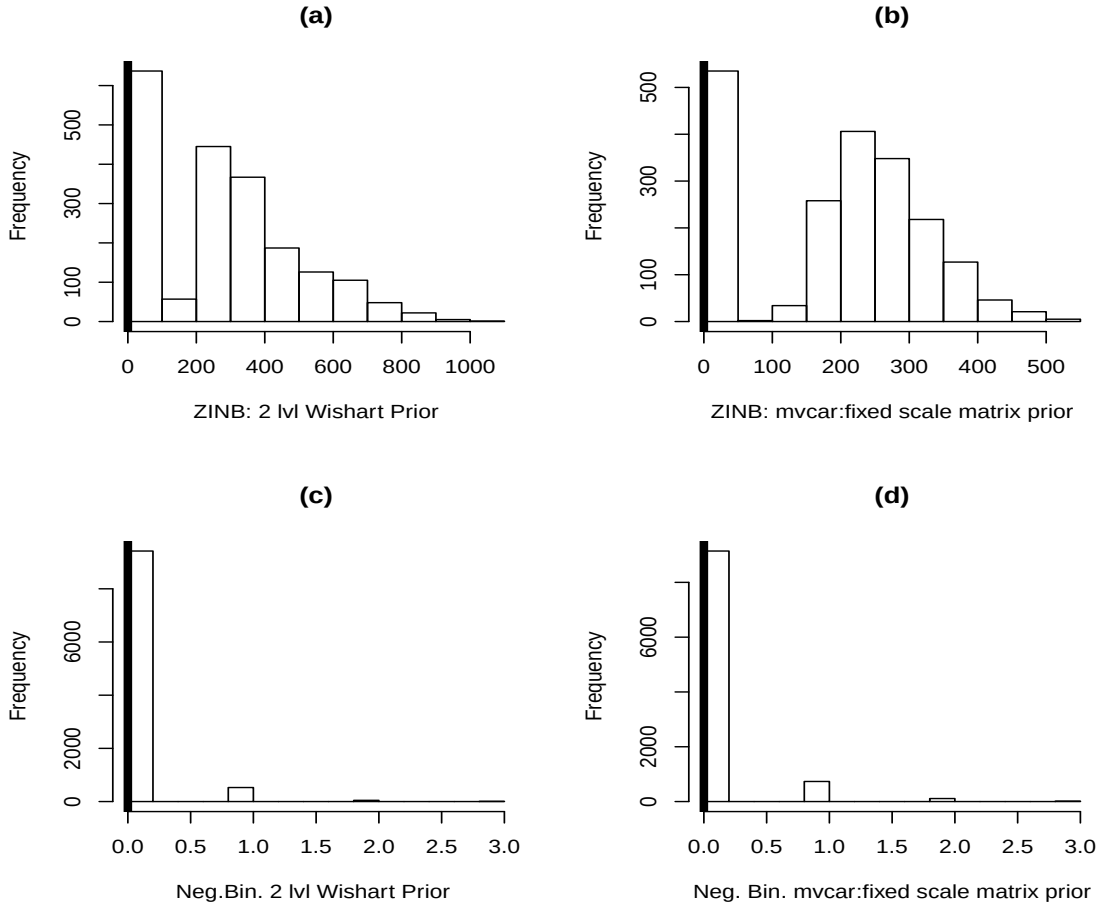
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 877, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 877, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



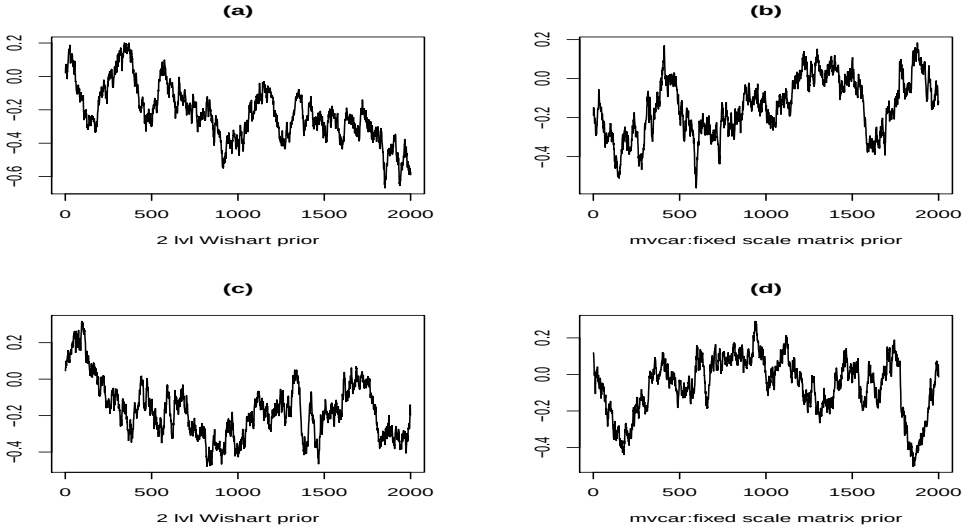
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 971, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 971, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



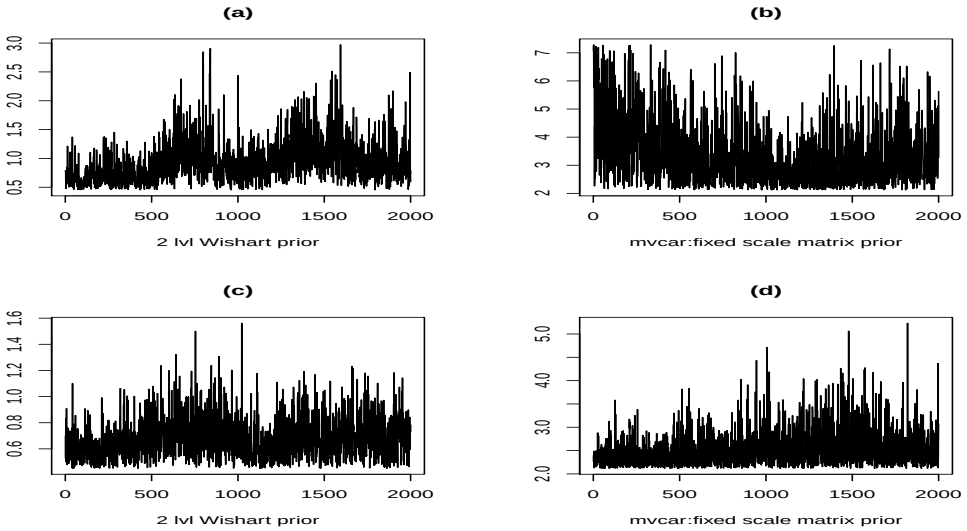
Posterior Predictive Plots for Simulated Data: (a), (c): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 992, respectively under two-level inverse-Wishart prior. (b), (d): The histograms show the posterior predictive distribution of simulated data of Y_2 for ZINBH and Negative Binomial at site 992, respectively under fixed inverse-Wishart prior. The vertical lines represent the observed values.



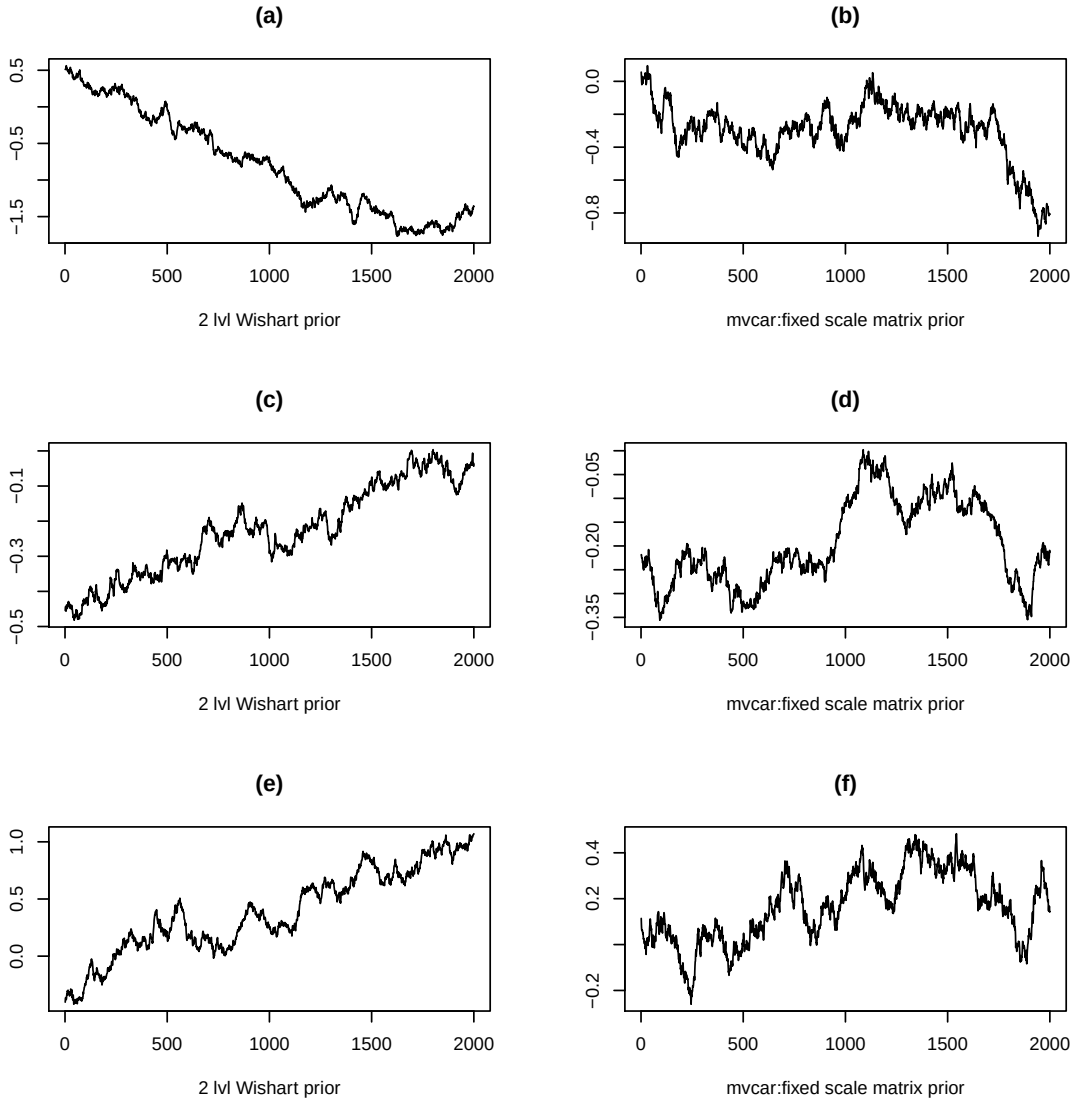
MCMC trace plots for occurrence regression coefficient for MI Crime Data: (a), (b): the graphs show the trace of iteration values of α for Murder and Non-negligent Manslaugher for ZINBH under TLHIW and fixed scale matrix prior models, respectively. (c), (d): show the trace of interation values of α for Buglary Forced Entry for ZINBH under TLHIW and fixed scale matrix prior models, respectively.



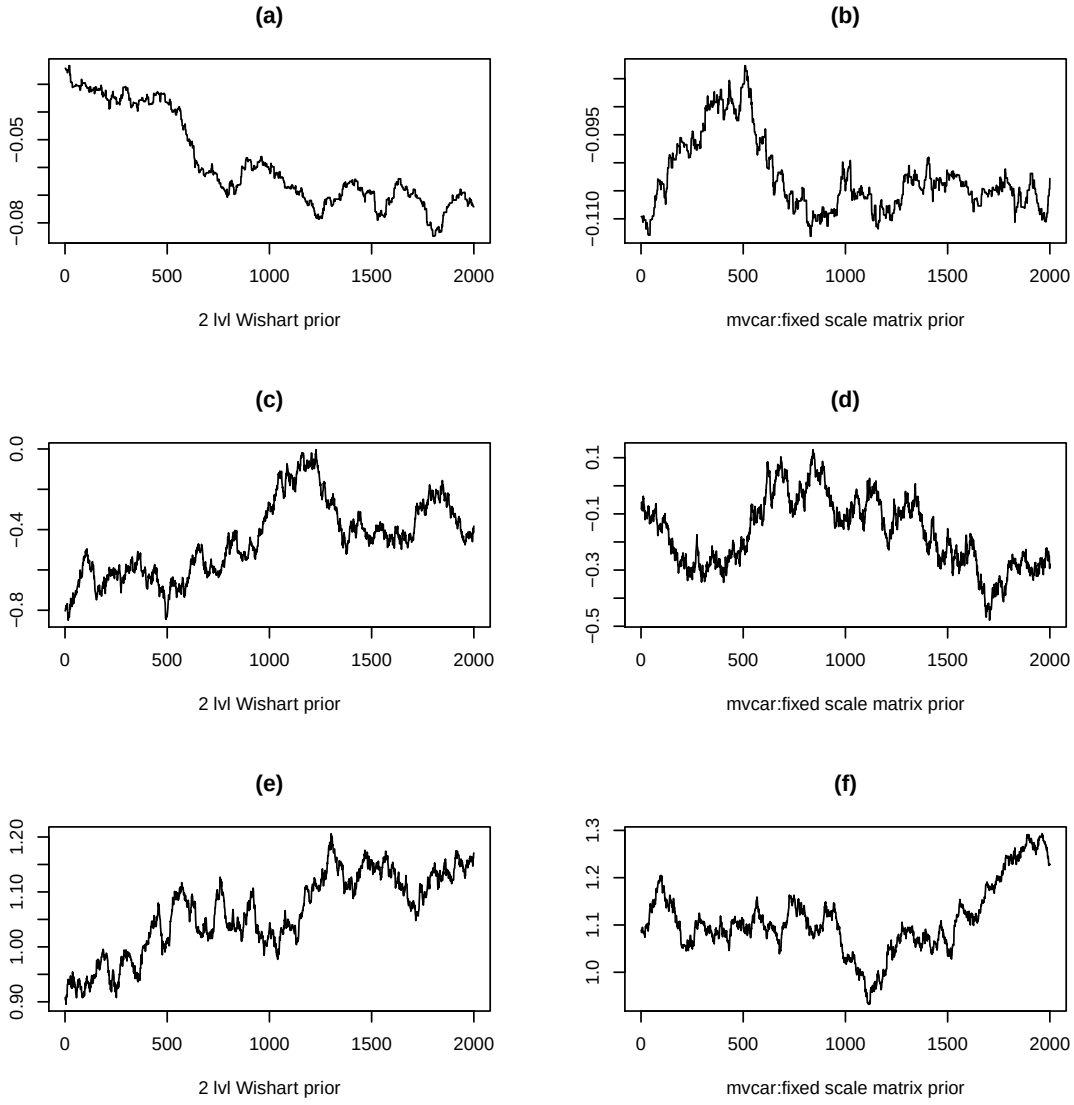
MCMC trace plots for dispersion parameters for MI Crime Data: (a), (b): the graphs show the trace of iteration values of α_{nb} for Murder and Non-negligent Manslaugher for ZINBH under TLHIW and fixed scale matrix prior models, respectively. (c), (d): show the trace of interation values of α_{nb} for Buglary Forced Entry for ZINBH under TLHIW and fixed scale matrix prior models, respectively.



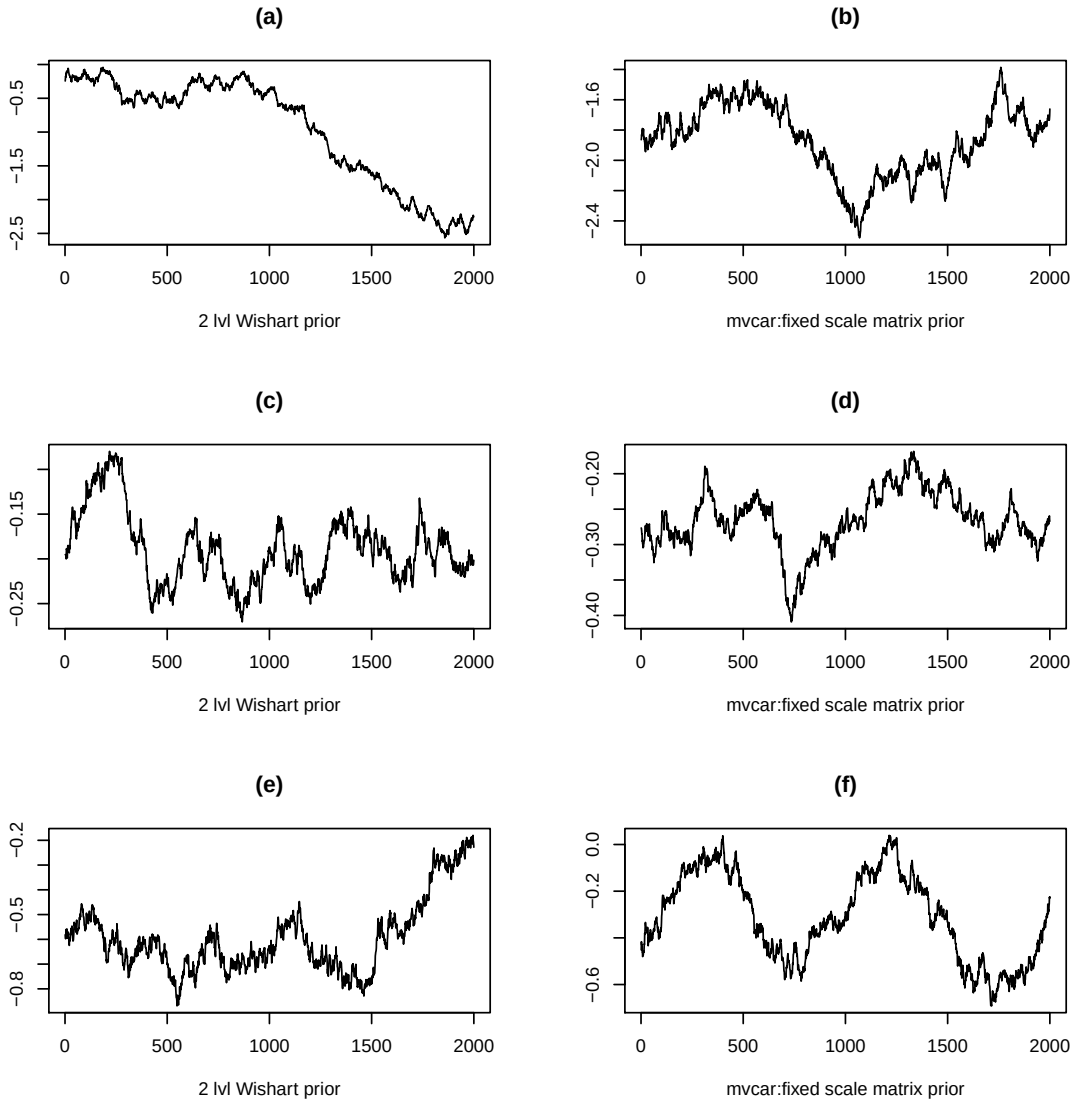
MCMC trace plots for abundance regression coefficients parameters for MI Crime Data, Murder and Non-negligent Manslaughter : (a), (b): the graphs show the trace of iteration values of β_1 for ZINBH under TLHIW and fixed scale matrix prior models, respectively. (c), (d): show the trace of interation values of β_2 for ZINBH under TLHIW and fixed scale matrix prior models, respectively.(e), (f): show the trace of interation values of β_3 for ZINBH under TLHIW and fixed scale matrix prior models, respectively.



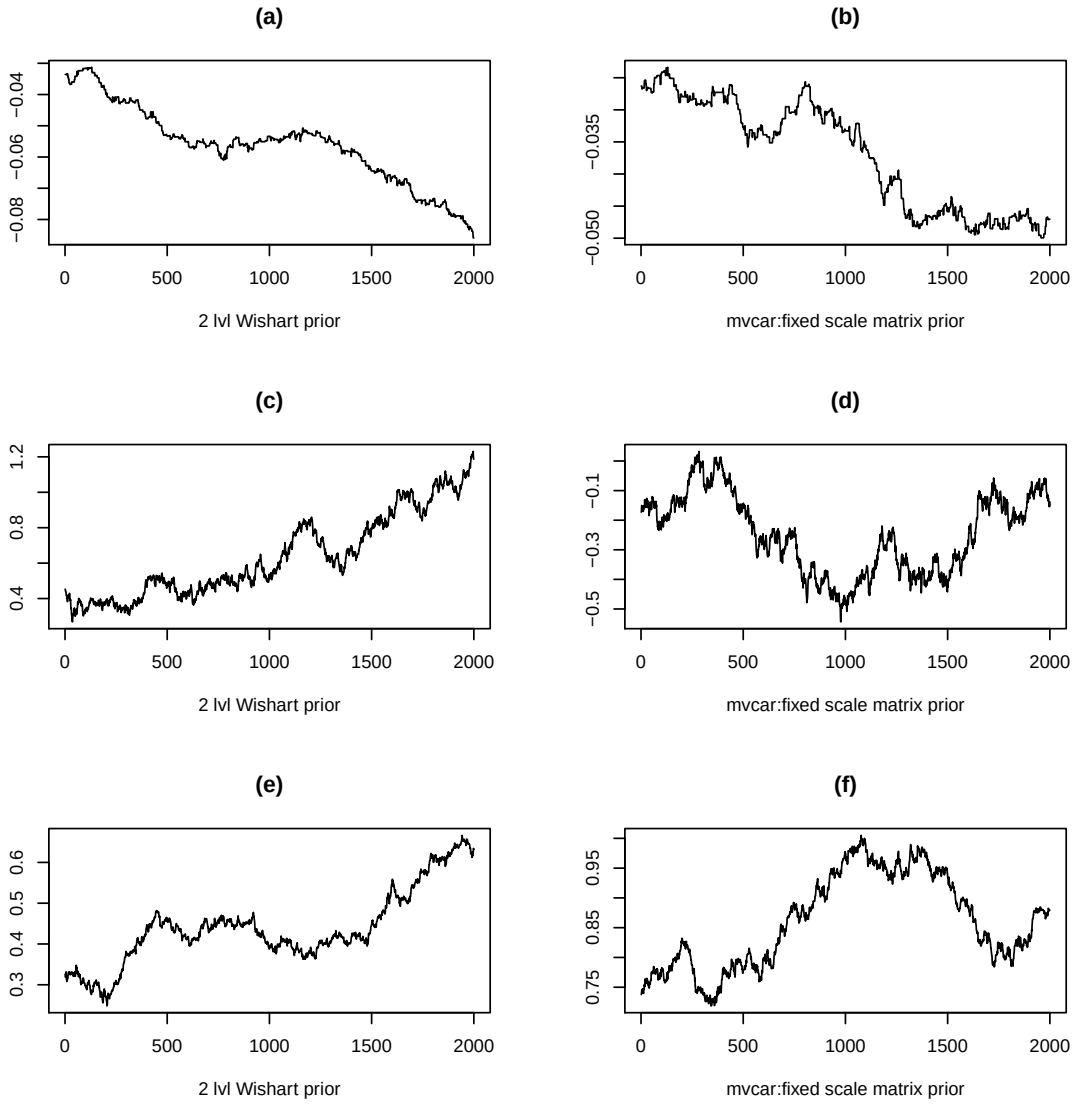
MCMC trace plots for abundance regression coefficients parameters for MI Crime Data, Murder and Non-negligent Manslaughter : (a), (b): the graphs show the trace of iteration values of β_4 for ZINBH under TLHIW and fixed scale matrix prior models, respectively. (c), (d): show the trace of interation values of β_5 for ZINBH under TLHIW and fixed scale matrix prior models, respectively.(e), (f): show the trace of interation values of β_6 for ZINBH under TLHIW and fixed scale matrix prior models, respectively.



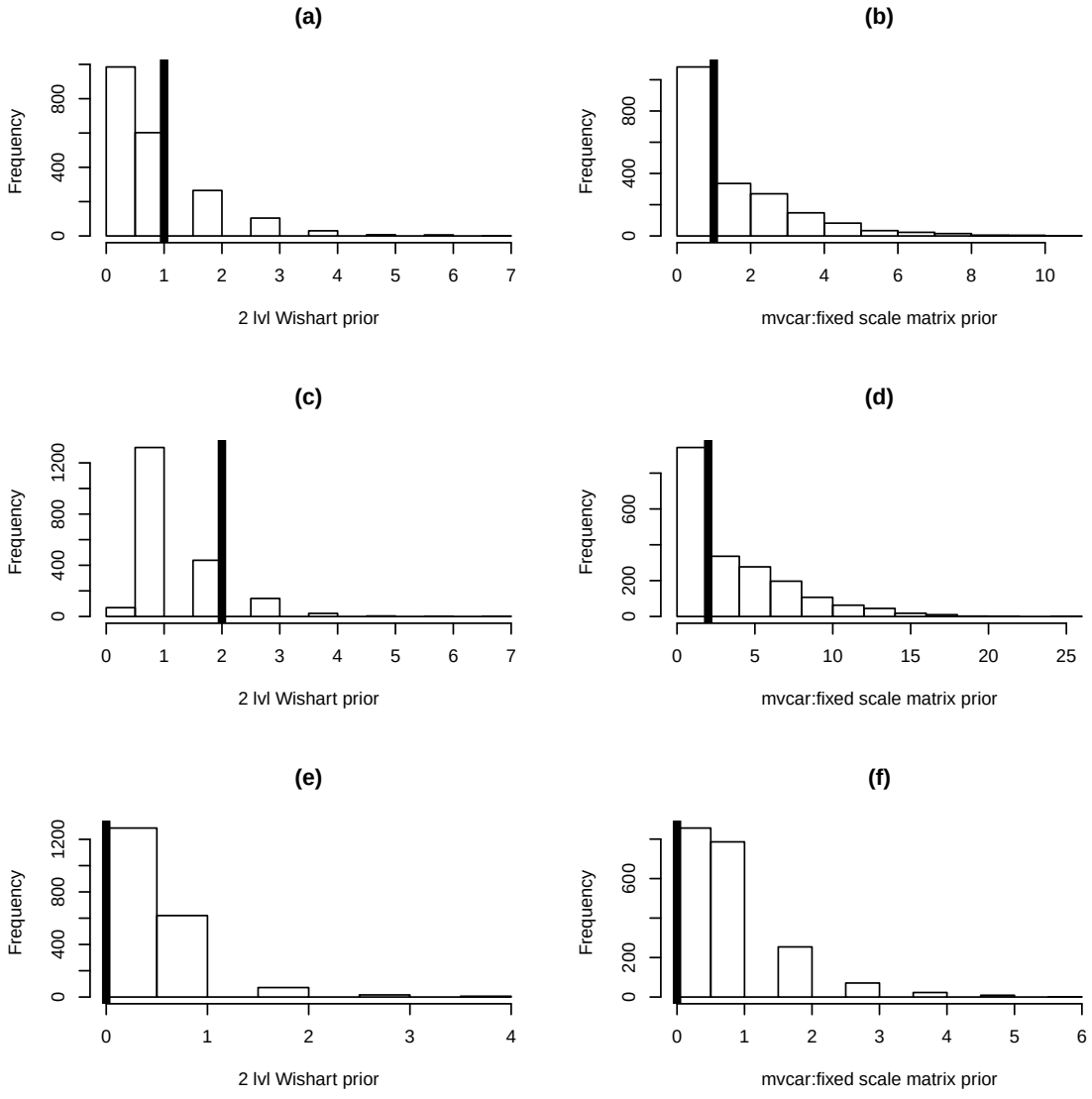
MCMC trace plots for abundance regression coefficients parameters for MI Crime Data, Burglary Forced Entry : (a), (b): the graphs show the trace of iteration values of β_1 for ZINBH under TLHIW and fixed scale matrix prior models, respectively. (c), (d): show the trace of interation values of β_2 for ZINBH under TLHIW and fixed scale matrix prior models, respectively.(e), (f): show the trace of interation values of β_3 for ZINBH under TLHIW and fixed scale matrix prior models, respectively.



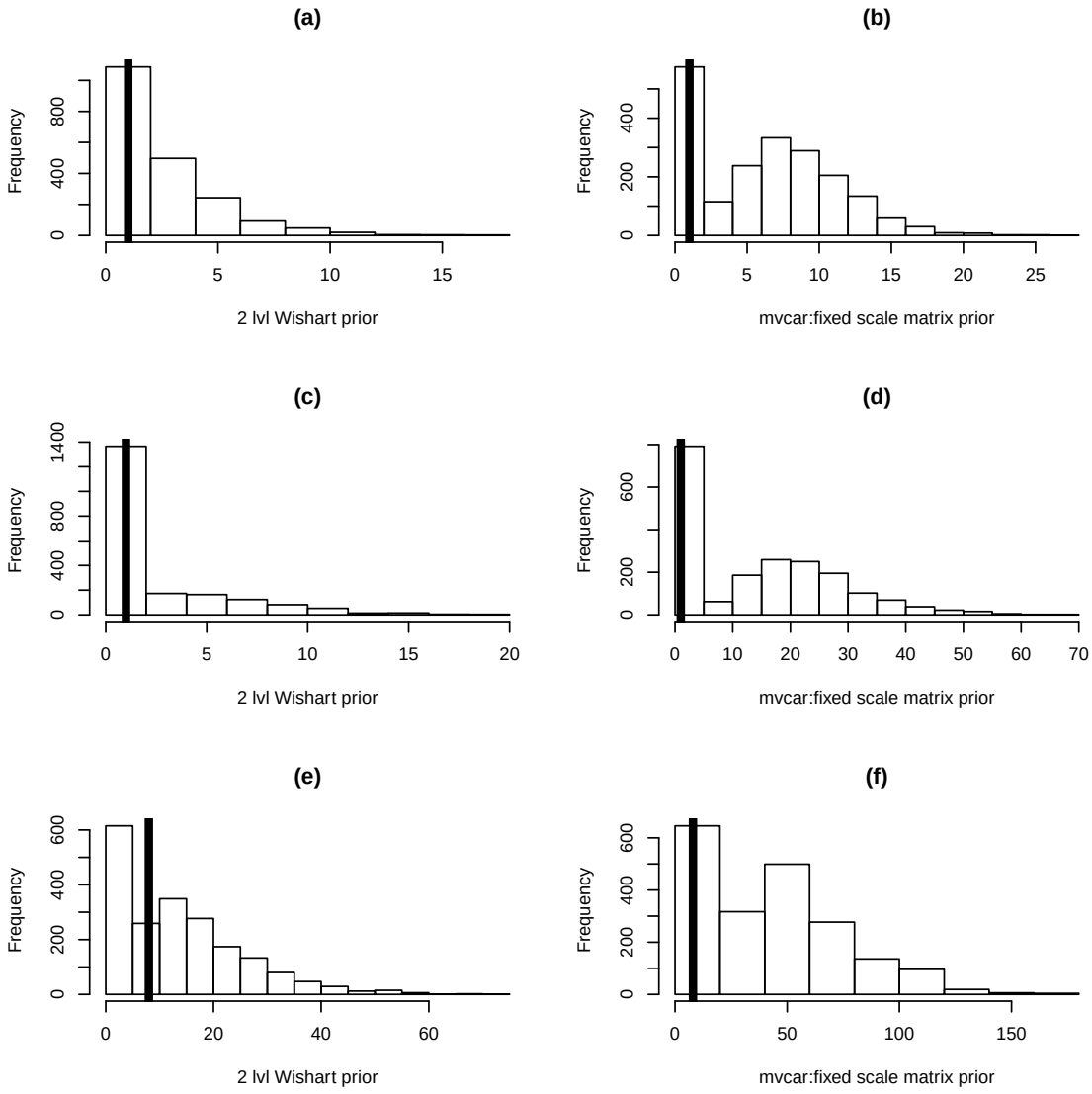
MCMC trace plots for abundance regression coefficients parameters for MI Crime Data, Burglary Forced Entry: (a), (b): the graphs show the trace of iteration values of β_4 for ZINBH under TLHIW and fixed scale matrix prior models, respectively. (c), (d): show the trace of interation values of β_5 for ZINBH under TLHIW and fixed scale matrix prior models, respectively.(e), (f): show the trace of interation values of β_6 for ZINBH under TLHIW and fixed scale matrix prior models, respectively.



Posterior Predictive Plots for Michigan Crime Data: Murder and Non-negligent Manslaughter (a), (c), and (f): The histograms show the posterior predictive distribution of the data for ZINBH at sites 4, 8 and 10 respectively, under TLHIW prior. (b), (d), and (f): The histograms show the posterior predictive distribution of the data for ZINBH at sites 4, 8 and 10 respectively, under fixed inverse-Wishart prior. The vertical lines represent the observed values.



Posterior Predictive Plots for Michigan Crime Data: Burglary Forced Entry. (a), (c), and (f): The histograms show the posterior predictive distribution of the data for ZINBH at sites 12, 56 and 81 respectively, under TLHIW prior. (b), (d), and (f): The histograms show the posterior predictive distribution of the data for ZINBH at sites 12, 56 and 81 respectively, under fixed inverse-Wishart prior. The vertical lines represent the observed values.



Bibliography

A. Agresti. *An Introduction to Categorical Data Analysis*. Wiley Interscience, New York, 2007.

P. Allison. Do we really need inflated models?, 2012a. URL `http://statisticalhorizons.com/zero-inflated-models`.

P. Allison. When can you safely ignore multicollinearity, 2012b. URL `http://statisticalhorizons.com/multicollinearity`.

K. Bae and B. K. Mallick. Gene selection using a two-level hierarchical bayesian model. *Bioinformatics*, 20:3423–3430, 2004.

A. Buu, N.J. Johnson, R. Li, and X. Tan. New variable selection methods for zero-inflated count data with applications to the substance abuse field. *Statistics in Medicine*, 30:2326–2340, 2011.

M. N. Tran C. Leng and D. Nott. Bayesian adaptive lasso. *Technical Report*, arXiv: 1009.2300, 2012.

G. Casella, J. Gilly, M Ghosh, and M. Kyung. Penalized regression, standard errors, and bayesian lassos. *Bayesian Analysis*, 5:369–412, 2010.

N. Cressie. *Statistics for Spatial Data,Revised Edition*. Wiley Interscience, New York, 1993.

A. E. Gelfand D. K. Agarwal and Citron-Pousty. Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics*, 9:341–355, 2002.

I. Frank and J. Friedman. A statistical view of some chemometrics regression tools. 1993.

A Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1:515–534, 2006.

Souparno Ghosh, Alan E Gelfand, Kai Zhu, and James S Clark. The k-zig: Flexible modeling for zero-inflated counts. *Biometrics*, 68(3):878–885, 2012.

Sujit K Ghosh, Pabak Mukhopadhyay, and Jye-Chyi JC Lu. Bayesian analysis of zero-inflated regression models. *Jounal of Statistical Planning and Inference*, 136: 1360–1375, 2006.

W. H. Greene. Some accounting for excess zeros and sample selection in poisson and negative binomial regression models. *Working paper EC-94-10:Department of Economics, New York University*, 1994.

R. Kronmal H. Liu, S. Ma and K. Chan. Semiparametric zero-inflated modeling in mulit-ethnic study of atherosclerosis(mesa). *Annals of Applied Statistics(to appear)*, 2012.

C. Hans. Model uncertainty and variable selection in bayesian lasso regression. *Statistics and Computing*, (2010:221–229, 2010.

A. Hoerl and R. Kennard. Ridge regression: Applications to nonorhtogonal problems. 1970.

M.C. Hu, M. Pavlicova, and E. V. Nunes. Zero-inflated and hurdle models of count data with extra zeros: Examples from an hiv-risk reduction intervention trial. *American Journal of Drug and Alcohol Abuse*, 37(5):367–375, 2011.

A. Huang and M. P. Wand. Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2):439–452, 2013.

N. Ismail and H. Zamani. Estimation of claim count data using negative binomial, generalized poisson, zero-inflated negative binomial and zero-inflated generalized poisson regression models. *Casualty Actuarial Soceity E-Forum, Springer*, 2013.

J. York J. Besag and A. Mollie. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–59, 1991.

G. K. Grunwald J. Tooze and R. H. Jones. Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research*, 11:341–355, 2002.

R. Winkelmann K. E. Staub and R. Adnan. Consistent estimation of zero-inflated count models. *Working Paper 9008, Socioeconomic Institute, University of Zurich*, 2011.

L. Kuo and B. K. Mallick. Variable selection for regression models. *Sankhya*, 60: 65–81, 1998.

P. A. Lachenbruch. Analysis of data with excess zeros. *Statistics in Medicine*, 11: 297–302, 2002.

R. Li and H. Lian. Variable selection in semiparametric regression modelling. *The Annals of Statistics*, 36(1):261–286, 2008.

D. Lord and B. Park. Appendix d: Negative binomial regression models and estimation methods. 2010. URL `http://www.icpsr.umich.edu/CrimeStat`.

A. Lykou and I. Ntzoufras. On bayesian lasso variable selection and the specification of the shrinkage parameter. *Statistics and Computing*, 23:361–390, 2013.

S. A. Padoan M. P. Wand, J. T. Ormerod and R. Fruhwirth. Mean field variational bayes for elaborate distributions. *Bayesiann Analysis*, 6:847–900, 2011.

D. Demetrio M. Ridout and J. Hinde. Models for count data with many zeros. In *booktitle*. International Biometric Conference, 1998.

D. Dunson M. Zhou, L. Li and L. Carin. Lognormal and gamma mixture negative binomial regression. 2012.

S.N. MacEachern and L. M. Berliner. Subsampling the gibbs sampler. *The American Statistician*, 48:188–190, 1994.

A. Majumdar and C. Gries. A bivariate zero-inflated regression model for count data: A bayesian approach with application to plant counts. *International Journal of Biostatistics*, 6(1):Article 27, 2010.

L.H. Moulton and N. A. Halsey. A mixture model with detection limits for regression analyzes of antibody response to vaccine. *Biometrics*, 51:1570–1578, 1995.

I. M. Nadiroh. Zero inflated negative binomial models in small area estimation. *Bachelors Project, Department of Statistics, Bogor Agricultural University*, 2009.

B. Neelon, A. O'Malley, and S.L. Normand. A bayesian model for repeated measures zero-inflated count data with application to psychiatric outpatient service use. *Statistical Modelling*, 10:421–439, 2010.

B. Neelon, P. Ghosh, and P. Loebs. A spatial poisson hurdle model for exploring geographica variation in emergency department visits. *Journal of the Royal Statistical Society A*, 175(4):1–25, 2012.

B Neelon, H.C. Chang, Q. Ling, and S.N. Hastings. A spatial temporal hurdle model for zero inflated count data: exploring trends in emergency department visits. *Statistical Methods in Medical Research*, 2014.

P. Osei. Statistical methods of disease mapping. *Thesis:African Institute of Mathematical Sciences*, 2010.

A. Schmidt and M. Rodriguez. Modelling multivariate counts varying continuously in space. 2010.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of Royal Statistical Society B*, 38:267–288, 1996.

P. Trevor and G. Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103:681–686, 2008.

J. M. VerHoef and J. K. Jansen. Space-time zero-inflated count models for harbor seals. *Envirometrics*, 18:697–712, 2007.

K. Yau Wang, A. Lee and P. Carrivick. A bivariate zero-inflated poisson regression model to analyzed occupational injuries. *Accident Analysis and Prevention*, 35: 625–629, 2003.

P. Wang. A bivariate zero-inflated negative binomial regression model for count data with excess zeros. *Economic Letters*, 78, 2003.

R. Williams. Alternatives to logistic regression (brief overview), 2015. URL `https://www3.nd.edu/ rwilliam/stats3/L09.pdf`.

M. Yaun and Y. Lin. Efficient empirical bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, 100:1215–1225, 2005a.

M. Yaun and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society B*, 68:49–67, 2005b.

H. Zhou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society B*, 67:301–320, 2005.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. 67(2):301–320, 2005.