This is the pre-peer reviewed version of the following article:

Sólymos, P., Lele, S. R. & Bayne, E. (in press): Conditional likelihood approach for analyzing single visit abundance survey data in the presence of zero inflation and detection error. *Environmetrics*, DOI: 10.1002/env.1149

which has been published in final form at http://onlinelibrary.wiley.com/doi/10.1002/env.1149/abstract

The abundance estimation method described in the paper has been implemented in the 'svabu' function of the 'detect' R package (available from the Comprehensive R Archive Network at http://cran.r-project.org/web/packages/detect/index.html). To install and start using the package, type into R console:

```
install.packages("detect")  # install the package
library (detect)            # load the package
help("detect-package")      # package help file
help("svabu")               # svabu help file
```

Conditional likelihood approach for analyzing single visit abundance survey data in the presence of zero inflation and detection error

Péter Sólymos[1], Subhash Lele[2] and Erin Bayne[3]

[1]Alberta Biodiversity Monitoring Institute, Department of Biological Sciences, University of Alberta, e-mail: solymos@ualberta.ca

[2]Department of Mathematical and Statistical Sciences, University of Alberta, e-mail: slele@ualberta.ca

[3]Department of Biological Sciences, University of Alberta, e-mail: bayne@ualberta.ca

Running title: Abundance estimation using single visit data

Word count in the abstract: 191

Word count in the manuscript as a whole: 6248

Word count in the main text: 4236 (from Introduction to Acknowledgements)

Number of references: 36

Number of figures and tables: 2 figures, 1 table

Address of correspondence: Péter Sólymos, Alberta Biodiversity Monitoring Institute, Department of Biological Sciences, CW 405, Biological Sciences Bldg., University of Alberta, Edmonton, Alberta, T6G 2E9, Canada, Phone: 780-492-8534, Fax: 780-492-7635, e-mail: solymos@ualberta.ca

Abstract

Current methods to correct for detection error require multiple visits to the same survey location. Many historical data sets exist that were collected using only a single visit and logistical/cost considerations prevent current research programs from collecting multiple visit data. In this paper we explore what can be done with single visit count data when there is detection error. We show that when appropriate covariates that affect both detection and abundance are available, conditional likelihood can be used to estimate the regression parameters of a binomial-zero inflated Poisson mixture model and correct for detection error. We use observed counts of Ovenbirds (*Seiurus aurocapilla*) to illustrate the estimation of the parameters for the Binomial-ZIP mixture model using a subset of data from one of the largest and longest ecological time series datasets that only has single visits. Our single visit method 1) does not require the assumptions of a closed population or adjustments caused by movement or migration; 2) is cost effective, enabling ecologists to cover a larger geographical region than possible when having to return to sites; and 3) resultant estimators appear to be statistically and computationally highly efficient.

Keywords: Closed populations, Conditional likelihood, Ecological Monitoring, Mixture models, Open populations, Pseudo-likelihood**.**

Introduction

Ecologists are fundamentally interested in understanding the environmental factors that influence variation in the size of populations. To understand variation in population size requires information on how the abundance of species changes in time and space. Many ecologists rely on relative differences in counts of the number of individuals observed to draw inferences about factors influencing populations (Krebs 1985). However, models that predict naïve estimates of abundance (e.g. Poisson regression) are known to underestimate true abundance because of detection error. Detection error for count data is the probability that an individual of a species is present during the period of observation but is not detected. Rarely is there no detection error in ecological data (Buckland *et al*. 1993, Yoccoz *et al*. 2001, Gu and Swihart 2004). Environmental factors that influence population size may also affect probability of detection. Thus, the issue of imperfect detection needs to be addressed if ecologists are to draw correct conclusions about factors influencing population change *per se* (MacKenzie *et al*. 2002, Tyre *et al*. 2003).

The last decade has seen an enormous growth in statistical methods to deal with detection error (MacKenzie *et al*. 2006, Royle and Dorazio 2008). One approach that has been widely adopted is that of multiple visit surveys that use an N-mixture approach to estimate detection error for count data (Royle 2004). In the N-mixture approach, true abundance has typically been modeled using a Poisson or a Negative Binomial (NB) distribution, while detection error has been modelled as a Binomial observation process. True abundance rates in the Poisson or Negative Binomial model and detection probabilities of individuals in the Binomial model are commonly modeled as a function of habitat and survey-specific characteristics. By accounting for detection error in the observed counts, N-mixture models differentiate between the two kinds of zeros: “false” zeros due to detection error where true abundance is greater than 0 but the

observed count is 0; and "true" zeros due to the state process where the true abundance is 0 and the observed count is also 0.

In many situations, a third type of zero can exist. When surveys take place on larger geographic scales, "true" zeros arise not only as zeros due to the Poisson or NB distribution but as a result of true zero-inflation (Martin *et al*. 2005). True zero-inflation can happen when a species' range is only partly covered by the extent of the area sampled, the species is quite rare, or the distribution of individuals is highly aggregated. Wenger and Freeman (2008) and Joseph *et al*. (2009) proposed zero-inflated Poisson (ZIP) and zero-inflated NB (ZINB) mixture models to account for this third type of true zeros. They used Binomial-ZIP and Binomial-ZINB models with a multiple visit sampling approach to account for detection error in over-dispersed counts.

The goal of all multiple visit methodologies is to provide a more accurate estimator of true abundance than the naïve estimator by adjusting for detection error. However, many historical data sets with a vast amount of information have been collected using only a single visit. As well, logistical and cost considerations preclude many current monitoring programs from collecting multiple visit data. Given the reality that many single visit datasets exist and will continue to be created, we explore the question, what can be done with single visit count data when there is detection error?

We show that detection error in count data can be corrected using only a single visit to a site provided some conditions are satisfied. Multiple visit methods assume a closed population, that is abundances do not change during the full survey period (Royle 2004), or assume certain types of migration/movement patterns (Dail and Madsen 2011, Chandler *et al*. 2011). We replace this assumption by requiring that covariates that affect detection and abundance are available. We argue such covariates are common in most ecological studies. For example, covariates that

affect detection of birds can often be obtained from the most basic characteristics of the surveys, i.e. time of day, time of year, and observer. Most research and monitoring projects are designed to compare abundance between different environmental conditions or times. We specify the conditions under which the parameters of the Binomial-ZIP N-mixture model, that account for all three kinds of zeros, can be consistently and efficiently estimated based on a single visit to sites. An important issue in complex models is the possibility of non-estimable parameters (Lele, 2010), so we also provide a simple diagnostic test for estimability of parameters for single visit models.

The Binomial-ZIP model

We consider the zero-inflated Poisson (ZIP) model for the true state. Our method can be extended to zero-inflated Negative Binomial (ZINB) model with minor algebraic manipulations. A hierarchical representation of the ZIP model is $(N_i \mid \lambda_i, A_i) \sim \text{Poisson}(\lambda_i A_i)$, $(A_i \mid \phi) \sim \text{Bernoulli}(1 - \phi)$, where N_i is the population abundance at location i ($i = 1, 2, \ldots, n$; the total number of sites), λ_i is the rate parameter of the Poisson distribution when the species is present at location i. The probability that $A_i = 0$ is ϕ, consequently the probability that at least one individual is present is $(1-\phi)(1-e^{-\lambda_i})$. The $\phi = 0$ case corresponds to a Poisson model for the true state. The Poisson rate parameter can be modelled as a function of covariates using the log link function: $\log(\lambda_i) = \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a vector of regression coefficients including the intercept (β_0), and $\boldsymbol{X}_i$ is the covariate matrix with n rows and as many columns as the number of variables in the model. Links other than the log-link for the Poisson model can also be used.

The observation process is modeled using the Binomial distribution as $(Y_i \mid N_i) \sim \text{Binomial}(N_i, p_i)$, where Y_i is the observed count at site i, and p_i is the probability of detecting an

individual given the true abundance N_i is greater than 0. The probability of detection can be modeled as a function of covariates using the logistic link function: $\text{logit}(p_i) = \boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is a vector of regression coefficients including the intercept (θ_0), and $\boldsymbol{Z}_i$ is a covariate matrix similar to $\boldsymbol{X}_i$. One can use links other than the logistic link in the Binomial model.

Parameter estimation

The likelihood function corresponding to the Binomial-Poisson mixture based on single visit is:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi | \ \boldsymbol{y}) = \prod_{i=1}^{n} \left\{ I(y_i = 0)\phi + (1-\phi) \sum_{N_i=y_i}^{\infty} \binom{N_i}{y_i} p_i^{y_i} (1-p_i)^{(N_i - y_i)} e^{-\lambda_i} \frac{\lambda_i^{N_i}}{N_i!} \right\},$$

where $I(.)$ is an indicator function. Because N_i is unknown, the likelihood involves summation over all possible values of N_i. Direct maximization of this function can lead to substantial confounding between the parameter ϕ and the intercept parameter θ_0 in the detection model. To reduce this confounding, we divide the problem in two parts. In the first part, we condition on a sufficient statistic for the parameter ϕ and use the conditional distribution of the data given the sufficient statistics to form a conditional likelihood function (Anderson, 1970) for the parameters ($\boldsymbol{\beta}$, $\boldsymbol{\theta}$). The conditional likelihood estimators are known to be consistent and asymptotically normal under fairly general conditions. To estimate ϕ, we construct a new random variable $W_i = I(Y_i > 0)$. Then, we write the likelihood function for ($\boldsymbol{\beta}$, $\boldsymbol{\theta}$, ϕ) based on the distribution of W_i. This likelihood function does not involve infinite summation and hence is easy to maximise. Further, it is a concave function of ϕ and hence has a unique solution. Based on the idea of pseudo-likelihood described in Gong and Samaniego (1981), we fix the values of ($\boldsymbol{\beta}$, $\boldsymbol{\theta}$) at their conditional likelihood based estimates $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}})$ and maximize the likelihood with respect to ϕ to

obtain its estimate. The results in Gong and Samaniego (1981) show that this pseudo-likelihood estimator is consistent and asymptotically normal. The derivation of the conditional and pseudo-likelihood functions is described in the Appendix. We use the bootstrap procedure (Efron and Tibshirani 1994) to calculate confidence intervals for the estimated parameters. The software implementation is available in the statistical package 'detect' (Sólymos *et al*. 2011) written in the free statistical software R (R Development Core Team 2011). We use probability plots to evaluate model fit under the Binomial-Poisson and Binomial-ZIP models. The model fit is adequate if the values of the empirical and fitted cumulative distribution function (CDF) fall along a line with intercept 0 and slope 1.

As pointed out by a referee, the marginal distribution of Y under the Binomial-ZIP model is identical to the marginal distribution of Y under a Zero Inflated Binomial-Poisson model. This result leads to some ambiguity in the interpretation of the zero-inflation parameter ϕ when using single survey methodology: Is it the zero inflation in the Poisson component or zero inflation in the Binomial component? We think the zero inflated Poisson model for the abundance distribution to be far more sensible than zero inflated Binomial model for the observation process. Nonetheless, from the scientific and management perspective, often the relationship between abundances and environmental covariates is more important than the zero inflation factor in either the Poisson or Binomial. As shown in the Appendix S2 of the Supplementary Information, the conditional likelihood for (β,θ) remains the same whether the model is Binomial-ZIP or ZIB-Poisson. Hence the estimators obtained using the conditional likelihood are valid under either model. Interpretation of the parameter ϕ is ambiguous. In our analysis, we interpret ϕ as zero inflation in the Poisson component because we do not think ZIB model for the observation process is sensible.

157

158 **Assumptions**

159 For most mixture models exact identifiability conditions are nearly impossible to specify. In the
160 single survey situation, exact mathematical proof of identifiability is not yet possible. We do,
161 however, know when they are not identifiable. For example, if the probability of detection and/or
162 abundance rate are constant (e.g. intercept only model, or if only discrete covariates are
163 available), single survey method leads to non-identifiability. Intuition and simulations suggest
164 that the parameters are estimable if there are continuous variables that affect both detection and
165 abundance. Furthermore, mathematics suggests and simulations indicate that we need to assume
166 that the covariate set for detection (A) and covariate set for abundance (B) should be such that
167 (A-B) and (B-A) are non-empty. That is, there should be covariates that affect only detection and
168 covariates that affect only abundance. So if the covariate vectors $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ have common
169 covariates, there needs to be at least one continuous covariate that is unique to either the
170 abundance or detection error vectors. According to our review of the detectability literature,
171 constant detection and/or constant abundance models are very rarely used in practice. The survey
172 also suggests that above assumptions on the covariate vectors are satisfied in many situations.
173 Although mathematical proof of identifiability is not possible in the N-mixture model
174 (without strong and possibly unrealistic assumptions), given a specific data set and a model,
175 estimability of the parameters can be checked using the data cloning algorithm (Lele *et al*. 2010;
176 Lele 2010). To protect against inappropriate analysis, we check for parameter estimability using
177 the diagnostics based on the data cloning algorithm as described in Appendix S3 of the
178 Supporting Information where we provide computer code for this diagnostic test.
179

Simulation study

To study properties of the estimation procedure, we performed several simulations. We considered situations where covariates that affect detection and abundance are distinct from each other and situations where some of the covariates are common, that is covariates that affected both detection and abundance. Furthermore, we considered eight different scenarios corresponding to combinations of low ($\bar{\lambda}$ = 2.13) vs. high abundance ($\bar{\lambda}$ = 5.25), zero-inflated (ϕ = 0.25) vs. non zero-inflated data (ϕ = 0), and low ($\bar{p}$ = 0.25) vs. high ($\bar{p}$ = 0.65) detection probability (for more details, see Appendix S1 in Supporting Information).

We fitted the Binomial-ZIP mixture model to each simulated data set using 100, 300, 500, 700, 1000 sites. All together we used 160 different settings (4 settings x 8 scenario x 5 sample sizes) and ran 100 simulations for each. Average of the true abundances varied between 1.6 to 5.2, while the average of the observed counts varied between 0.4 to 3.4 depending on the parameter settings and the covariates used in the simulations to describe detection error. These settings represented a wide range of ecologically plausible situations.

With single survey estimation, abundance parameters ($\boldsymbol{\beta}$) were consistently estimated (converged to the true values as the sample size increased), and reliable estimates were obtained with n = 100 in most situations. Detection parameters ($\boldsymbol{\theta}$) were also estimated consistently. The zero inflation parameter ϕ was well estimated even at small sample sizes. Predicted $\bar{\lambda}$ values were somewhat overestimated for n = 100; otherwise for larger sample sizes they were consistent with the true values. Predicted $\bar{p}$ values were consistent for all sample sizes. The correlation between the true and predicted $\bar{\lambda}$ and $\bar{p}$ values were high ranging from 0.8 to 1 for sample sizes n = 300 and above. Even when the data were simulated under no zero-inflation (ϕ = 0), the parameter ϕ was well estimated. Figure 1 represents the worst case scenario with a common

discrete covariate for the abundance and detection models, and low abundance – zero-inflated data – low detectability scenario. Even in this difficult situation, it is clear that the conditional likelihood method works well. A complete summary of the results obtained for the 160 cases is available in Appendix S1 in Supporting Information.

Analysis of the Ovenbird data

We used observed counts of Ovenbirds (*Seiurus aurocapilla*) to illustrate the estimation of the parameters for the Binomial-ZIP model. Data were collected in 1999 using Breeding Bird Survey (BBS) Protocols (Downes and Collins 2003) in the boreal plains eco-region of Saskatchewan. The goal of the study was to determine whether the abundance of this species was influenced by the amount of forest around each survey point. Data were collected along 36 BBS routes each consisting of 50 survey locations with survey locations separated by 800 meters. To increase independence of observations we used every second survey point along each route in our analysis (n = 766 survey locations). Attributes about the forest type and amount of forest remaining with a 400 meter radius were estimated from the Saskatchewan Digital Land Cover Project (MacTavish 1995).

The habitat requirements of the Ovenbird are well understood in the boreal forest (Hobson and Bayne 2002) and we expected that Ovenbird abundance would be positively influenced by the amount of forest, or deciduous forest remaining and negatively by amount of agricultural land. The zero-inflation component is likely to be present because of the marked difference in habitat suitability for the species along the agricultural area gradient. We also included latitude-longitude as the study covered an east-west gradient over 1000 kilometers and

225 a 400 km north-south gradient in length although *a priori* we were not sure what effect this
226 would have on abundance.
227 We expected three continuous and one discrete variable to influence detection
228 probability: time of day, time of year, amount of forest and observer. In general, male songbirds
229 sing very regularly early in the breeding season making it easy to detect individuals that are
230 present. As the breeding season progresses however, males spend less time singing as they focus
231 on other activities. This often results in lower detectability later in the breeding season. We
232 included Julian date as a variable influencing detection error. Male songbirds also have a
233 tendency to sing earlier in the day, shortly after sunrise, and then later in the morning focus on
234 mate guarding or foraging. To account for this, we included time of the day as a factor
235 influencing detectability. Detectability can also be influenced by habitat attributes. In more open
236 environments where forest loss has occurred it is plausible that birds can be heard from long
237 distances increasing the likelihood that an individual is detected (Schieck 1997). Alternatively, in
238 areas with more forest the chance of multiple males singing simultaneously may be higher.
239 Ovenbirds often countersing with each other, whereby one individual choosing to sing results in
240 all other individuals in close proximity singing in response to that individual. In less forested
241 areas with fewer individuals this behavior may be less likely to occur. Different observers has
242 different abilities to detect birds, thus this covariate is often used in detectability corrections.
243 Observer had two levels referring to two observers (with acronyms RDW, SVW), with a
244 relatively balanced contribution to the whole sampling effort (44 and 56 % respectively).
245 All covariates were scaled to unit variance and centered. We performed backward
246 stepwise model selection starting with the full model including all abundance and detection
247 covariates, and dropped insignificant terms until all remaining terms were significant on the 0.1

248 alpha level. Then we applied backward stepwise model selection based on Wald-tests to remove
249 non-significant terms from the model. We also calculated Akaike's Information Criterion (AIC)
250 and 90% confidence limits based on 199 nonparametric bootstrap samples for the final model.
251 We fitted the Binomial-ZIP mixture model to the single visit Ovenbird data set. We
252 started with the full model including habitat characteristics and geographic coordinates for the
253 abundance model, and observer, Julian day, time of day and observer for the detection model.
254 Proportion of forest area was used in both the abundance and detection model, because it was *a*
255 *priori* assumed to influence both processes (Model 1; Table2). We started by simplifying the
256 detection model first. We dropped the time of day, because that term was not significant based
257 on a Wald test (Model 2). All remaining terms in the detection model were significant ($p < 0.05$).
258 Then we started dropping terms from the abundance model. Proportion of deciduous forest,
259 proportion of area converted to agriculture and longitude were not significant. All terms in our
260 final model (#5) were significant based on asymptotic Wald-tests ($p < 0.05$, for latitude: $p < 0.1$).
261 This model could not be further simplified without the loss of parameter estimability (see
262 numerical proof in Appendix S3 of Supporting Information). The AIC value corresponding to the
263 Binomial-Poisson mixture with the same covariates as Model 5 was 1353.9. This is much higher
264 than the AIC value 1025.4 of the Binomial-ZIP model. Aside from better AIC value, the
265 probability plot clearly shows that the Binomial-ZIP model fit is better than the Binomial-
266 Poisson model (Fig. 2B).
267 Proportion of forest area had positive effect on Ovenbird abundance. Latitude was only a
268 marginally significant predictor of abundance suggesting that there was a slight spatial pattern
269 that explained some of the variation in Ovenbird abundance. Ovenbird abundance increased
270 further north in the study area. Julian date had significant negative effect on detectability of

individuals probably because of decreased singing activity later in the season. Proportion of forest area had significant negative effect on detectability. This indicates that individuals are more detectable in open habitats, in spite of lower abundances in such habitats. Observer effect was also significant with associated average detection probabilities of 0.52 and 0.65 for the two observers. The zero-inflation component was 0.31, and the average probability of Poisson zeros ($P(N = 0) = \text{mean}\{(1-\phi)\, e^{-\lambda_i}\}$) was 0.27 (Table2, Fig. 2A). The probability of occurrence ($\text{mean}\{(1-\phi)(1-e^{-\lambda_i})\}$) was 0.41 and predicted mean abundance for the entire study area was $(1-\phi)\,\bar{\lambda} = 2.32$. This translates into the population estimate of 5.69-10.88 male birds per point count station at point count stations where the entire area was forested (100% forest cover) depending on latitude, including true zero inflation. Mean probability of detection of individual Ovenbirds was 0.65.

Given that Breeding Bird Survey uses an unlimited sampling distance to count birds, absolute density cannot be directly estimated from the Ovenbird example. However, Rosenberg and Blancher (2004), as part of the Partners in Flight planning process, estimated that the maximum distance over which Ovenbirds could be heard on BBS routes was 200 metres. Using this as the area sampled by BBS counts, our mean count when a point count station has 100% forest cover converted to a density of 0.661-1.262 male Ovenbirds per hectare depending on latitude. This is close to the density estimate of 0.99 (95% confidence limits (CL): 0.85-1.12) found by Bayne (2000) who mapped the territories of color-banded male Ovenbirds and determined absolute density in the same region.

Discussion

The N-mixture models that account for detection error in wildlife studies represent an important class of models. According to Royle *et al*. (2005): “It is not possible to estimate or

294 model variation in abundance free of detection probability without additional information. In
295 many animal sampling problems, a simple way to acquire this additional information is to
296 generate replicate counts (in time) under the conventional 'closed population' assumption that no
297 gains or losses occur over the duration of the replicate sampling". As such, most studies have
298 relied on replicate sampling to correct for detection error. We show that if non-overlapping set of
299 covariates exist that influence detection and abundance rate, detection error can be corrected with
300 single visit survey data for occupancy and abundance studies. The single survey methodology
301 requires neither the assumption of closed population nor assumptions about types of migration
302 and movement patterns to correctly estimate population abundance. Thus, single-visit approach
303 provides a means for correcting detection error for large-scale long-term historic datasets like the
304 Breeding Bird Survey for which multiple visit data is not and will not be available.

305 An objection raised against the use of single survey method is the requirement of the
306 covariates. For example, it is argued that if proper covariates have not been collected, the entire
307 single survey dataset become useless. While this objection is valid, similar objections can be
308 raised against naïve models or multiple survey estimators. If the closed population assumption is
309 not satisfied, entire multiple survey datasets can also be viewed as useless. Furthermore, if
310 proper covariates are not collected then naïve and multiple visit models will both be
311 inappropriate for prediction. This is a general problem with regression methodology, not single
312 survey methods.

313 The use of conditional likelihood reduces the confounding among the parameters with
314 respect to the zero-inflation coefficient with a possible loss of *asymptotic* statistical efficiency.
315 The conditional likelihood separates the parameter space and hence reduces the extent of
316 confounding in these situations. This leads to numerical stability in small samples. Simulations

317 indicate that there is hardly any loss of efficiency in using conditional likelihood. Use of
318 conditional likelihood to eliminate nuisance parameters has a long history in statistical inference
319 (e.g. Kalbfleish and Sprott 1973). The phenomenon that use of conditional likelihood improves
320 stability of the estimators of the parameter of interest is commonly observed. For example, use of
321 REML (Restricted Maximum Likelihood) stabilizes the estimation of variance components in
322 linear mixed models. Conditional likelihood estimators have also been used in the wildlife
323 ecology literature (Buckland *et al*. 1993, Farnsworth *et al*. 2002).

324 Many sampling methods and statistical analyses have been developed to estimate species
325 abundance. Even when it is possible to measure abundance/density, the economics of doing so
326 can be prohibitive for large-scale applications. As a result, collecting presence/absence
327 (detection/non-detection) data at a series of locations to get coarse measures of species
328 abundance has become a preferred method of evaluating ecological status and trends because of
329 the simplicity of data collection (MacKenzie *et al*. 2006). A companion paper (Lele *et al*. in
330 press) and a PhD thesis (Moreno 2011) shows that one can estimate detection error with a single
331 survey for presence/absence (detection/non-detection) data but need substantially larger sample
332 sizes. When abundance data are available, the estimators are stable and efficient, at much smaller
333 sample sizes. Furthermore, using the zero-inflated Poisson model for the true abundance, one can
334 differentiate between zero-inflation and Poisson zeros. This is not possible when using
335 detected/not-detected data to model site occupancy. Hence, we encourage ecologists to collect
336 count data whenever possible.

337 N-mixture models based on multiple visits can be misleading when the assumption of
338 closure is violated. For example, Rota *et al*. (2009) found that 71-100% of bird species showed
339 violation of closure across time periods of 3 weeks and 8 days. Chandler *et al*. (2011) found that

a multiple visit N-mixture model for the Chesnut-sided Warbler (*Dendroica pensylvanica*) overestimated density by ~400 % if random temporal emigration was not taken into account. Dail and Madsen (2011) found similarly high bias with simulations. Because of this bias, changes have been recommended in survey designs that maximize the chance of getting a closed population. This has been done by redefining the time or space interval over which multiple surveys need to be done to obtain a closed population (e.g. Kendall and White 2009). All of these corrections to survey design may be useful in situations when the closure assumption of Royle's (2004) original N-mixture model is violated, but require additional information that is not always available. Many ecologists already have multiple survey datasets that violate the closed population assumption and for which the modified survey intervals cannot be corrected *post-hoc*. What ecologists should do with such data has not been addressed in the literature and we suggest that our single visit methodology provides an alternative to simply relying on naïve estimators of abundance.

Acknowledgements

Comments from Editor Walter W. Piegorsch, the Associate Editor, two anonymous referees and Marc Kéry greatly improved the manuscript. We would like to thank Stan Boutin, Steve Cumming, Steve Matsuoka, Dave Huggard, Monica Moreno, Jim Schieck, Fiona Schmiegelow, Samantha Song, and the Boreal Avian Modeling Project Team and Technical committee for helpful discussions on the issue of detection error. Special thanks to Dr. Keith Hobson of Environment Canada for providing access to the data for the Ovenbird example. Funding for this research was provided by the Alberta Biodiversity Monitoring Institute, Environment Canada, North American Migratory Bird Conservation Act, and Natural Sciences and Engineering Research Council.

363 **References**

364 Anderson EB. 1970. Asymptotic properties of conditional maximum likelihood estimators. *J.*
365 *Royal Stat. Soc. B* **32**: 283-301.

366 Bayne EM. 2000. *Effects of forest fragmentation on the demography of ovenbirds (*Seiurus
367 aurocapillus*) in the boreal forest*. University of Saskatchewan, Saskatoon, Canada. PhD
368 Thesis.

369 Buckland ST, Anderson DR, Burnham KP, Laake JL. 1993. *Distance Sampling*. Chapman and
370 Hall.

371 Casella G, Berger RL. 2002. *Statistical inference*. 2nd edn. Australia, Pacific Grove, CA.
372 Thomson Learning. 660 p.

373 Chandler RB, Royle A, King DI. 2011. Inference about density and temporary emigration in
374 unmarked populations. *Ecology* **92**: 1429-1435.

375 Dail D, Madsen L. 2011. Models for estimating abundance from repeated counts of an open
376 metapopulation. *Biometrics* **67**: 577-587.

377 Downes CM, Collins BT. 2003. *The Canadian breeding bird survey, 1967-2000*. Canadian
378 Wildlife Service, Progress Notes No. 219. National Wildlife Research Centre, Ottawa,
379 ON.

380 Efron B, Tibshirani R. 1994. *An introduction to the bootstrap*. Chapman & Hall/CRC. 436 p.

381 Farnsworth GL, Pollock KH, Nichols JD, Simons TR, Hines JE, Sauer JR. 2002. A removal
382 model for estimating detection probabilities from point count surveys. Auk 119:414–425.

383 Gong G, Samaniego FJ. 1981. Pseudo-likelihood estimation: theory and applications. *Annals of*
384 *Statistics* **9**: 861-869.

385 Gu W, Swihart RK. 2004. Absent or undetected? Effects of non-detection of species occurrence
386 on wildlife-habitat models. *Biol. Conserv.* **116**: 195-203.
387 Hobson KA, Bayne EM. 2002. Breeding bird communities in boreal forest of Western Canada:
388 Consequences of "unmixing" the mixed woods. *Condor* **102**: 759-769.
389 Joseph LN, Elkin C, Martin TG, Possingham HP. 2009. Modeling abundance using N-mixture
390 models: the importance of considering ecological mechanisms. *Ecol. Appl.* **19**: 631-42.
391 Kalbfleish JD, Sprott DA. 1973. Marginal and Conditional likelihoods. *Sankhya* **35**: 311-328.
392 Kendall WL, White GC. 2009. A cautionary note on substituting spatial subunits for repeated
393 temporal sampling in studies of site occupancy. *J. Appl. Ecol.* **46**: 1182-1188.
394 Krebs CJ. 1985. *Ecology: The experimental analysis of distribution and abundance*. 3rd edn.
395 Harper and Row, New York, USA.
396 Lele SR. 2010. Model complexity and information in the data: could it be a house built on sand?
397 *Ecology*, **91**: 3503-3514.
398 Lele SR, Moreno M, Bayne E. in press. Dealing with detection error in site occupancy surveys:
399 What can we do with a single survey? *Journal of Plant Ecology*, …
400 Lele SR, Nadeem K, Schmuland B. 2010. Estimability and likelihood inference for generalized
401 linear mixed models using data cloning. *Journal of the American Statistical Association*,
402 **105**: 1617-1625.
403 MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA. 2002. Estimating
404 site occupancy rates when detection probabilities are less than one. *Ecology* **83**: 2248-
405 2255.

MacKenzie DI, Nichols JD, Royle AJ, Pollock KH, Bailey LL, Hines JE. 2006. *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence.* Elsevier, Amsterdam, Netherlands. 324 pp.

MacTavish P. 1995. *Saskatchewan digital landcover mapping project*. Report I-4900-15-B-95. Saskatchewan Research Council, Saskatoon, SK.

Martin TG, Wintle BA, Rhodes JR, Kuhnert PM, Field SA, Low-Choy SJ, Tyre AJ, Possingham HP. 2005. Zero tolerance ecology: improving ecological inference by modeling the source of zero observations. *Ecol. Lett.* **8**: 1235-1246.

Moreno M. 2011. *Site occupancy models*. Ph.D. thesis, University of Alberta, Edmonton AB, pp. 206

Moreno M, Lele SR. 2010. Improved estimation of site occupancy using penalized likelihood. *Ecology*, **91**: 341-346.

R Development Core Team. 2011. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org [accessed 1 January 2011]

Rosenberg KV, Blancher PJ. 2005. Setting numerical population objectives for priority landbird species. In: *Bird Conservation and Implementation in the Americas: Proceedings of the Third International Partners in Flight Conference* (eds. Ralph CJ, Rich TD). U.S. Department of Agriculture, Forest Service, General Technical Report PSW-GTR-191. Vol. 1, pp. 57-67.

Rota CT, Fletcher RJ, Dorazio RM, Betts MG. 2009. Occupancy estimation and the closure assumption. *Journal of Applied Ecology* **46**: 1173-1181.

Royle JA. 2004. N-mixture models for estimating population size from spatially replicated counts. *Biometrics* **60**: 108-115.

Royle JA, Dorazio RM. 2008. *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*. Academic Press, San Diego, CA. xviii, 444 pp.

Royle JA, Nichols JD, Kéry M. 2005. Modelling occurrence and abundance of species when detection is imperfect. *Oikos* **110**: 353-359.

Schieck J. 1997. Biased detection of bird vocalizations affects comparisons of bird abundance among forested habitats. *The Condor* **99**: 179-190.

Sólymos P, Moreno M, Lele SR. 2011. 'detect': analyzing wildlife data with detection error. R package version 0.1-0. URL: http://cran.r-project.org/package=detect [accessed 19 October 2011]

Tyre AJ, Tenhumberg B, Field SA, Niejalke D, Parris K, Possingham HP. 2003. Improving precision and reducing bias in biological surveys: estimating false negative error rates. *Ecol. Appl.* **13**: 1790-1801.

Wenger SJ, Freeman MC. 2008. Estimating species occurrence, abundance, and detection probability using zero-inflated distributions. *Ecology*, **89**: 2953-2959.

Yoccoz NG, Nichols JD, Boulinier T. 2001. Monitoring of biological diversity in space and time. *Trends in Ecol. Evol.* **16**: 446-453.

449 **Appendix: Conditional and pseudo-likelihood estimation for the Binomial-Zero Inflated**

450 **Poisson mixture model**

451 Let $Y_i \mid N_i \sim Binomial(N_i, p_i)$ where $p_i = p(Z_i, \theta)$ is a function of detection covariates Z_i. Let

452 $N_i \mid A_i \sim Poisson(A_i \lambda_i)$ where $\lambda_i = \lambda(X_i, \beta)$ is a function of abundance covariates X_i. Further

453 $A_i \sim Bernoulli(1-\phi)$. Then the random variable Y_i is said to follow a Binomial-Zero Inflated

454 Poisson distribution. We first derive some elementary mathematical statistics results related to

455 this distribution.

456 **Result 1**: Consider the conditional distribution

457 $$P(Y_i = y_i \mid Y_i > 0) = \frac{P(Y_i = y_i)}{1 - P(Y_i = 0)} \text{ for } y_i = 1,2,3,....$$

458 The probability mass function for this conditional distribution is given by:

459 $$P(Y_i = y_i \mid Y_i > 0) = \frac{\sum_{N_i = y_i}^{\infty} \binom{N_i}{y_i} p_i^{y_i} (1-p_i)^{N_i - y_i} e^{-\lambda_i} \lambda_i^{N_i} / N_i!}{1 - e^{-\lambda_i p_i}} \text{ for } y_i = 1,2,3,...$$

460 Notice that this conditional distribution does not depend on the parameter ϕ.

461 *Proof:* This proof follows elementary probability theory (e.g. Casella and Berger, 2002).

462 $$P(Y_i = y_i \mid Y_i > 0) = \frac{P(Y_i = y_i)}{1 - P(Y_i = 0)} = \frac{(1-\phi) \sum_{N_i = y_i}^{\infty} \binom{N_i}{y_i} p_i^{y_i} (1-p_i)^{N_i - y_i} e^{-\lambda_i} \lambda_i^{N_i} / N_i!}{1 - P(Y_i = 0)} \quad \ldots\ldots\ldots(1)$$

463 Further,

$$P(Y_i = 0) = \phi + (1-\phi)\sum_{N_i=0}^{\infty}\binom{N_i}{0} p_i^{\,0}(1-p_i)^{N_i-0} e^{-\lambda_i}\lambda_i^{\,N_i}/N_i!$$

464
$$= \phi + (1-\phi)e^{-\lambda_i}\sum_{N_i=0}^{\infty}\left[(1-p_i)\lambda_i\right]^{N_i}/N_i!$$
$$= \phi + (1-\phi)e^{-\lambda_i}e^{(1-p_i)\lambda_i}$$
$$= \phi + (1-\phi)e^{-\lambda_i p_i}$$

465 Hence, we can write

466
$$1 - P(Y_i = 0) = (1-\phi)\left(1 - e^{-\lambda_i p_i}\right) \quad \text{.....(2)}$$

467 Combining equations (1) and (2), it follows that:

468
$$P(Y_i = y_i \mid Y_i > 0) = \frac{\sum_{N_i=y_i}^{\infty}\binom{N_i}{y_i} p_i^{\,y_i}(1-p_i)^{N_i-y_i} e^{-\lambda_i}\lambda_i^{\,N_i}/N_i!}{1-e^{-\lambda_i p_i}}.$$

469 **Result 2**: The binary random variable defined by $W_i = I_{(Y_i>0)}$ has the following distribution:

470
$$P(W_i = 0) = \phi + (1-\phi)e^{-\lambda_i p_i}$$

471
$$P(W_i = 1) = (1-\phi)\left(1 - e^{-\lambda_i p_i}\right).$$

472 *Proof:* Follows from equation (2) in the proof of the previous result.

473 **Conditional likelihood estimation of (β,θ):**

474 To estimate the parameters (β,θ), we use the likelihood using only those sites that have at

475 least one individual observed. This is called the conditional likelihood function (Anderson 1970).

476 The conditional likelihood is given by: $CL(\beta,\theta) = \prod_{y_i>0} P(Y_i = y_i \mid Y_i > 0)$ where the product is only

477 on those sites where $y_i > 0$. We maximize this function to obtain the estimates of the parameters

478 (β,θ). The conditional likelihood estimators are known to be consistent (Anderson 1970) as the

479 number of sites that have at least one individual observed increases.

480 **Pseudo-likelihood estimation of ϕ:**

To estimate the parameter ϕ, we consider the likelihood based on the random variables W_i where parameters (β,θ) are fixed at their conditional likelihood estimates $(\hat{\beta},\hat{\theta})$. Gong and Samaniego (1981) call such likelihood 'pseudo-likelihood'.

$$PL(\phi;W,\hat{\beta},\hat{\theta}) = \prod_{i=1}^{n} \left\{(1-\phi)(1-e^{-\hat{\lambda}_i \hat{p}_i})\right\}^{W_i} \left\{\phi + (1-\phi)e^{-\hat{\lambda}_i \hat{p}_i}\right\}^{1-W_i}$$

Because the conditional likelihood estimates $(\hat{\beta},\hat{\theta})$ are consistent, the pseudo-likelihood estimator of ϕ obtained by maximizing the pseudo-likelihood is also consistent (Gong and Samaniego 1981).

SUPPORTING INFORMATION

The following Supporting Information is available for this article:

Appendix S1 *Simulation results*

Appendix S2 *Conditional likelihood under ZIB-Poisson model*

Appendix S3 *Identifiability diagnostics*

Table 1. Model selection results for the Ovenbird data set based on the Binomial-ZIP mixture (n = 766 survey locations). Model terms not significant (based on Wald test) were backward dropped until only significant ($p < 0.1$) terms remained (Model 5). Bootstrap based 90% confidence intervals are provided in parentheses for most parsimonious Model 5 (see Appendix S3 for numerical proof of identifiability of model parameters for Model 5).

	Model 1	Model 2	Model 3	Model 4	Model 5	CI for Model 5
Abundance						
Intercept	0.453	0.333	0.239	0.349	0.140	(-0.103, 0.478)
Proportion of forest area	1.028	1.498	1.016	1.031	1.384	(0.947, 1.487)
Proportion of deciduous area	0.045	0.008				
Proportion of agricultural area	-0.267	0.449	-0.136			
Latitude	0.112	0.304	0.198	0.205	0.133	(-0.019, 0.224)
Longitude	0.075	0.282	0.166	0.099		
Detection						
Intercept	0.065	0.821	0.949	0.693	0.785	(0.207, 1.710)
Proportion of forest area	-1.562	-1.749	-1.813	-1.592	-1.873	(-2.405, -1.332)
Julian day	-0.303	-0.458	-0.380	-0.428	-0.359	(-0.507, -0.249)
Time of day	0.072					
Observer (SVW)	0.516	0.545	0.592	0.523	0.553	(0.135, 0.899)
ϕ	0.346	0.389	0.363	0.391	0.314	(0.257, 0.438)
$P(N = 0)$	0.217	0.207	0.230	0.197	0.272	(0.171, 0.324)
$\bar{\lambda}$	4.092	3.169	2.844	2.849	3.380	(1.996, 3.598)
$(1-\phi)\bar{\lambda}$	2.676	1.936	1.810	1.734	2.318	(1.238, 2.398)
$\bar{p}$	0.566	0.664	0.678	0.654	0.654	(0.575, 0.735)

503 **Figure captions**

504 **Figure 1.** Simulation results with a common discrete covariate used both for the
505 abundance (β_2) and the detection (θ_2) model. Each box and whiskers correspond to 100
506 simulations; horizontal axes give the sample size (*n*) used for estimation. As *n* increases, medians
507 (thick black lines) are getting closer to the true parameter values (thick grey lines), and estimates
508 are getting accurate (inter-quartile boxes and range whiskers getting narrower). The low
509 abundance – zero inflated data – low detectability scenario was used. β, θ, and ϕ are model
510 parameters (see text), $\bar{\lambda}$ is the mean of the predicted rate parameter of the Poisson distribution, $\bar{p}$
511 is the mean of the detection probabilities. Correlations between true and predicted λ and *p* values
512 are shown in the lowest row. Right bottom insert represents the count distribution for an example
513 data set out of the 100 simulated ones, black bars are true, grey bars are observed counts (one
514 pair of bars for each count).

515 **Figure 2.** Count distribution for the Ovenbird data set (A) and probability plot (B) for the
516 N-mixture model fitted to the data set. Ovenbird abundances are actual counts from 891
517 locations (grey bars), the estimated proportion of zero-inflation (black) and Poisson zeros (white)
518 are shown beside the zero point mass bar, the difference between the observed and predicted zero
519 point mass is due to non-detection zeros. The probability plot shows the values of the empirical
520 and fitted cumulative distribution functions (CDF) based on the Binomial-Poisson (filled circles)
521 and the Binomial-ZIP (open circles) mixtures. Scattered line represents the line with slope 1;
522 values closer to this line indicate better fit.
523

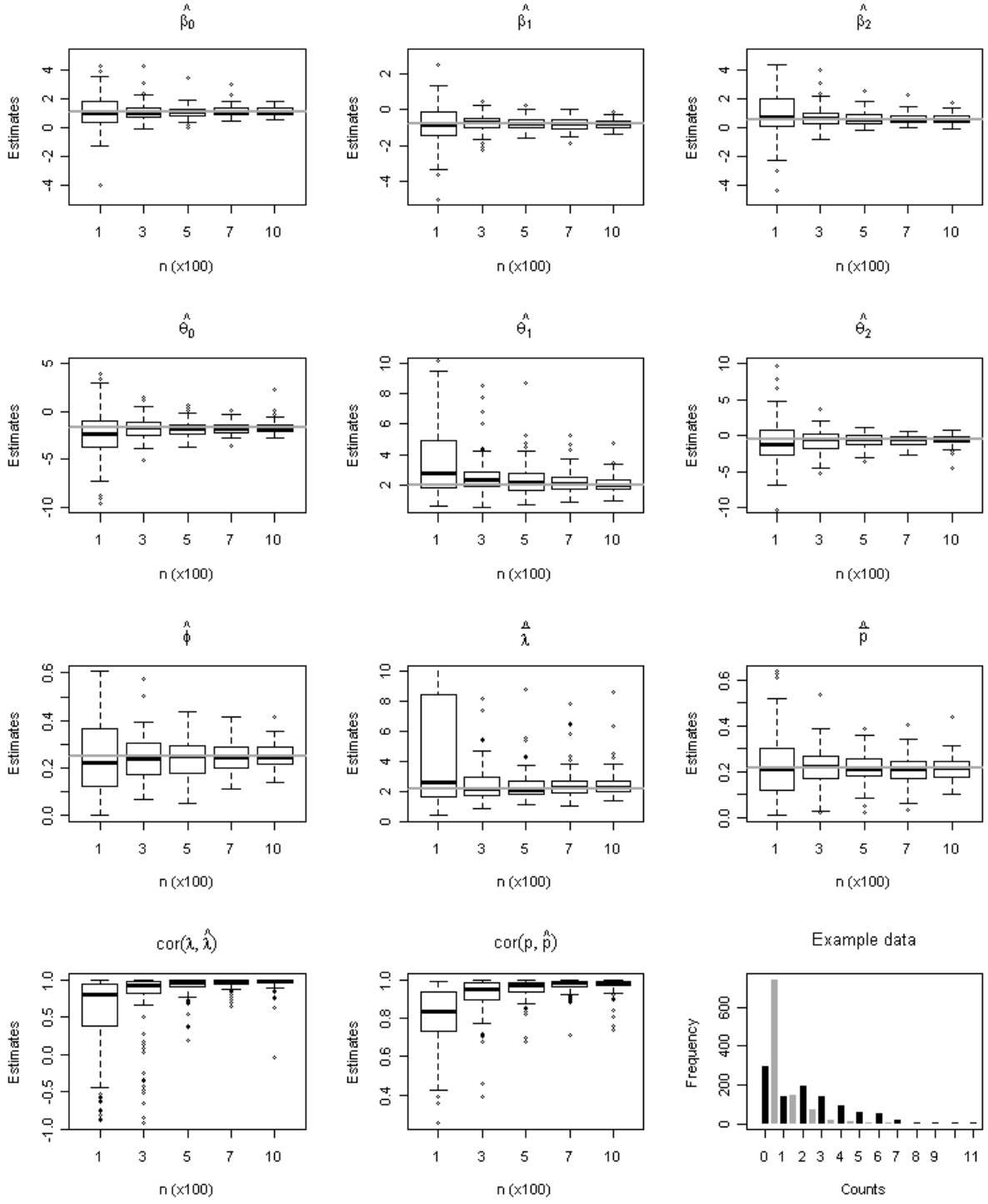
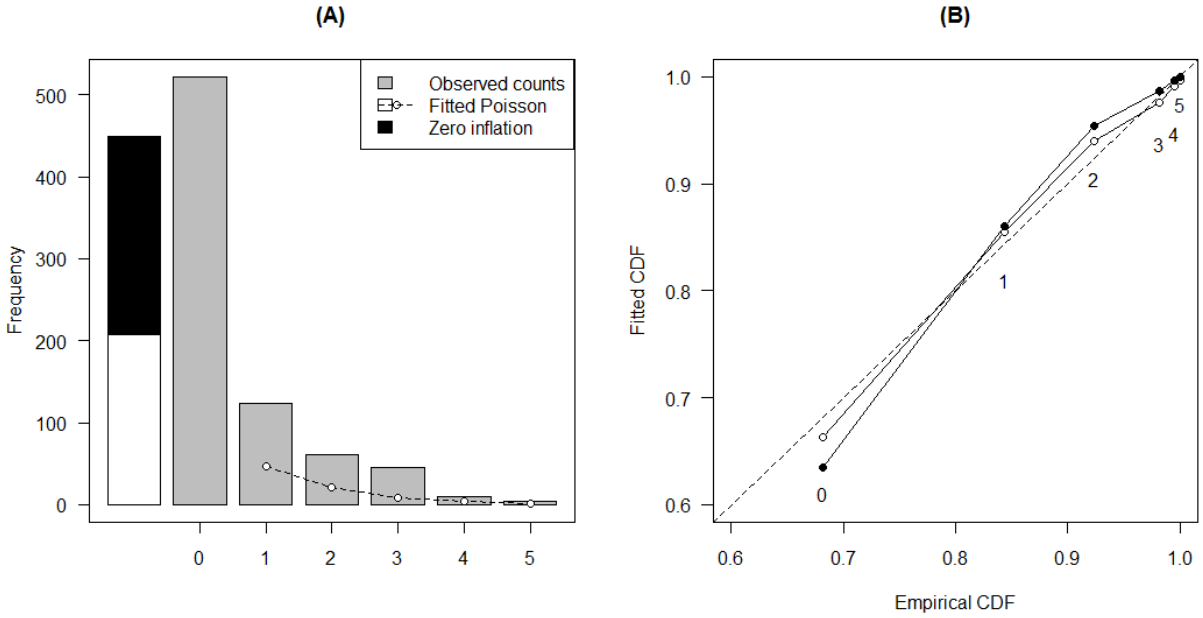

β̂0
β̂1
β̂2
θ̂0
θ̂1
θ̂2
ϕ̂
λ̂
p̂
cor(λ̂, λ̂)
cor(p̂, p̂)
Example data
Estimates
n (x100)
Frequency
Counts

Fig. 1



526

527 Fig. 2

528

Appendix S1: Simulation Results

October 12, 2011

This document provides supporting information to the paper:
Péter Sólymos, Subshash Lele and Erin Bayne, Conditional likelihood approach for analyzing single visit abundance survey data in the presence of zero inflation and detection error

Contents

1 Introduction **1**

2 Separate continuous covariate (Setup 1) **4**

3 Separate discrete covariate (Setup 2) **12**

4 Common continuous covariate (Setup 3) **20**

5 Common discrete covariate (Setup 4) **28**

1 Introduction

To study the properties of the estimating procedure described in the previous section, we performed simulations. We used six randomly generated covariates in four different setups (Table 1). In all simulation setups, we used continuous covariates that were unique to either to the abundance or the detection model. Besides these, we used separate continuous covariates (Setup 1), separate discrete covariates (Setup 2), a common continuous covariate (Setup 3), and a common discrete covariate (Setup 4).

Besides the covariate setups, we established eight different scenarios corresponding to combinations of low ($\beta_0 = 1.1$, $\bar{\lambda} = 2.13$) vs. high abundance ($\beta_0 = 2$, $\bar{\lambda} = 5.25$), zero-inflated ($\phi = 0.25$) vs. non zero-inflated data ($\phi = 0$), and low ($\theta_0 = -1.7$, $\bar{p} = 0.25$) vs. high ($\theta_0 = 1$, $\bar{p} = 0.65$) detection probability. Other abundance ($\beta_1 = -0.8$, $\beta_2 = 0.5$) and detection parameters ($\theta_1 = 2$, $\theta_2 = -0.5$) were the same for all simulations.

For each scenario in each setup, we generated true abundances (N_i) and observed counts (Y_i) for $n = 1000$ sites, and repeated this 100 times. We fitted

the Binomial-ZIP mixture model to each simulated data set using 100, 300, 500, 700, 1000 sites. All together we used 160 different settings (4 settings x 8 scenario x 5 sample size) and fitted the Binomial-ZIP model to 16000 random data sets. Average of the true abundances varied between 1.6–5.2, average of the observed counts varied between 0.4–3.4 depending on the parameter settings and the covariates used in the simulations. These settings represented a wide range of ecologically plausible situations.

Because it is a concern, that multiple surveys (given the assumptions are met) provide better inference compared to the single visit approach, we took our worst case setup (common discrete covariate to the abundance and detection model; all eight scenarios), and generated four independent visits to each location. We then compared the single visit $n = 500$ results with the 2 visits $n = 250$ results, and the single visit $n = 1000$ case with the 2 visits $n = 500$, and 4 visits $n = 250$ results.

Figures are composed of 12 subplots in each. Boxplots represent the values of the 100 simulations, horizontal grey line represent the true values.

- $\hat{\beta_0}$, $\hat{\beta_1}$, $\hat{\beta_2}$: abundance parameter estimates,
- $\hat{\theta_0}$, $\hat{\theta_1}$, $\hat{\theta_2}$: detection parameter estimates,
- $\hat{\phi}$: estimate of the zero inflation parameter,
- $\hat{\bar{\lambda}}$: estimate of the mean Poisson rate parameter,
- $\hat{\bar{p}}$: estimate of the mean probability of detection parameter,
- $cor(\lambda, \hat{\lambda})$: correlation between true and predicted Poisson rate parameter values,
- $cor(p, \hat{p})$: correlation between true and predicted probability of detection parameter values,
- Example data: histogram with an example data set, where black bars are true (N_i) and grey bars are observed (Y_i) count frequencies.

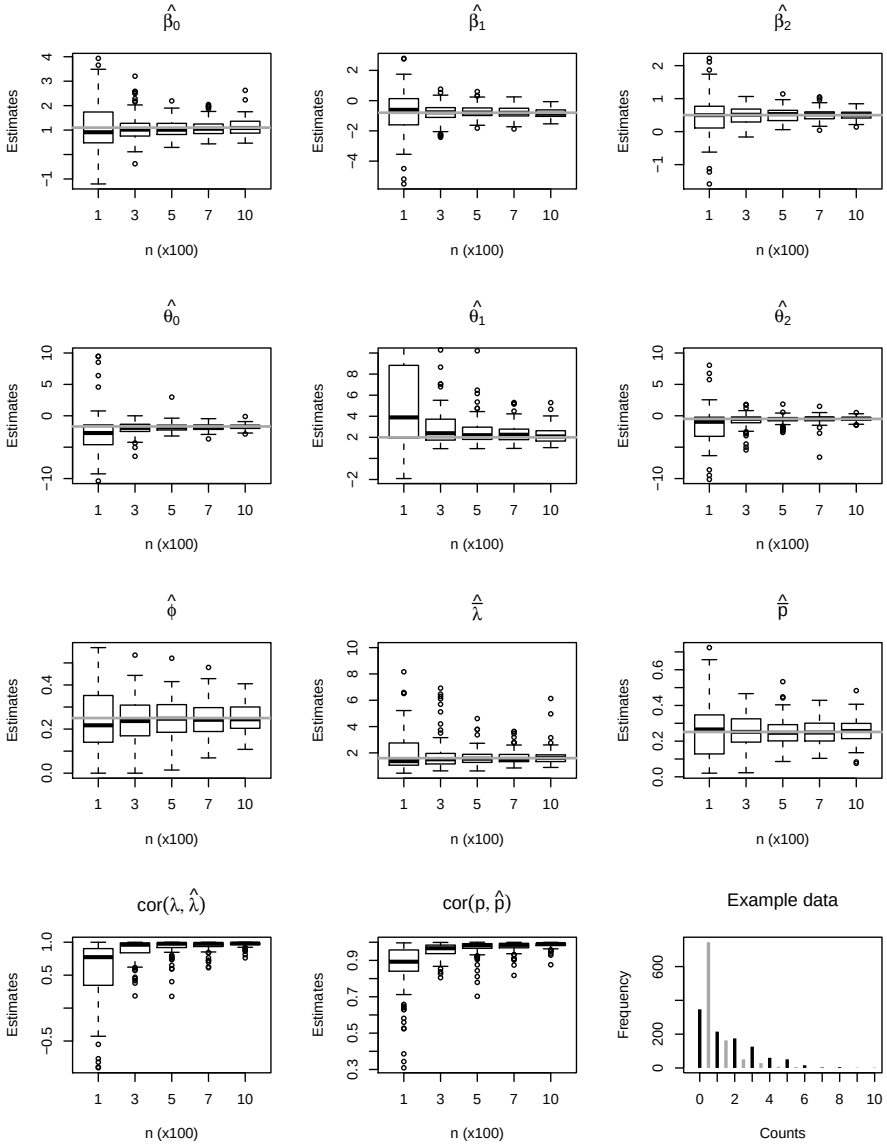
Otherwise, all the figures are similar to the figures presented in the paper, but shows different simulation settings.

Table 1: Settings for simulations. Covariates used for the abundance and detection models in the four different simulation setups. β and θ symbols refer to abundance and detection effects, respectively, that were used in the simulations as described in the text.

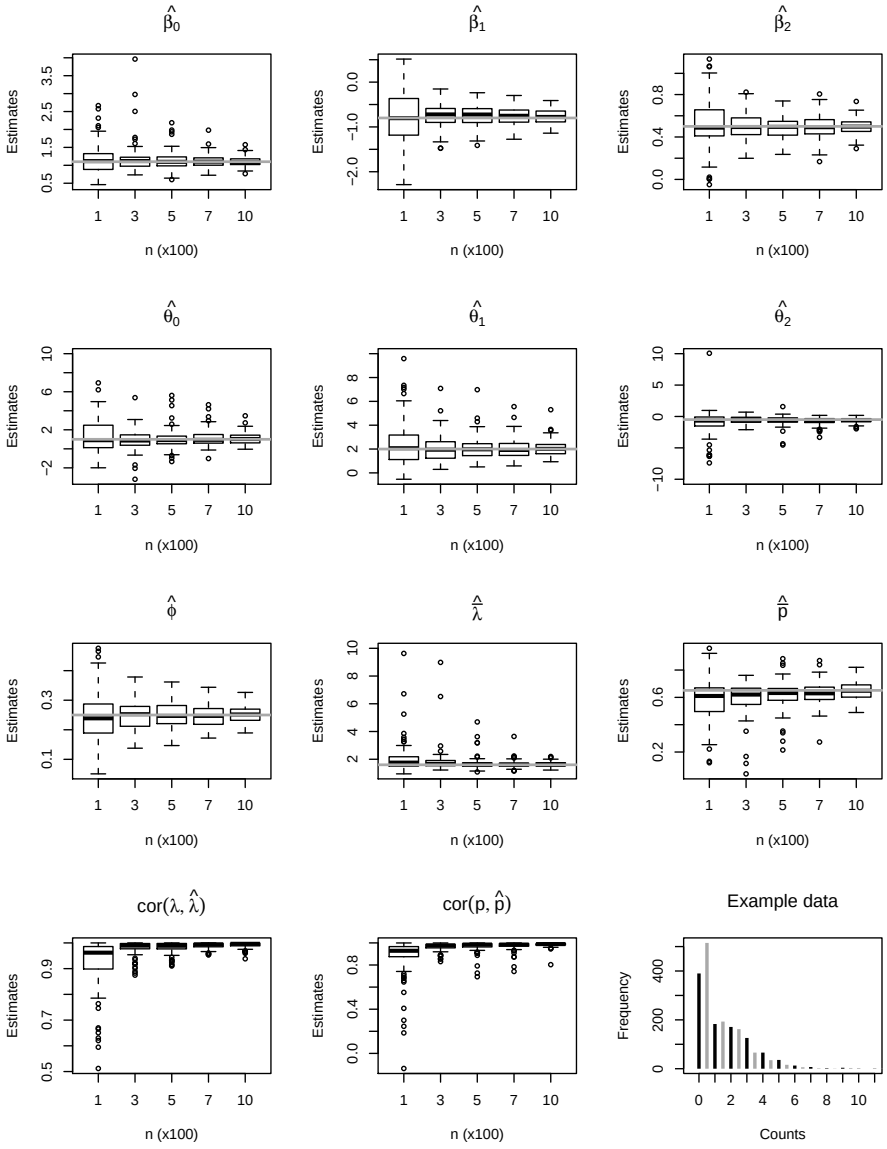
Covariates	Setup 1	Setup 2	Setup 3	Setup 4
$x_1 \sim \text{Uniform}(0, 1)$	β_1	β_1	β_1	β_1
$x_2 \sim \text{Normal}(0, 1)$	θ_1	θ_1	θ_1	θ_1
$x_3 \sim \text{Uniform}(-1, 1)$	β_2	–	β_2, θ_2	–
$x_4 \sim \text{Uniform}(-1, 1)$	θ_2	–	–	–
$x_5 \sim \text{Bernoulli}(0.6)$	–	β_2	–	β_2, θ_2
$x_6 \sim \text{Bernoulli}(0.4)$	–	θ_2	–	–

2 Separate continuous covariate (Setup 1)

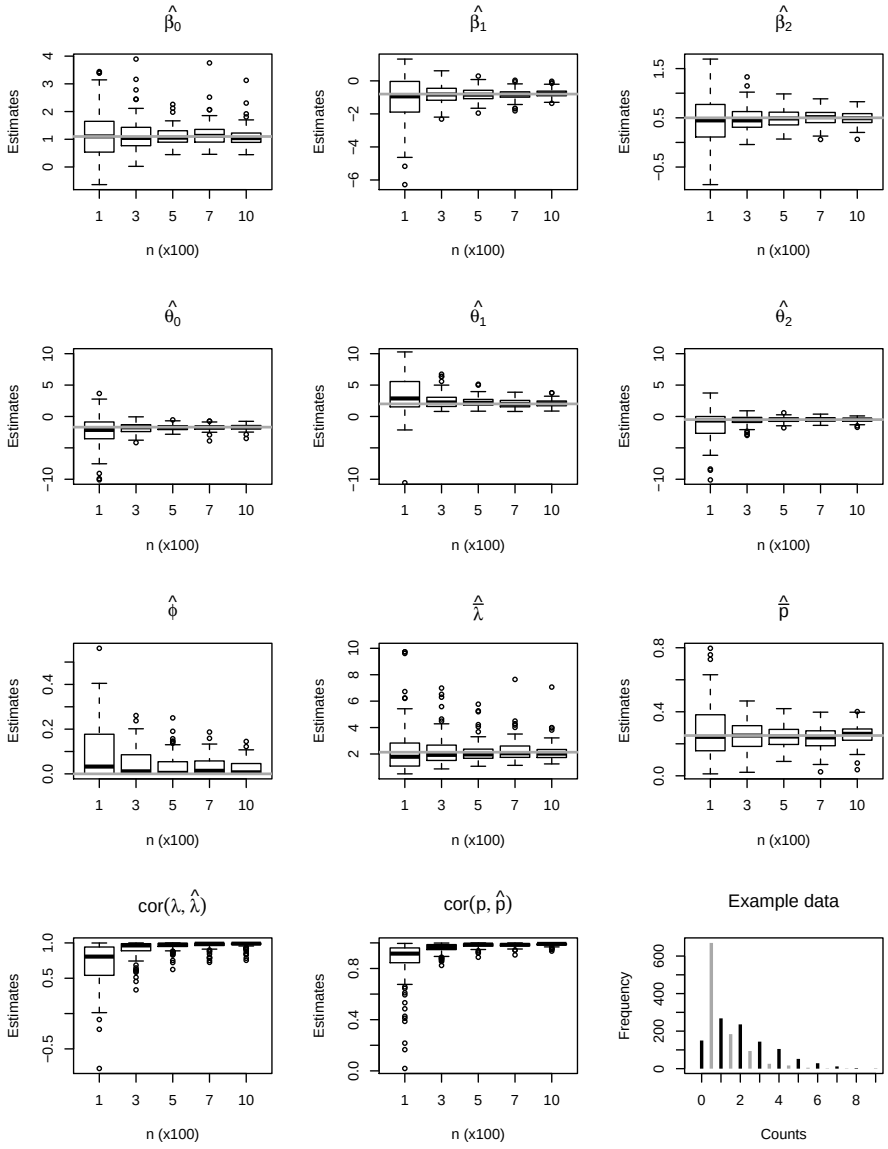
Setup 1, low abundance, zero inflated data, low probability of detection.



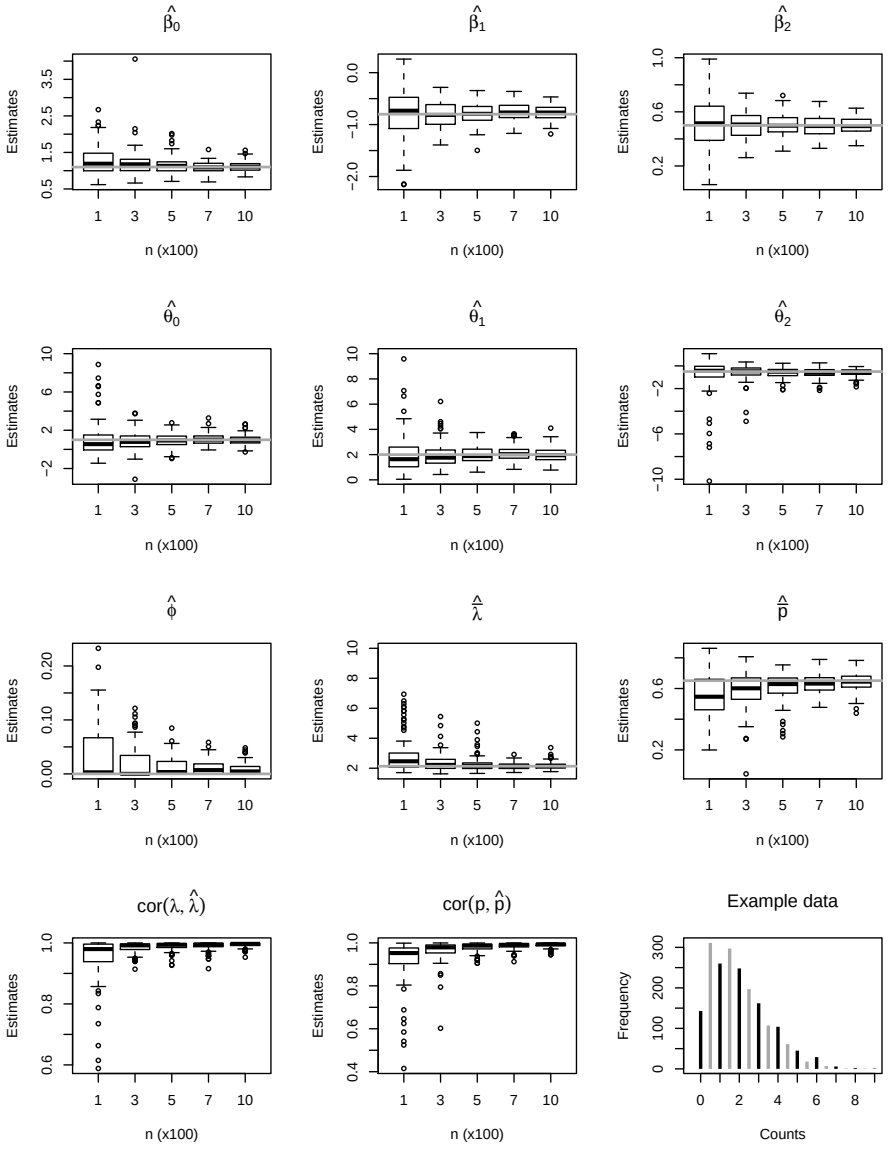
Setup 1, low abundance, zero inflated data, high probability of detection.



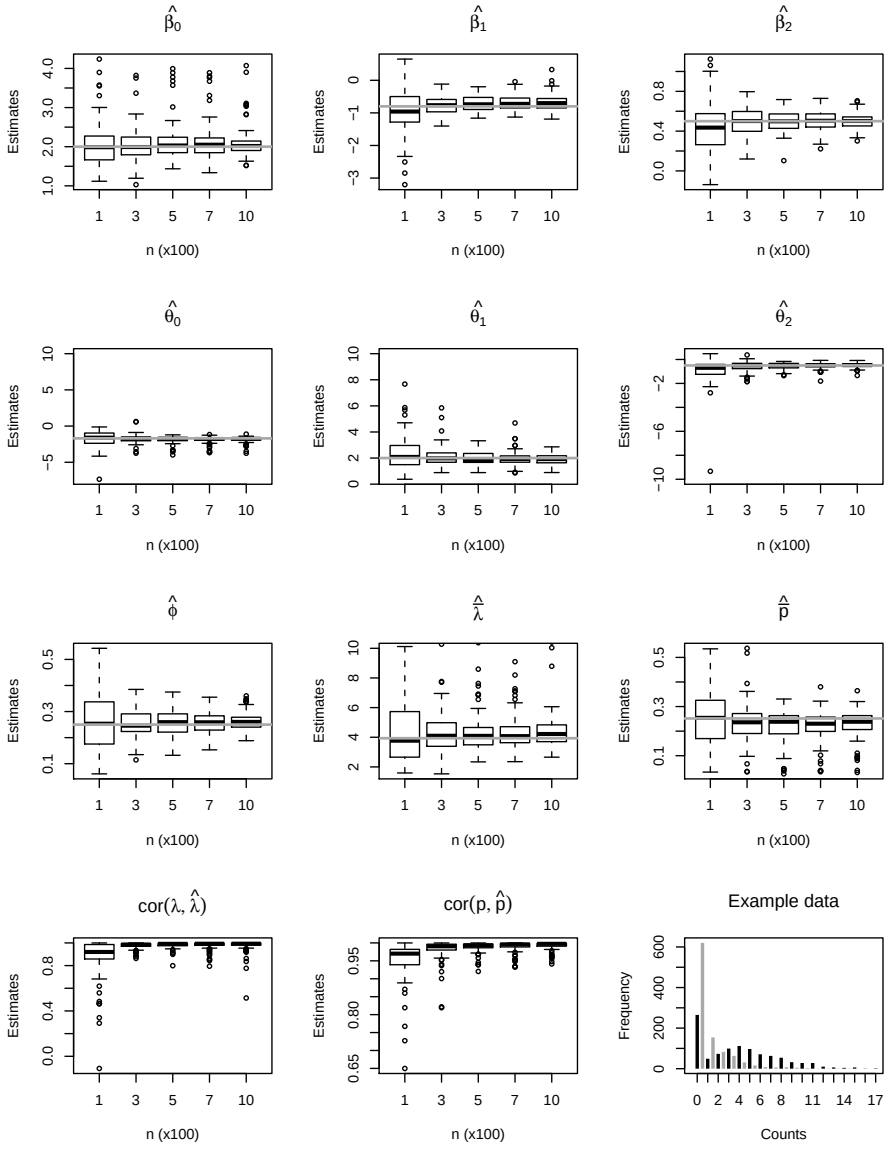
Setup 1, low abundance, not zero inflated data, low probability of detection.



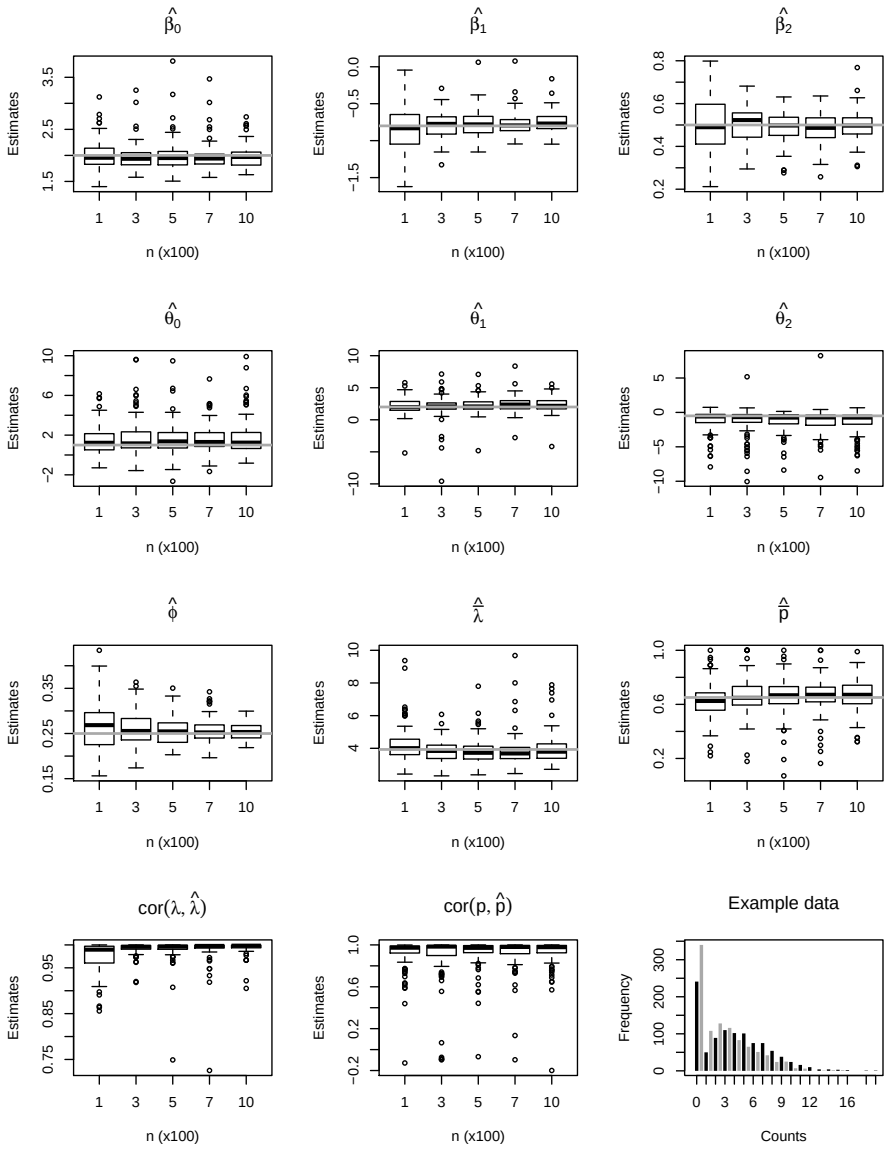
Setup 1, low abundance, not zero inflated data, high probability of detection.



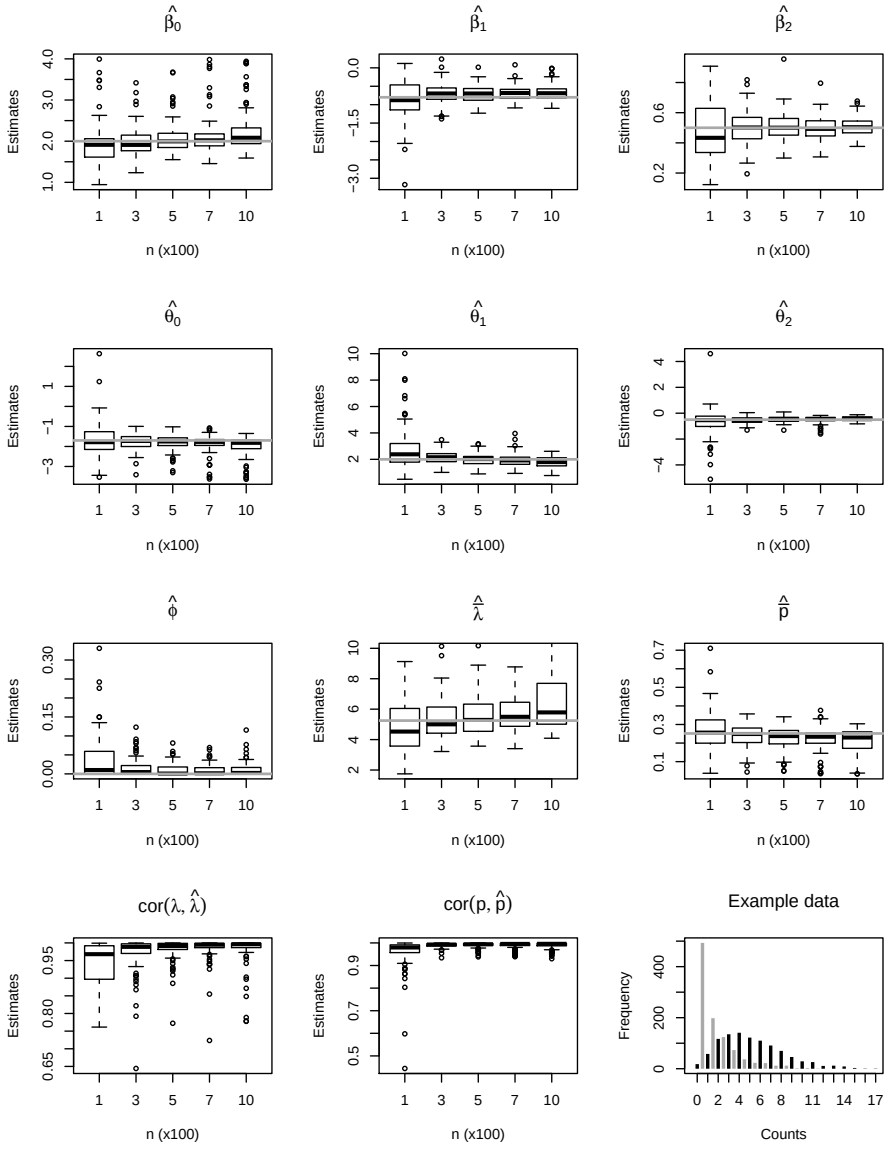
Setup 1, high abundance, zero inflated data, low probability of detection.



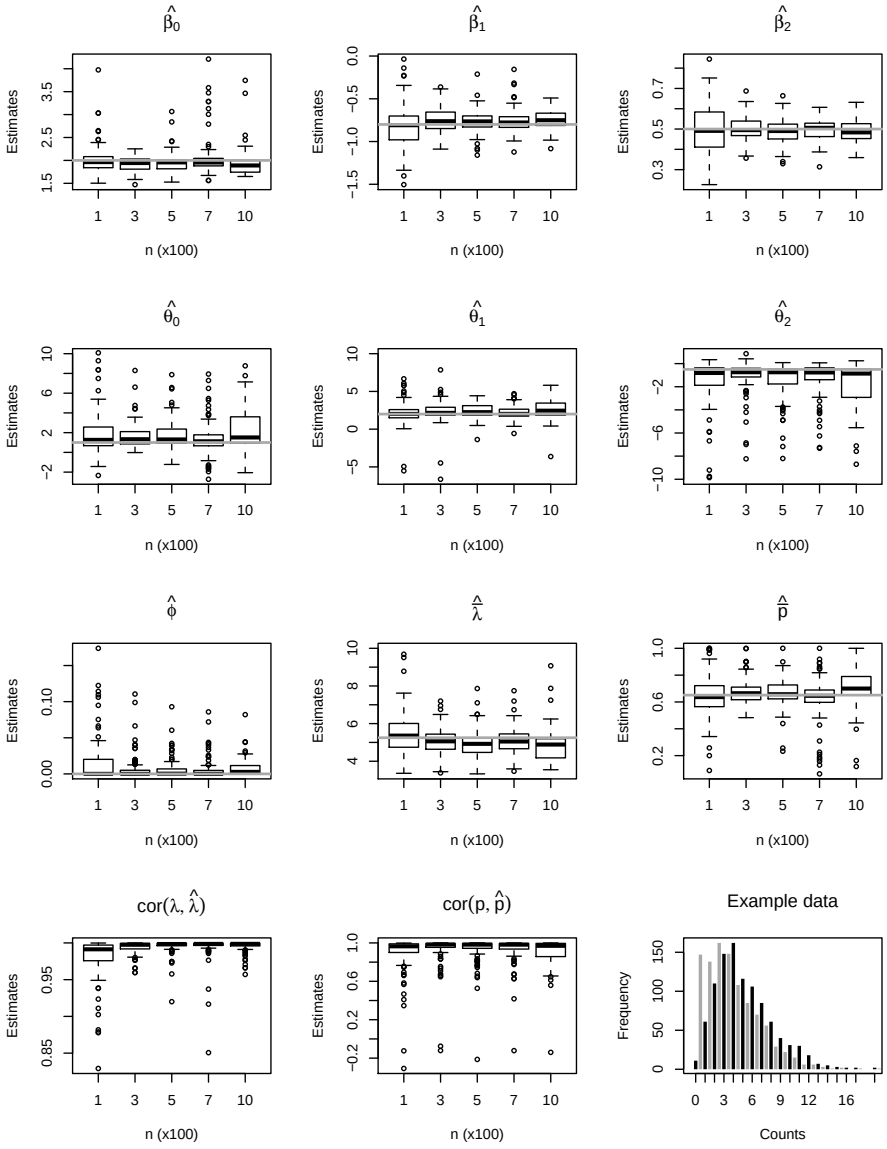
Setup 1, high abundance, zero inflated data, high probability of detection.



Setup 1, high abundance, not zero inflated data, low probability of detection.

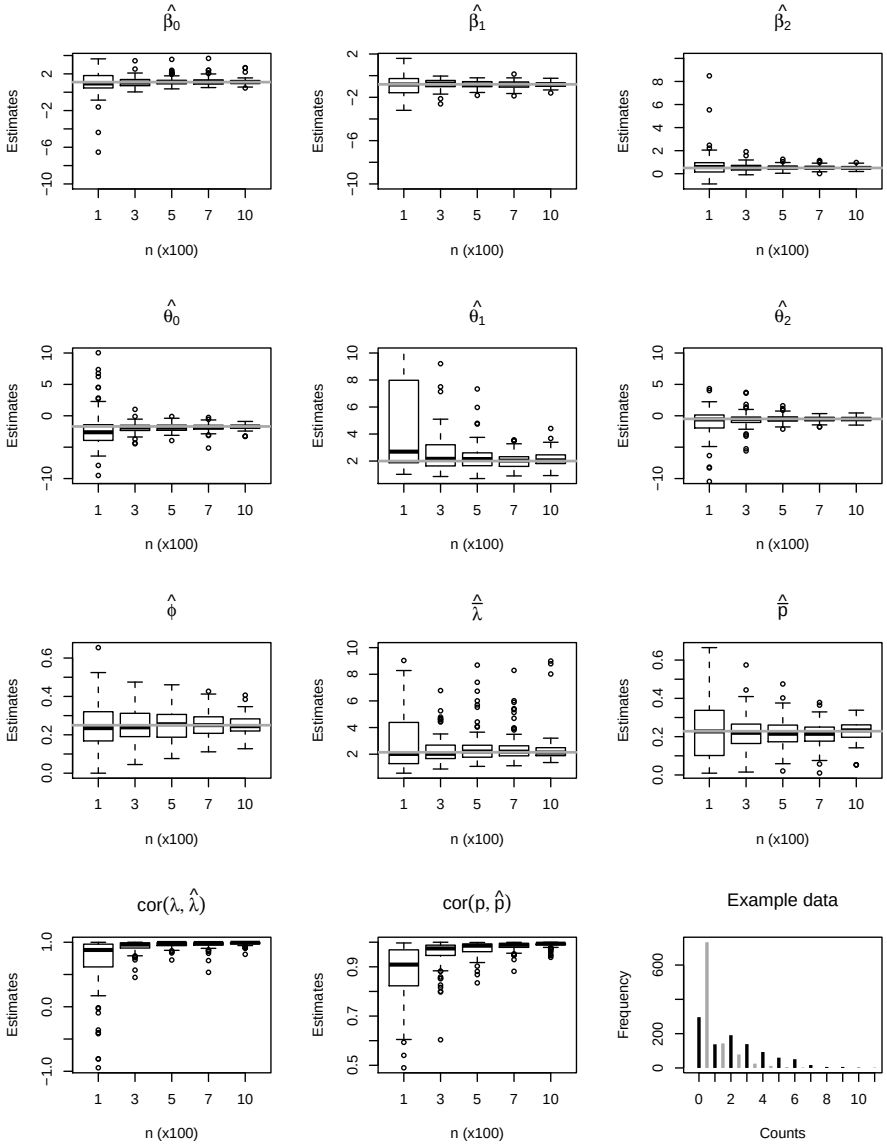


Setup 1, high abundance, not zero inflated data, high probability of detection.

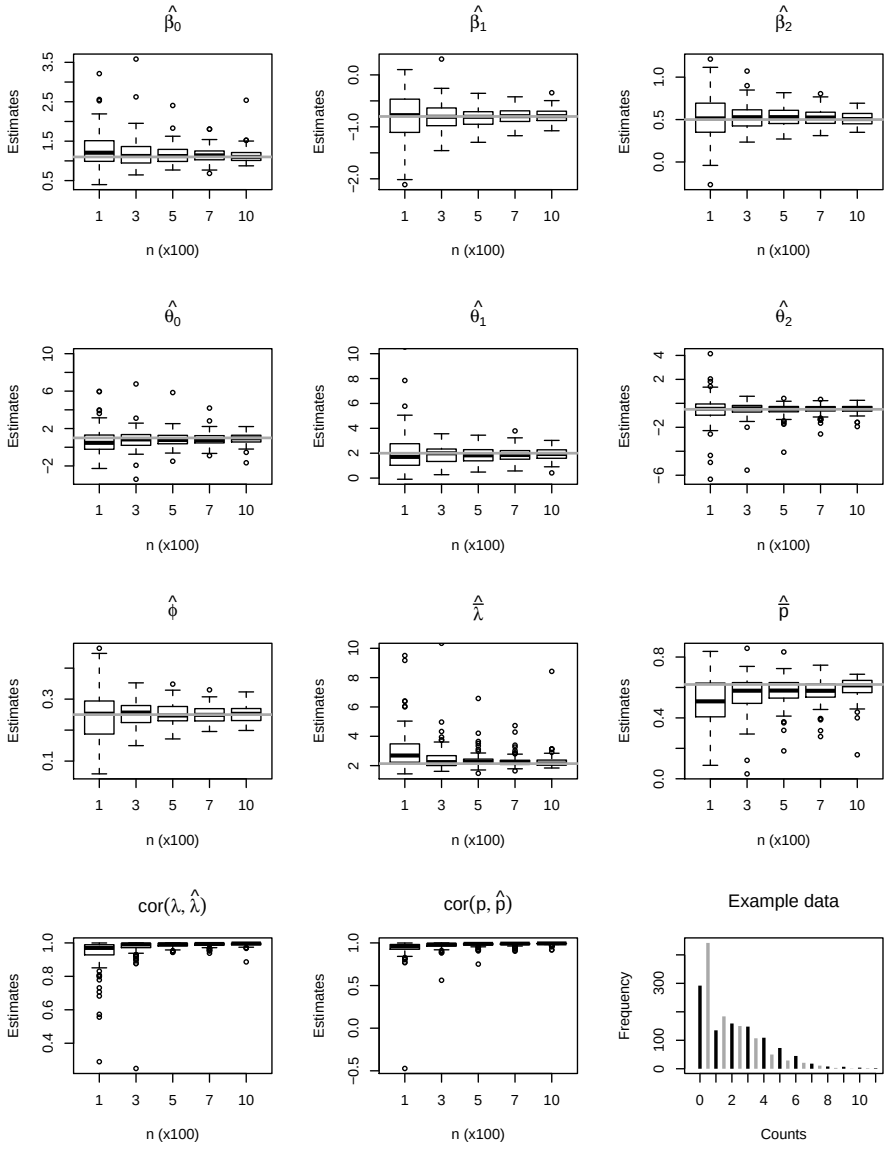


3 Separate discrete covariate (Setup 2)

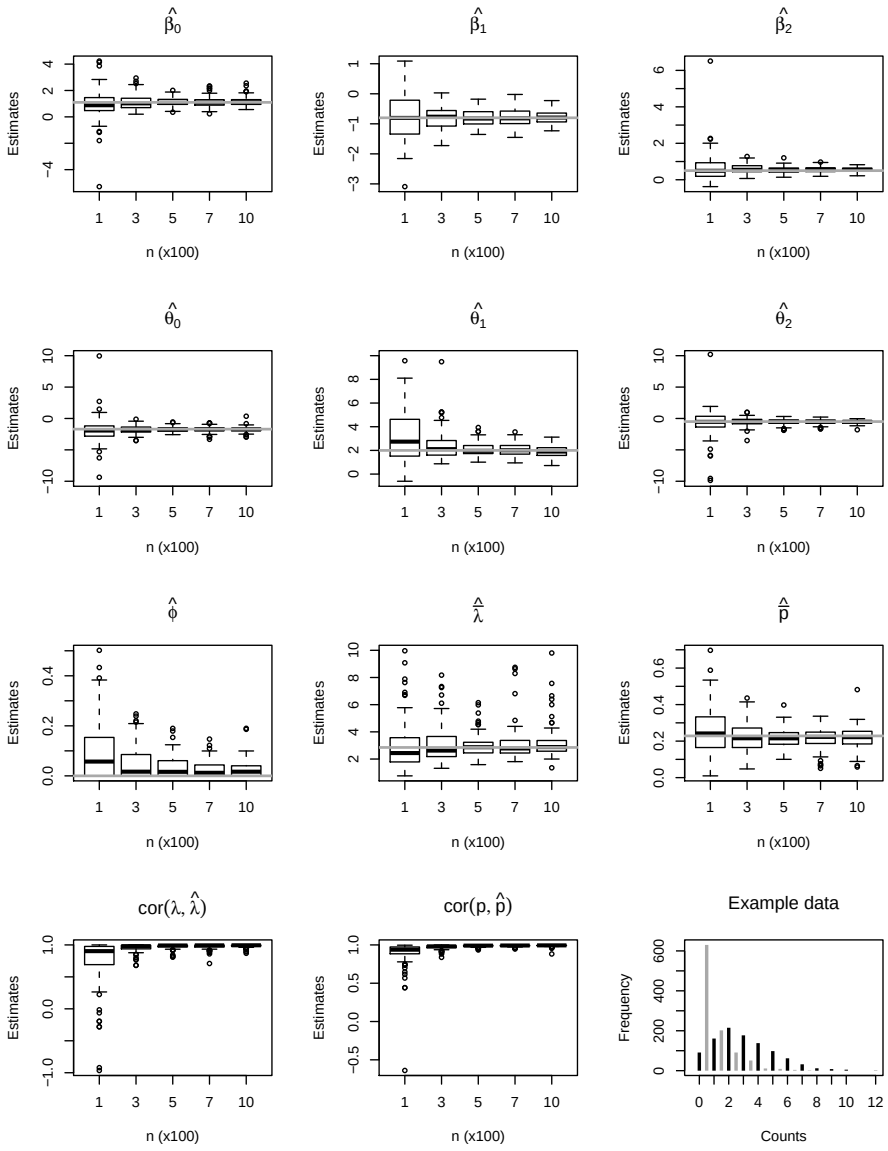
Setup 2, low abundance, zero inflated data, low probability of detection.



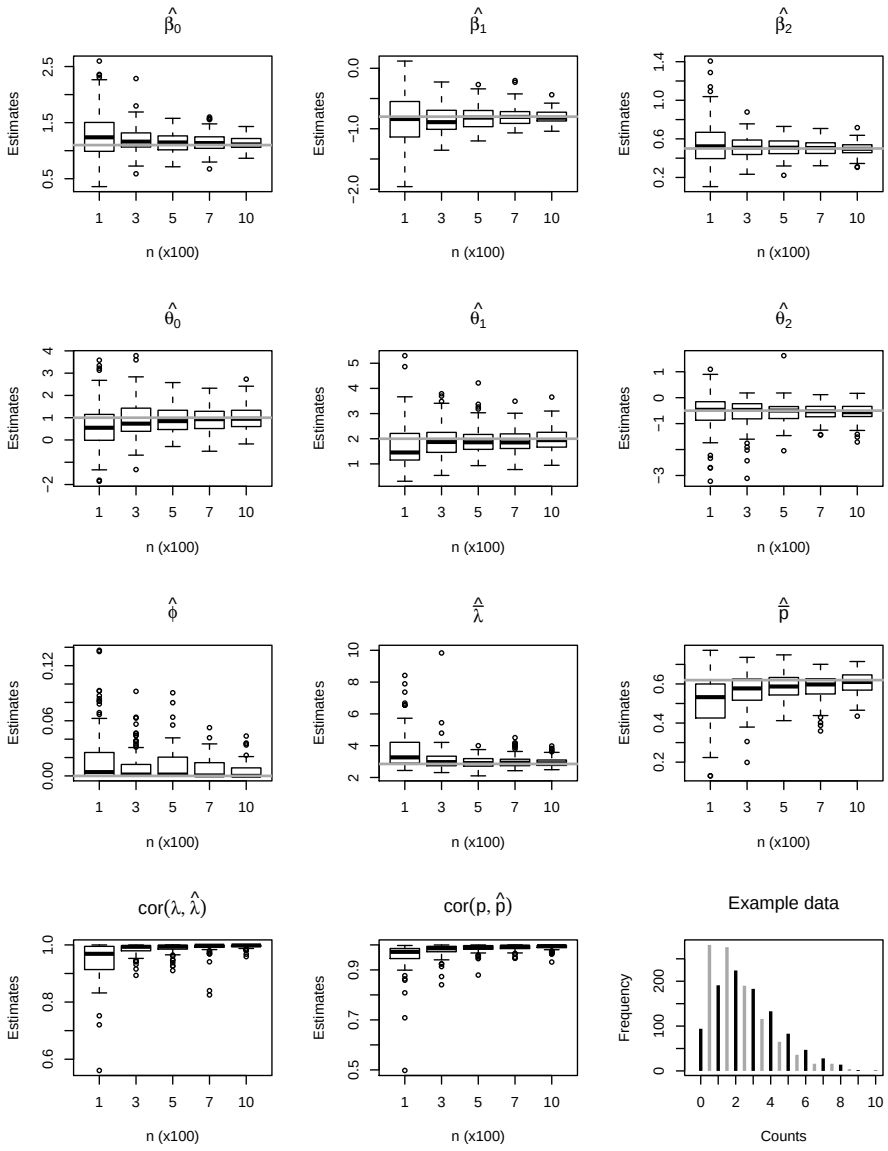
Setup 2, low abundance, zero inflated data, high probability of detection.



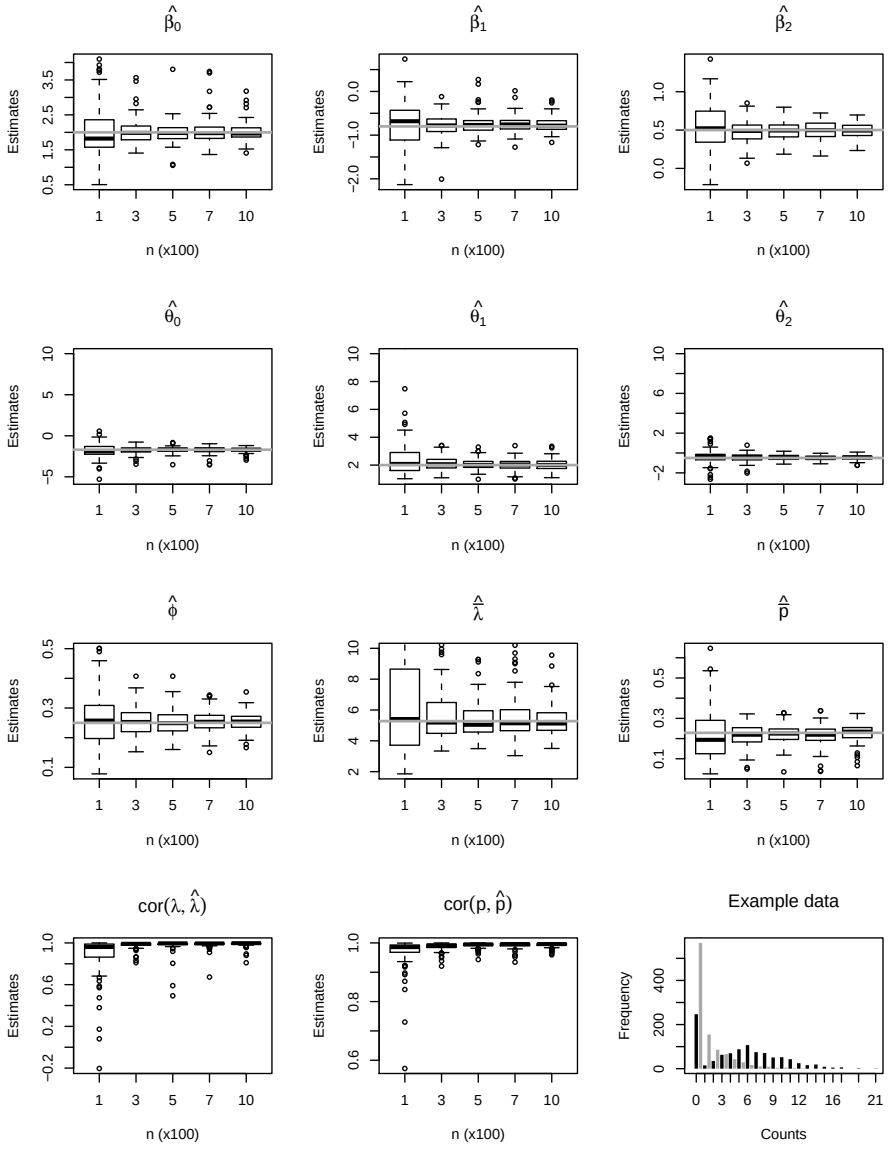
Setup 2, low abundance, not zero inflated data, low probability of detection.



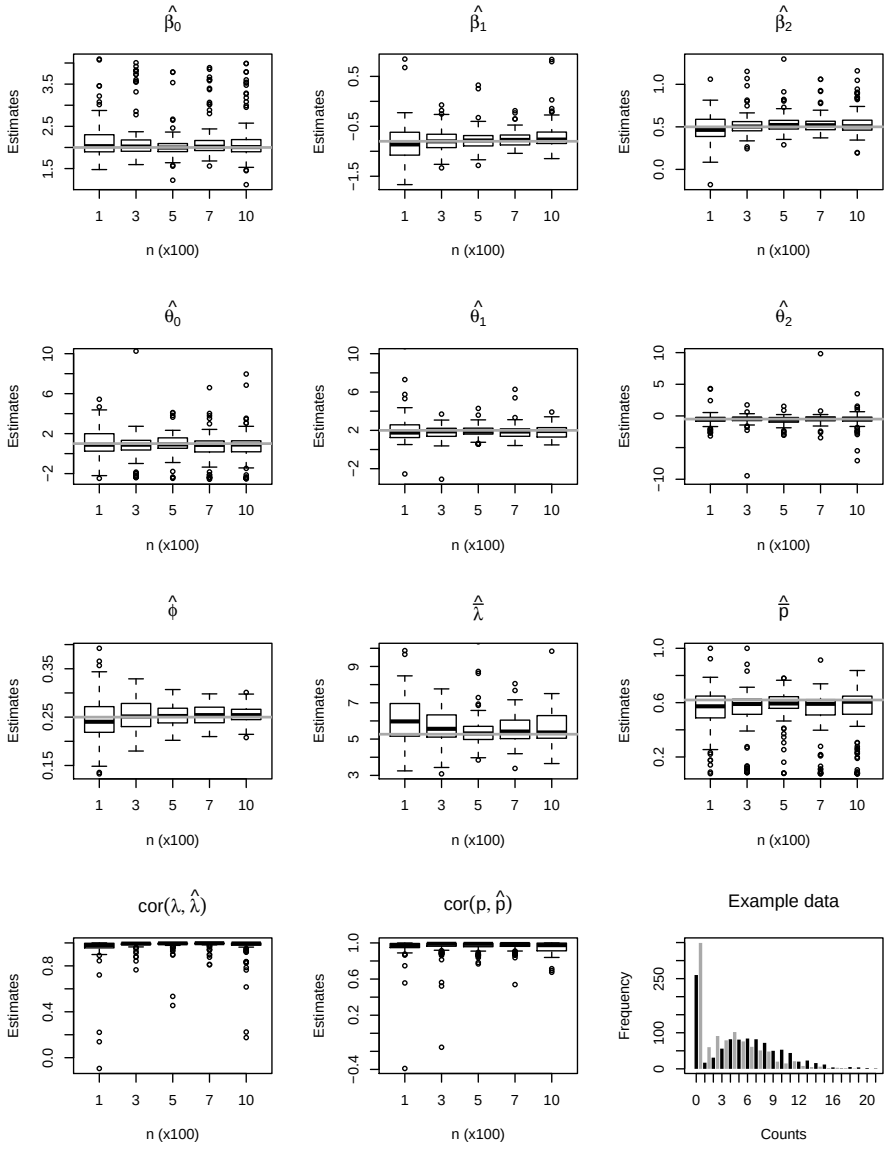
Setup 2, low abundance, not zero inflated data, high probability of detection.



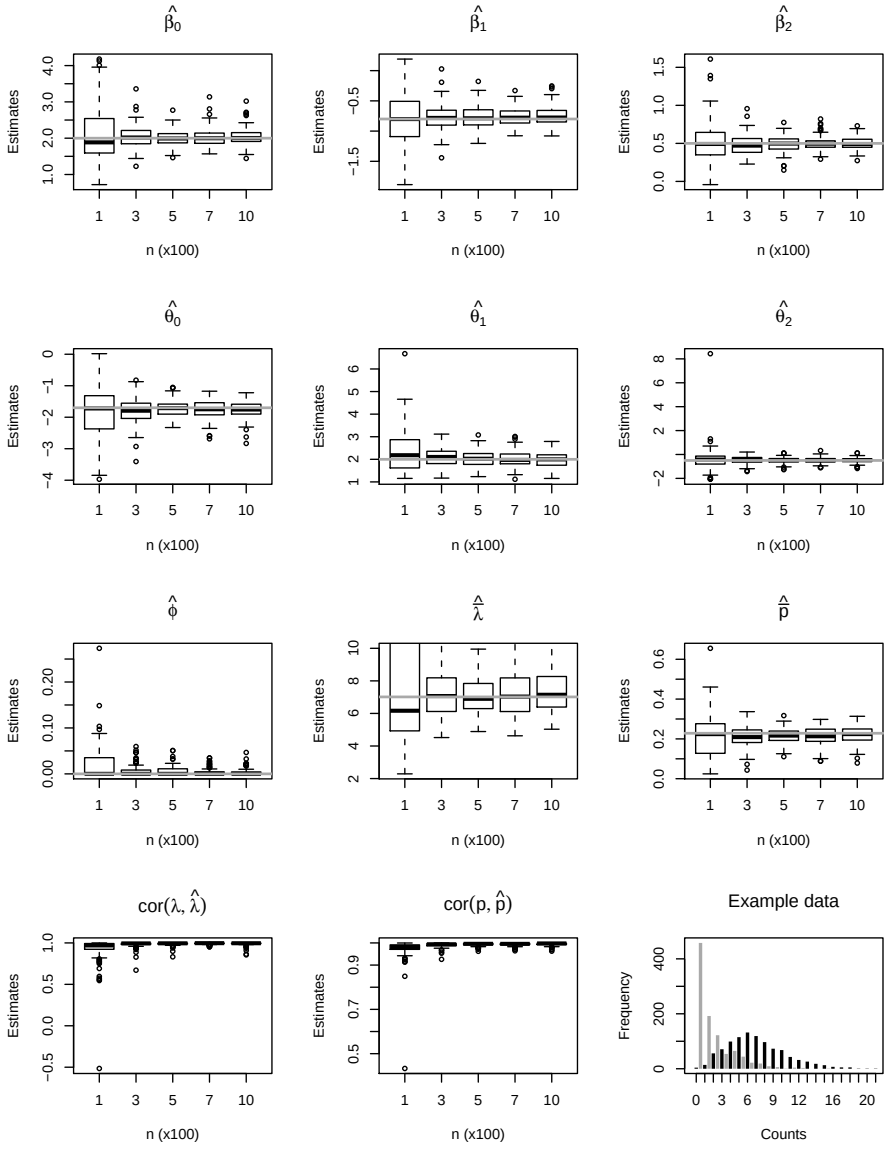
Setup 2, high abundance, zero inflated data, low probability of detection.



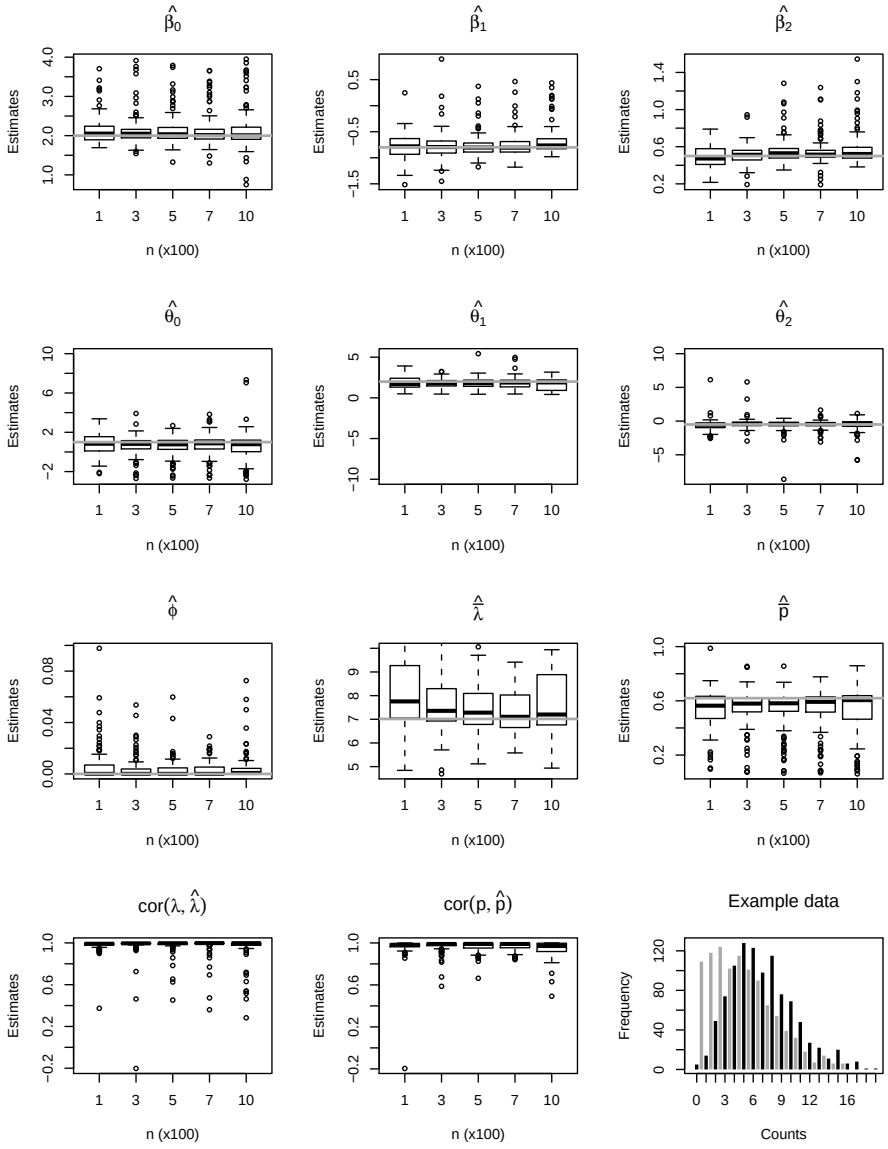
Setup 2, high abundance, zero inflated data, high probability of detection.



Setup 2, high abundance, not zero inflated data, low probability of detection.

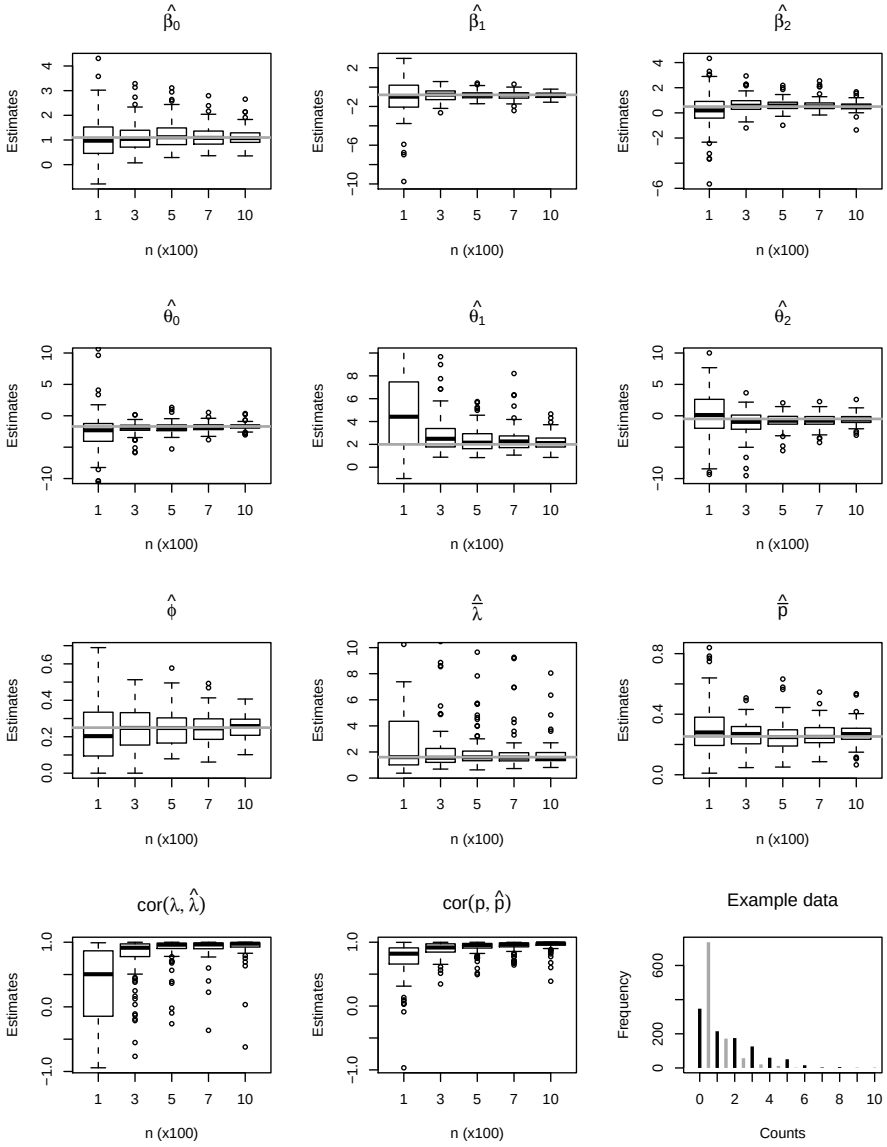


Setup 2, high abundance, not zero inflated data, high probability of detection.

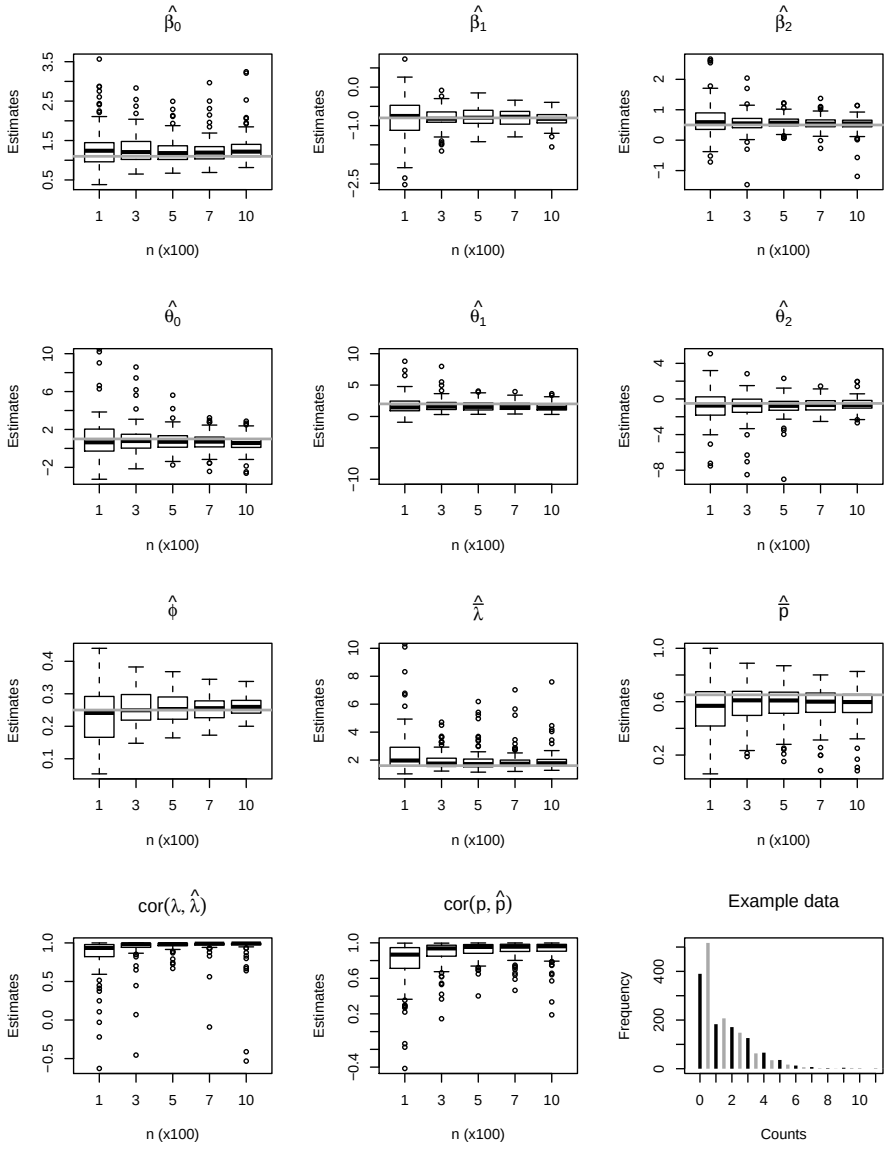


4 Common continuous covariate (Setup 3)

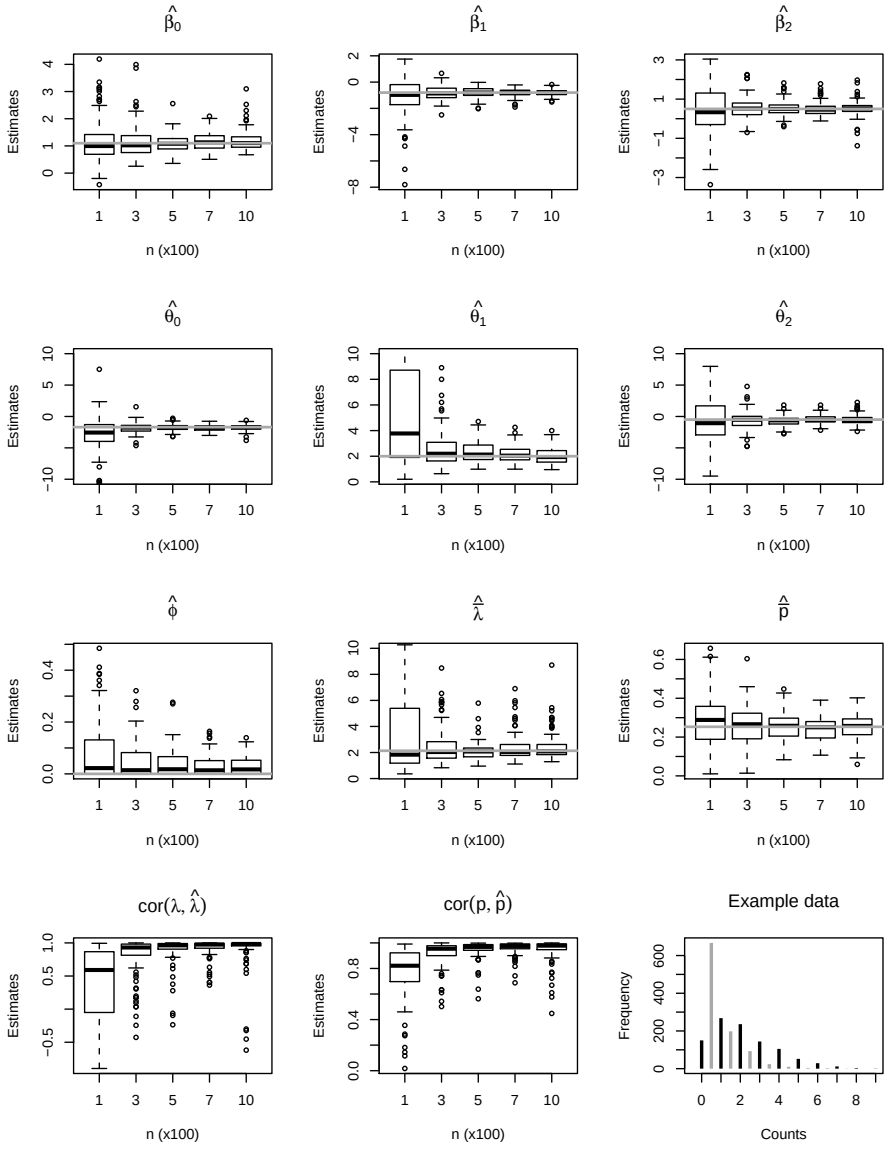
Setup 3, low abundance, zero inflated data, low probability of detection.



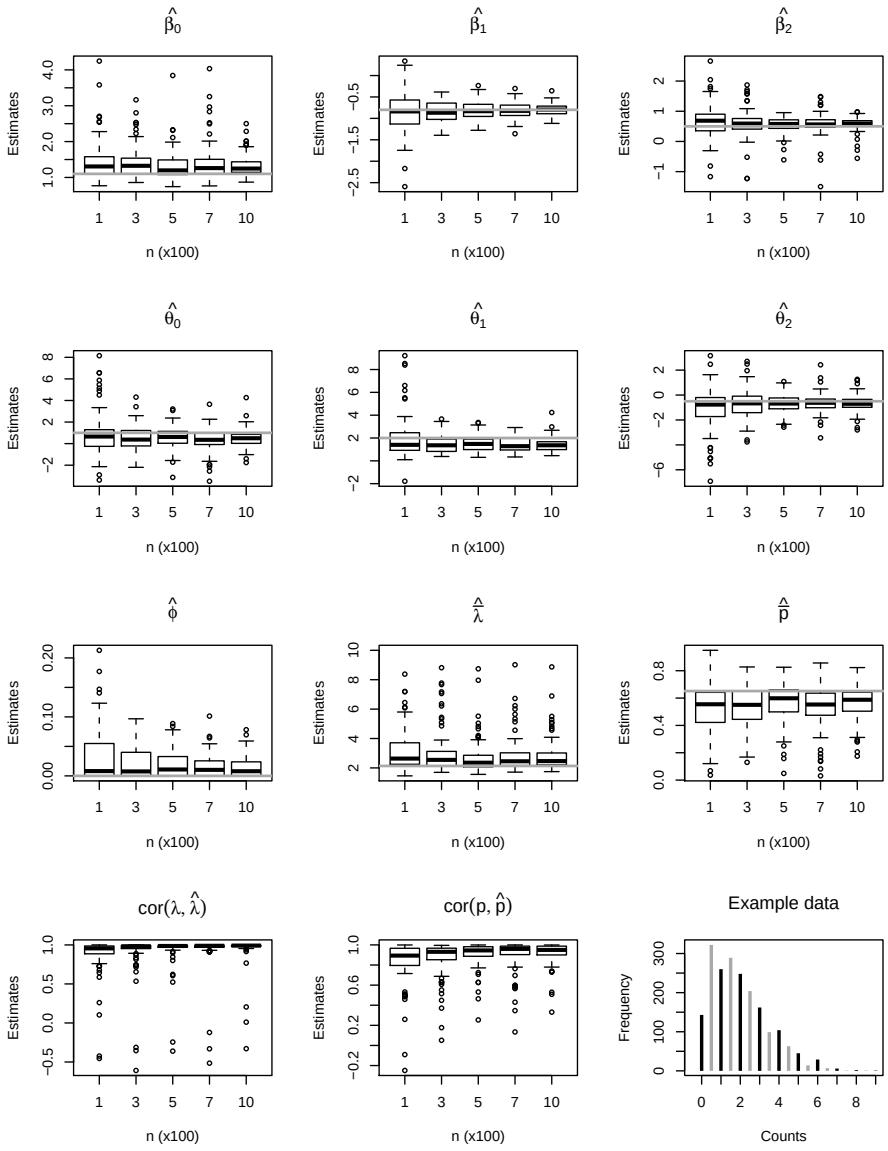
Setup 3, low abundance, zero inflated data, high probability of detection.



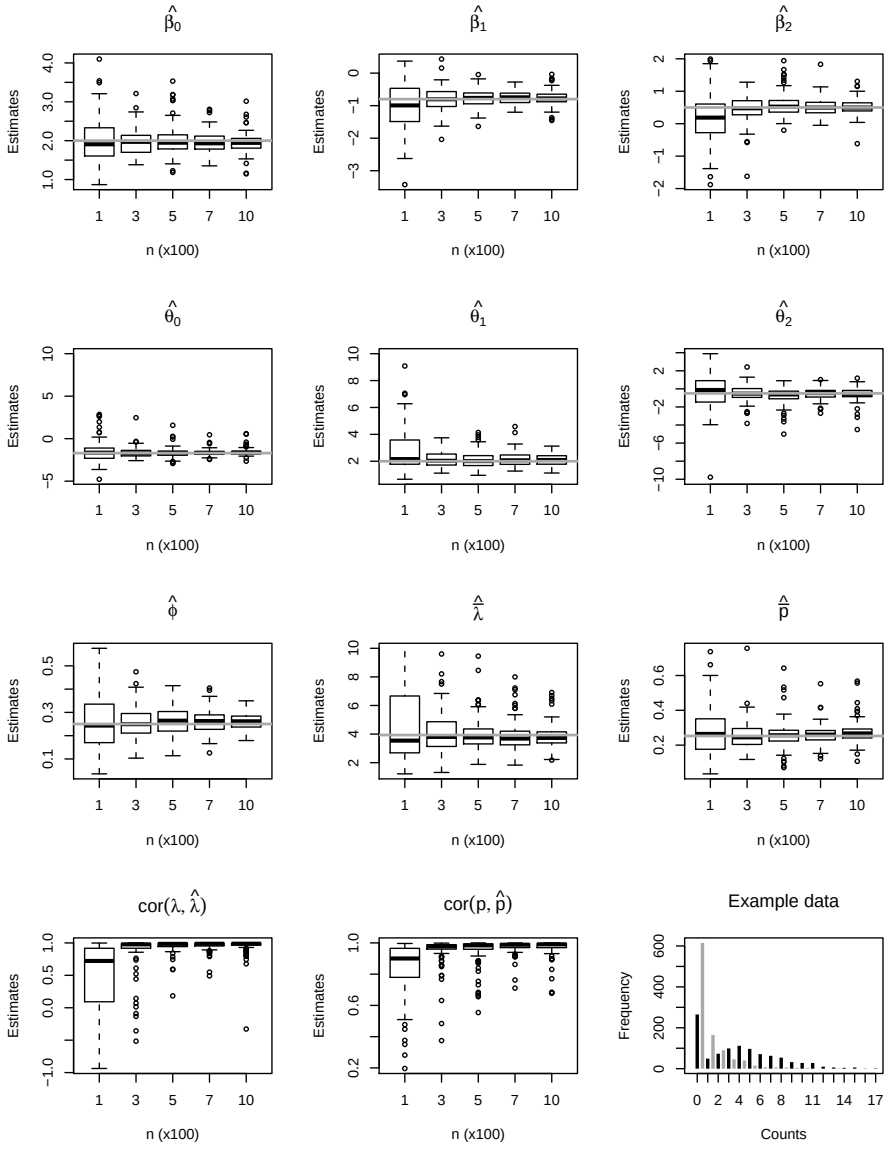
Setup 3, low abundance, not zero inflated data, low probability of detection.



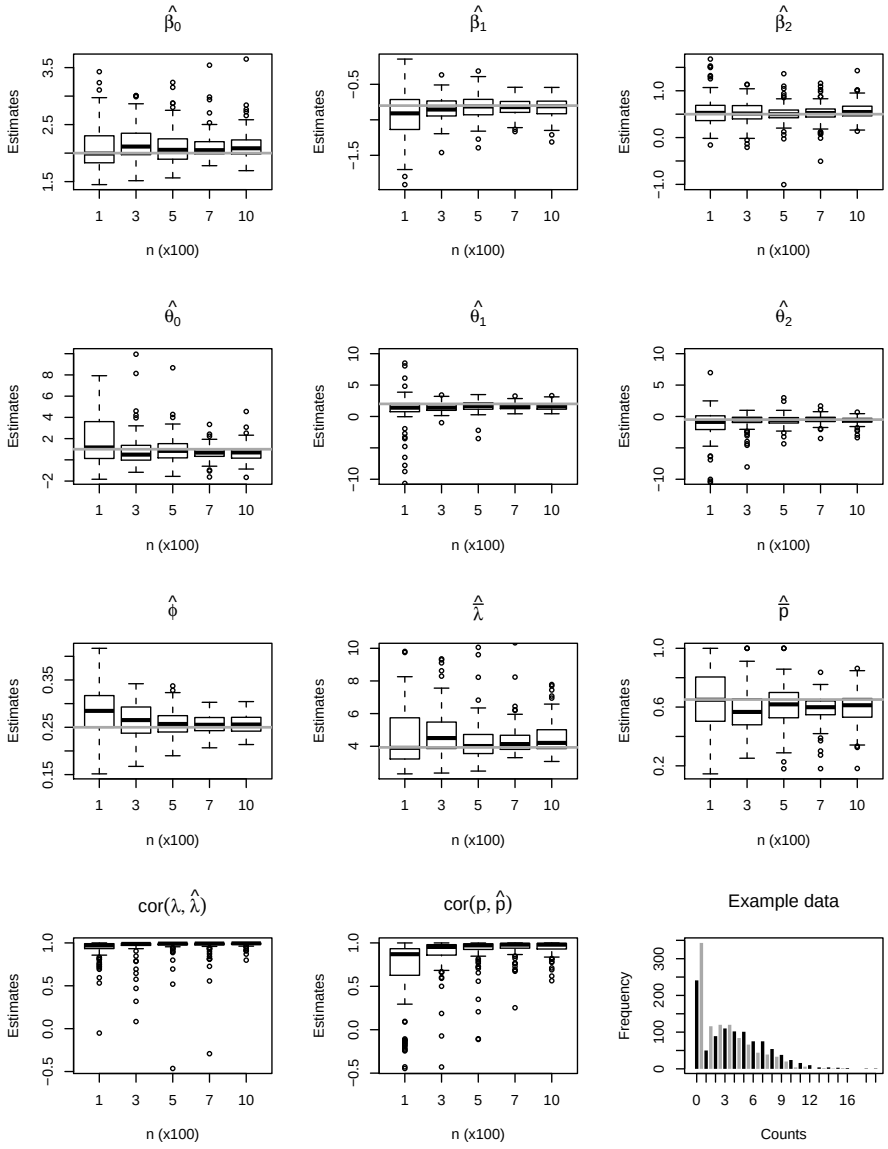
Setup 3, low abundance, not zero inflated data, high probability of detection.



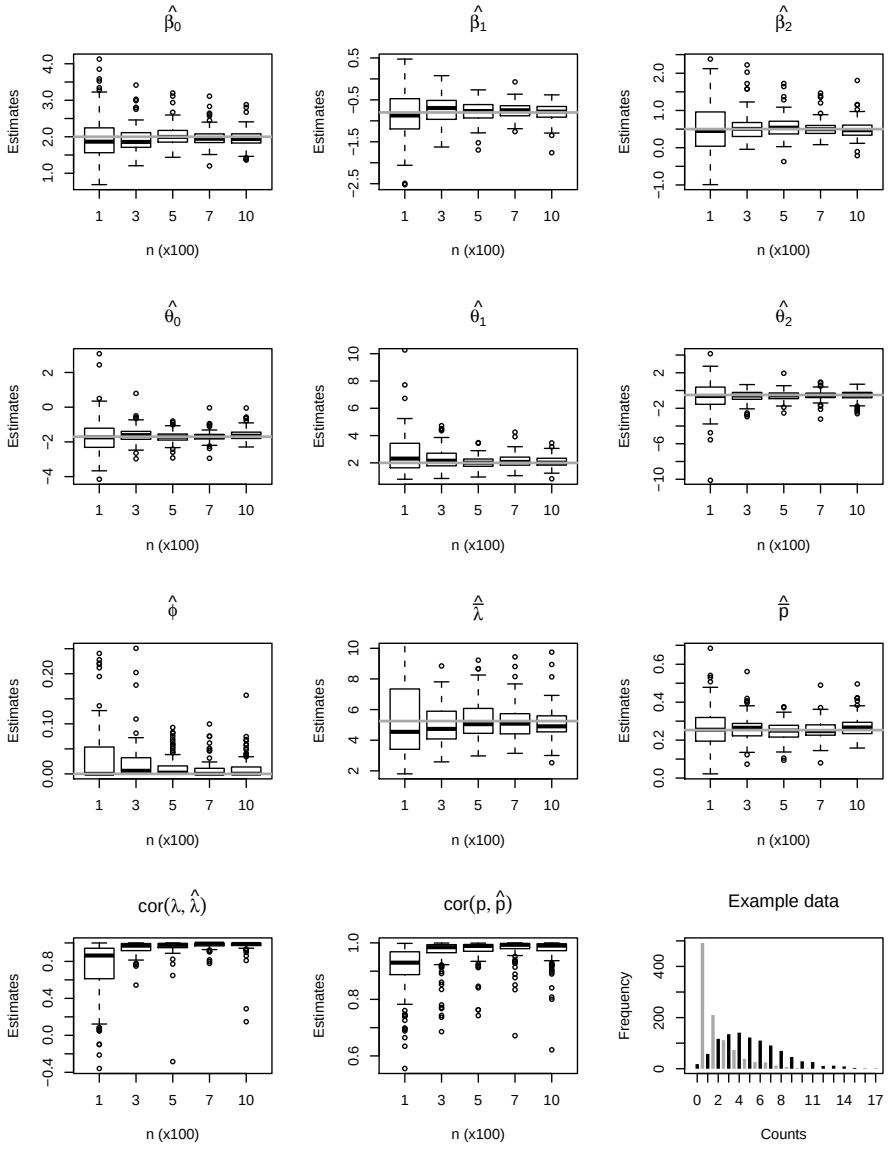
Setup 3, high abundance, zero inflated data, low probability of detection.



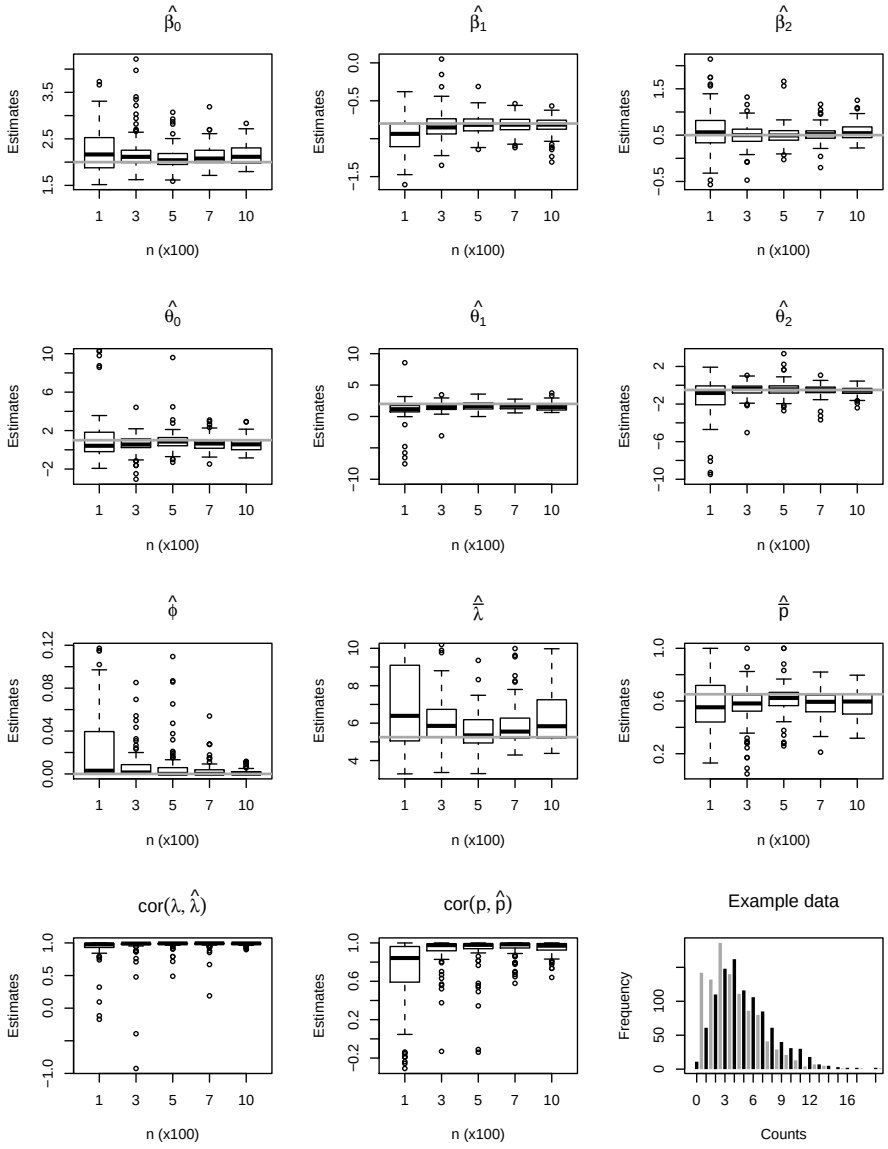
Setup 3, high abundance, zero inflated data, high probability of detection.



Setup 3, high abundance, not zero inflated data, low probability of detection.

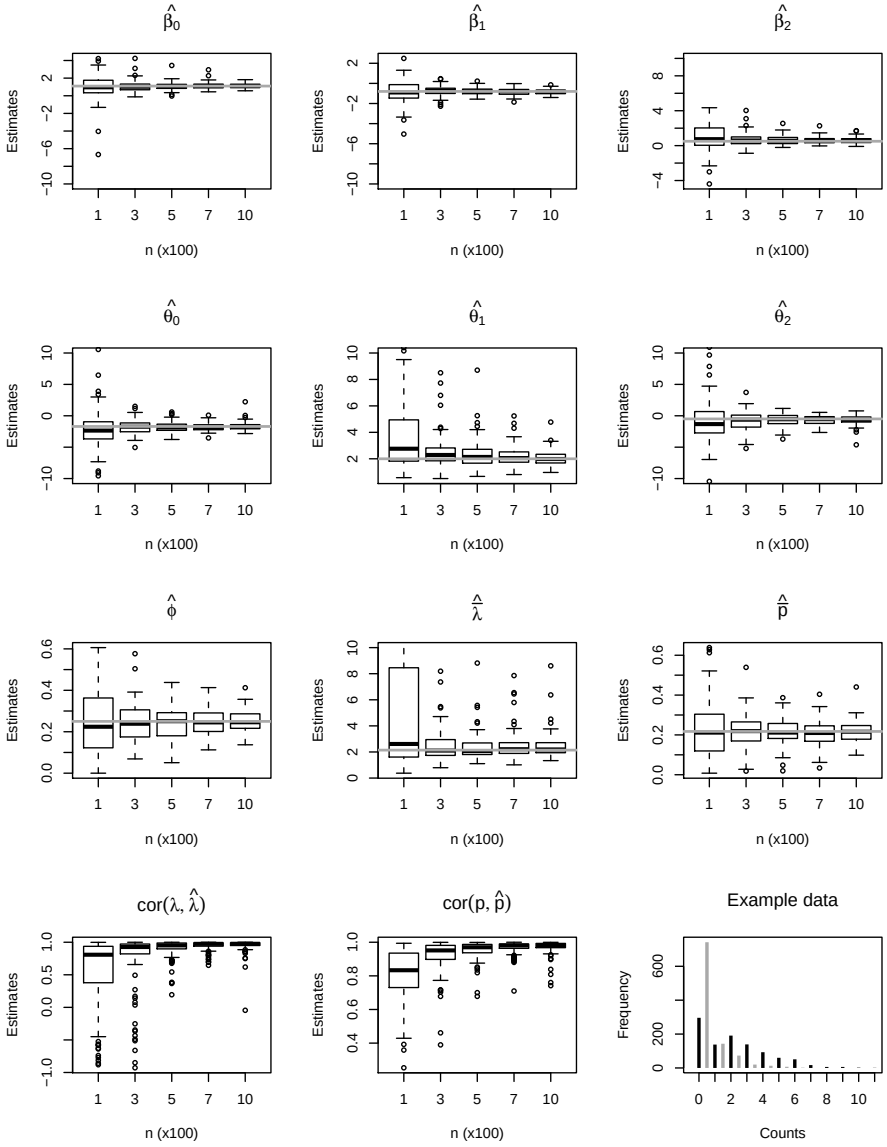


Setup 3, high abundance, not zero inflated data, high probability of detection.

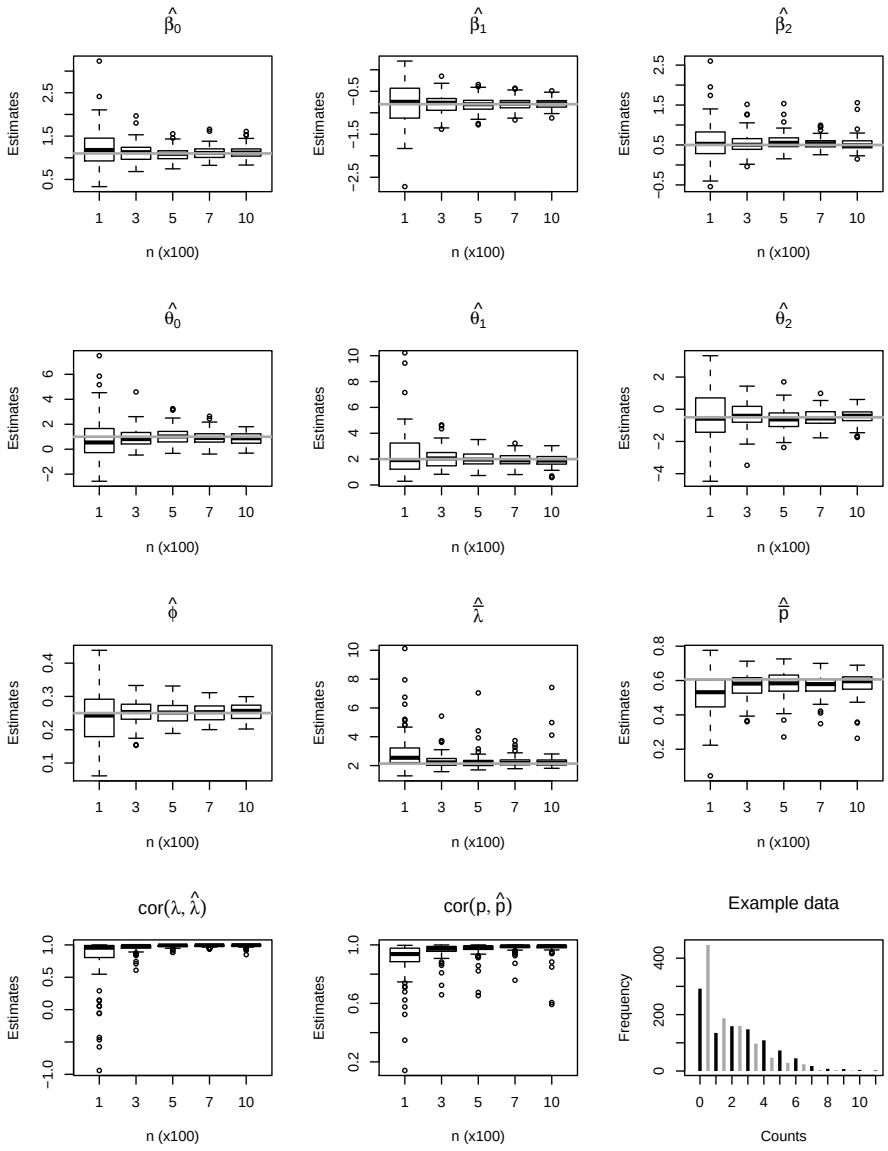


5 Common discrete covariate (Setup 4)

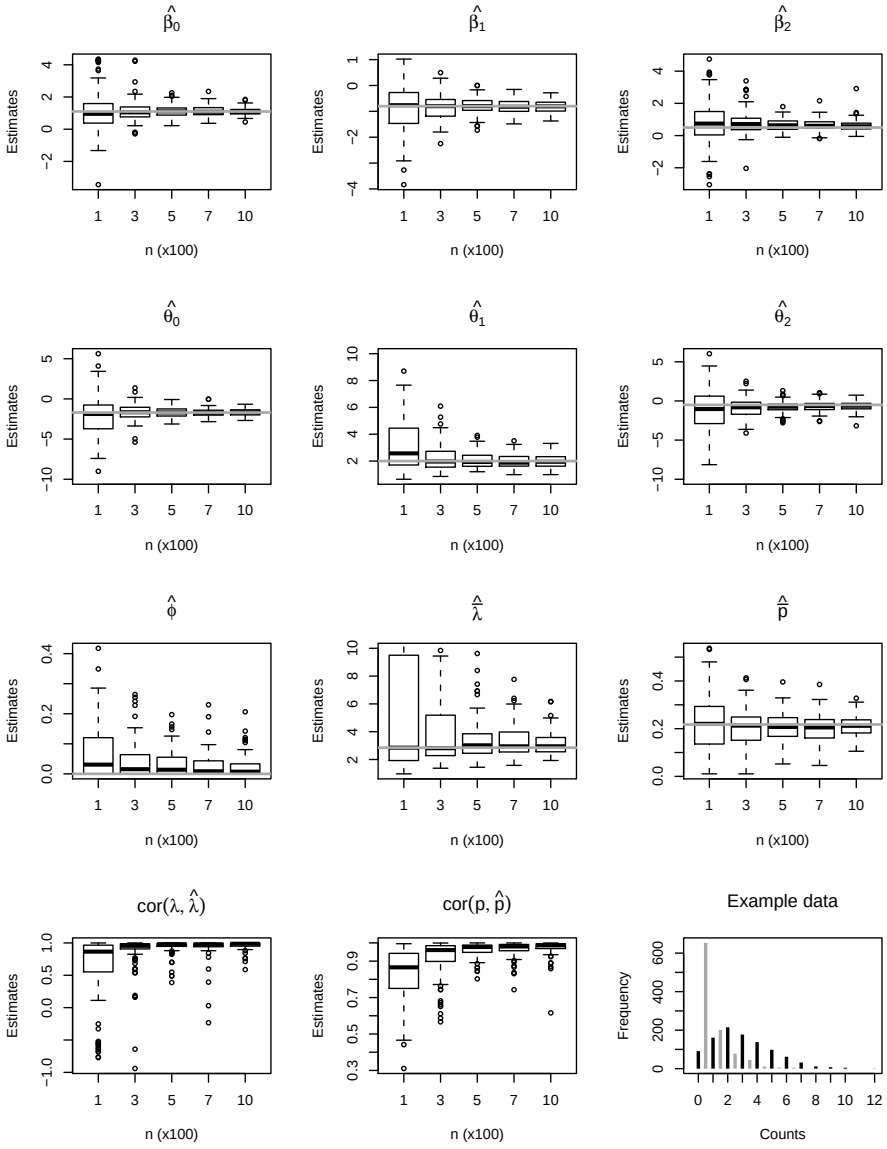
Setup 4, low abundance, zero inflated data, low probability of detection.



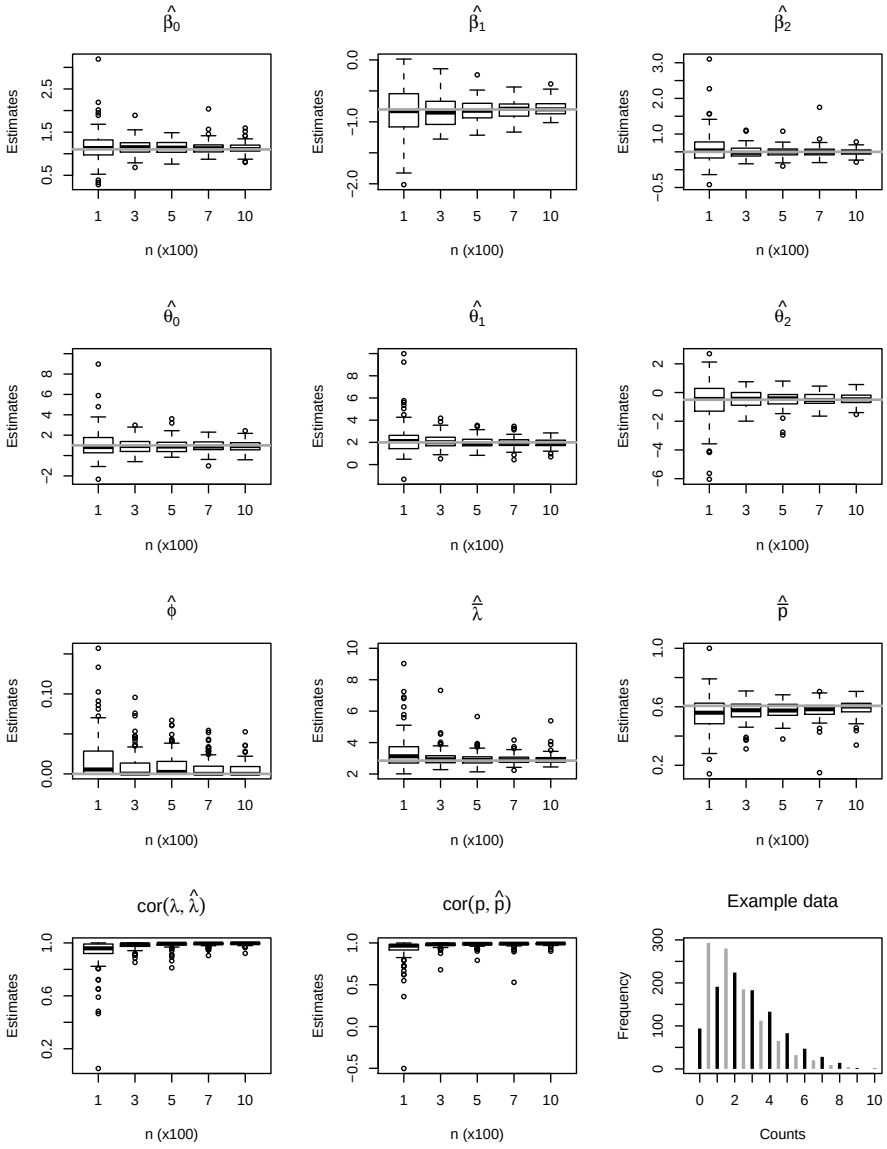
Setup 4, low abundance, zero inflated data, high probability of detection.



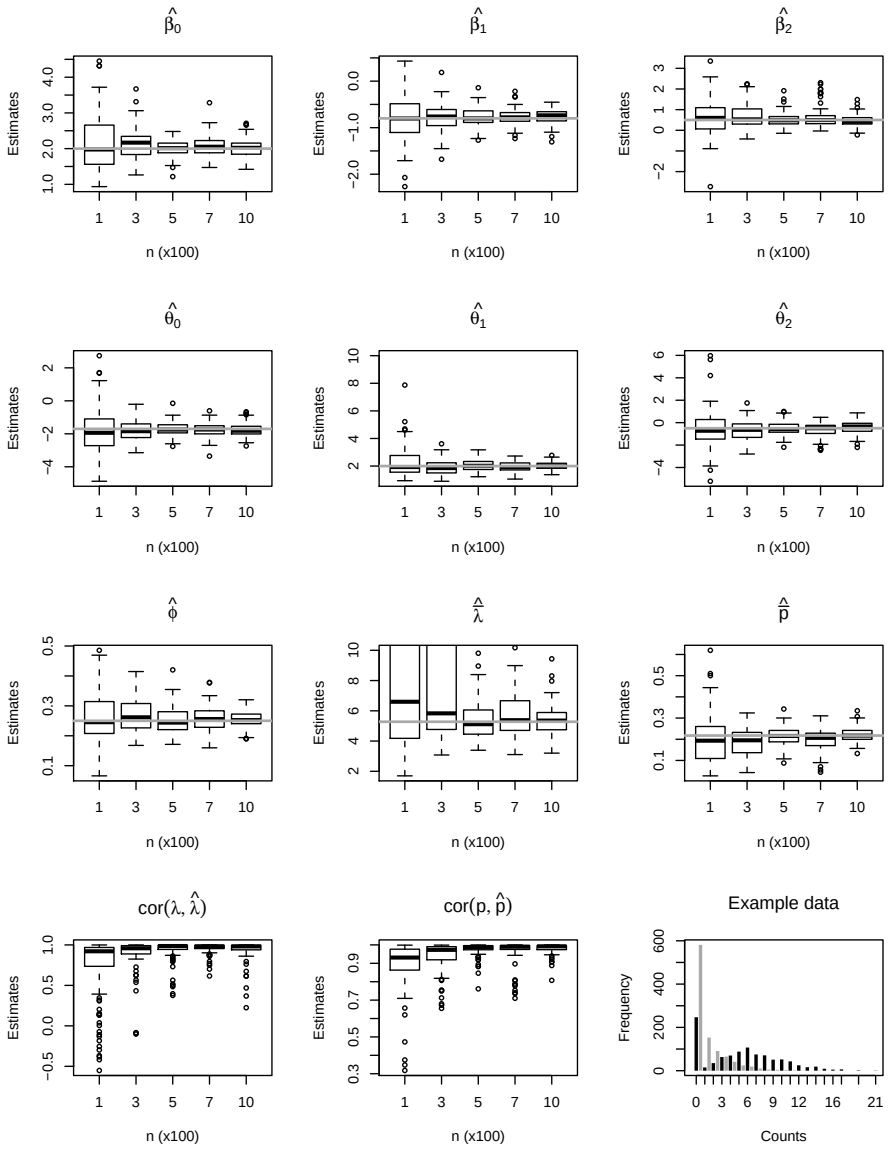
Setup 4, low abundance, not zero inflated data, low probability of detection.



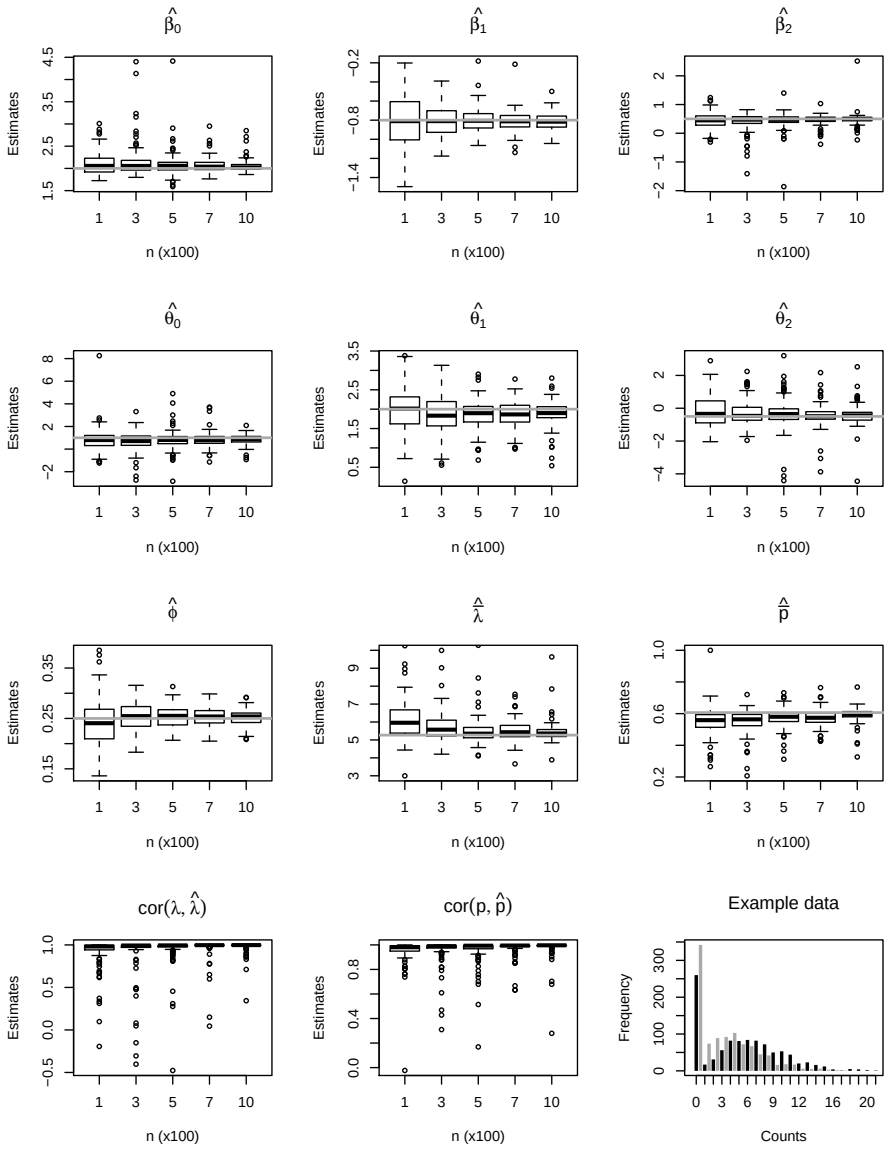
Setup 4, low abundance, not zero inflated data, high probability of detection.



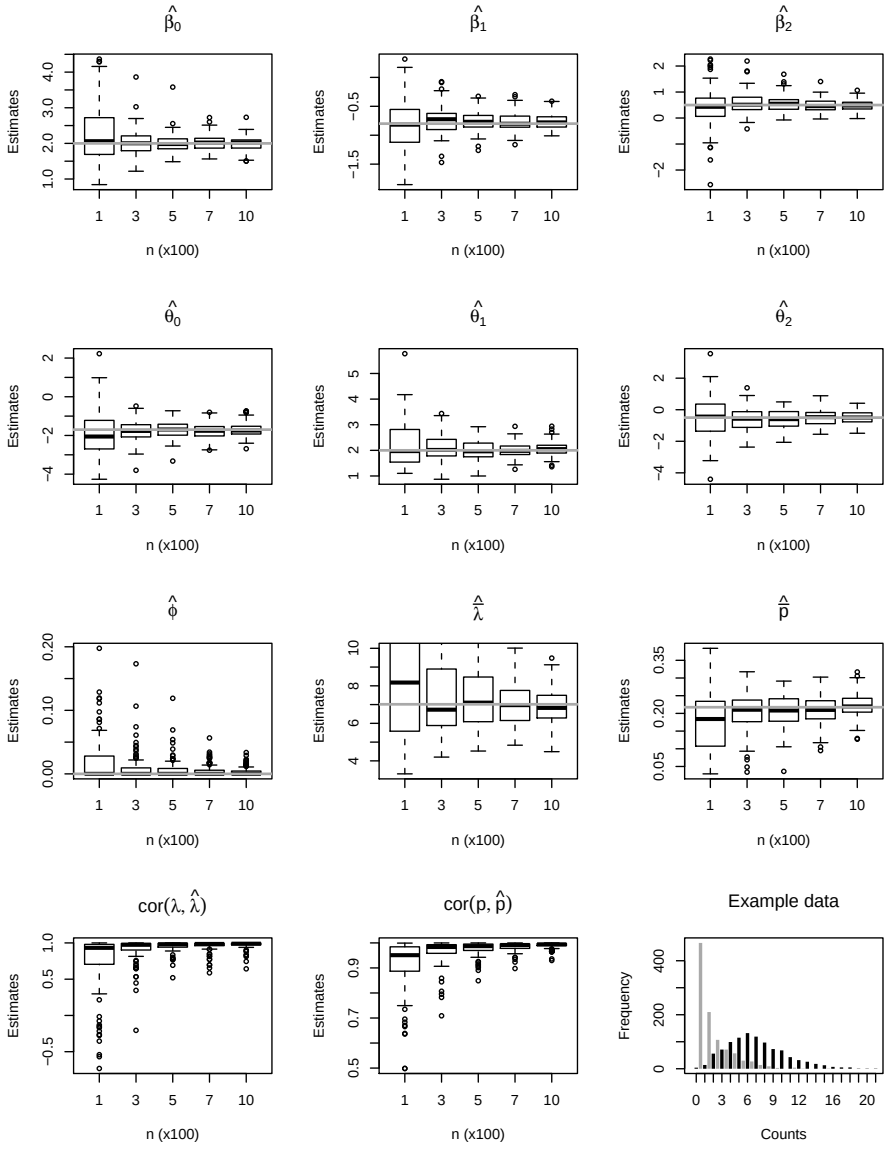
Setup 4, high abundance, zero inflated data, low probability of detection.



Setup 4, high abundance, zero inflated data, high probability of detection.



Setup 4, high abundance, not zero inflated data, low probability of detection.



Setup 4, high abundance, not zero inflated data, high probability of detection.