

# SUPPLEMENTARY MATERIALS: SPARSE POISSON LOG NORMAL MODEL FOR ZERO-INFLATED MULTIVARIATE COUNT DATA

Togo Jean Yves KIOYE<sup>1</sup>, Paul-Marie  
GROLLEMUND<sup>1,2</sup>, Jocelyn CHAUVET<sup>3,4</sup>,  
Christophe CHASSARD<sup>1</sup>

<sup>1</sup> *Unité Mixte de Recherche sur le Fromage  
(UMRF), Université Clermont Auvergne,  
INRAE, France*

*togo-jean.yves.kioye@uca.fr,  
christophe.chassard@inrae.fr*

<sup>2</sup> *Laboratoire de Mathématiques Blaise Pascal  
(LMBP), Université Clermont Auvergne, France*

*paul.marie.grollemund@uca.fr,  
pierre.druilhet@uca.fr*

<sup>3</sup> *Centre de recherche de l'ICES, France*

<sup>4</sup> *Laboratoire Angevin de Recherche en  
Ingénierie des Systèmes (LARIS), France*  
*jchauvet@ices.fr*

## 1 Experiment 3: Performance evaluation across species sizes

### Experiment design

In this study, we consider a fixed coefficient matrix  $B$  in which 25% of the entries are randomly set to

zero. The sample size is fixed at  $n = 2000$ , while the number of columns, or species size, varies from 5 to 40. For each species size, we estimate the parameters and assess the RMSE between the true values of some parameters and their corresponding estimates, as well as the TPR and TNR of the penalized coefficients. Additionally, we compute the BIC and ICL to evaluate model performance. This experiment is repeated over 100 datasets.

### Experiment results

The left plot in Figure 1 presents the violin plots of the RMSE of  $B$ , while the right plot presents the RMSE of  $Y$ . The left plot of Figure 2 shows the violin plots of the TPR, while the right plot shows those of the TNR. Finally, Figure 3 presents the violin plots of the BIC and ICL on the left and right, respectively.

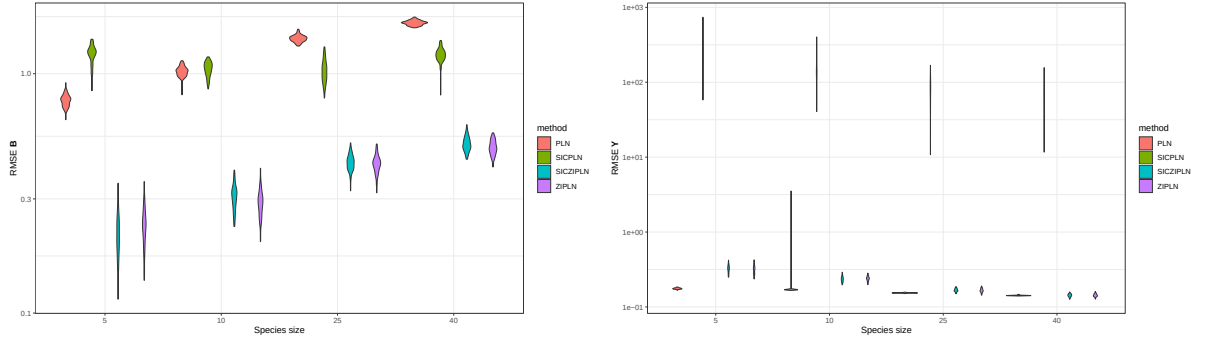


**Fig. 1:** Violin plots of the RMSE values over 100 datasets, comparing the true values (denoted by  $B$  and  $Y$ ) with their corresponding estimates for varying numbers of species,  $p \in \{5, 10, 25, 40\}$ . Left plot shows the violin plots of the RMSE for coefficients matrix  $B$ . Right plot presents the violin plots for the prediction errors RMSE

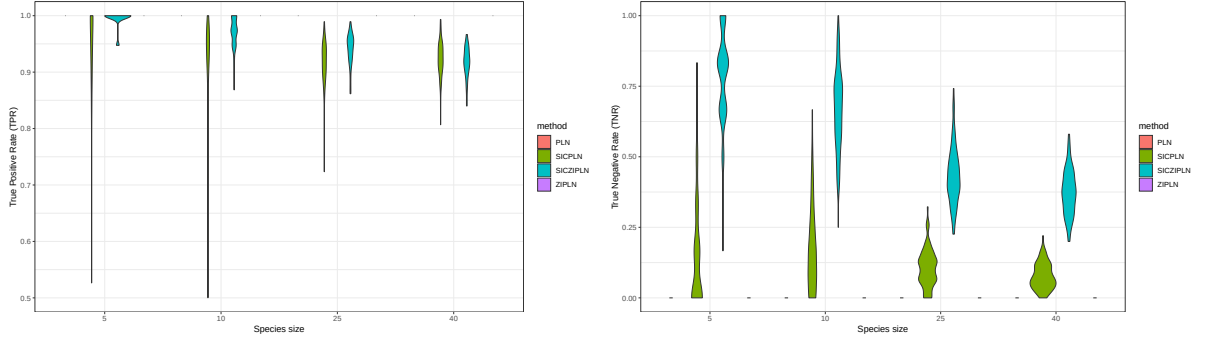


**Fig. 2:** Violin plots of the True Positive Rate (TPR) and True Negative Rate (TNR) over 100 datasets for varying numbers of species,  $p \in \{5, 10, 25, 40\}$  for PLN, SICPLN, SICZIPLN and ZIPLN. Left plot shows the violin plots for the TPR. Right plot presents the violin plots for the TNR

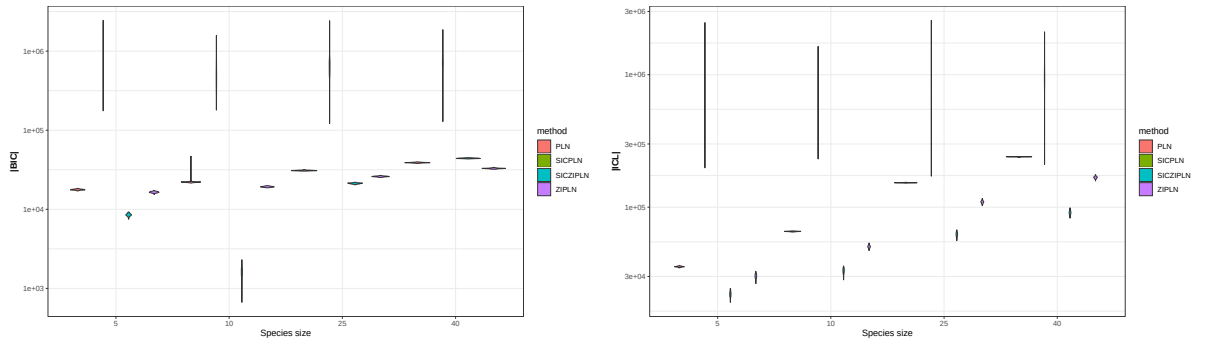
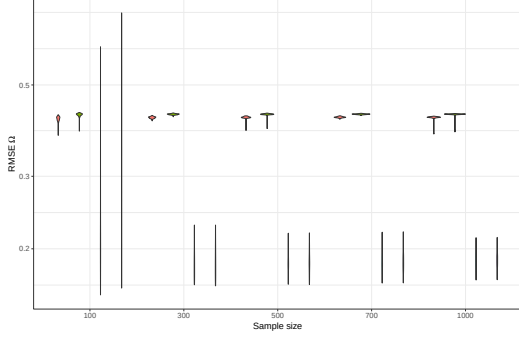


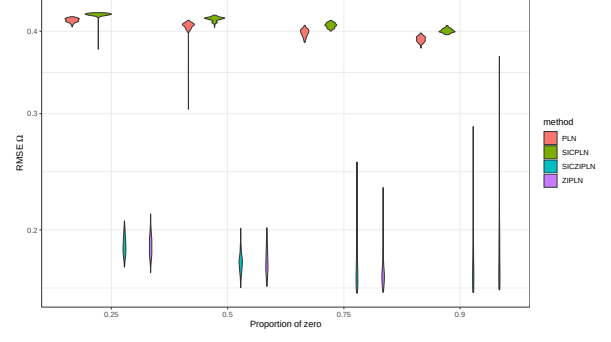
**Fig. 3:** Violin plots depicting the BIC and ICL over 100 datasets for varying numbers of species  $p \in \{5, 10, 25, 40\}$ . Left plot shows the violin plots of the BIC. Right plot presents the violin plots for the ICL

## 2 RMSE of the precision matrix $\Omega$

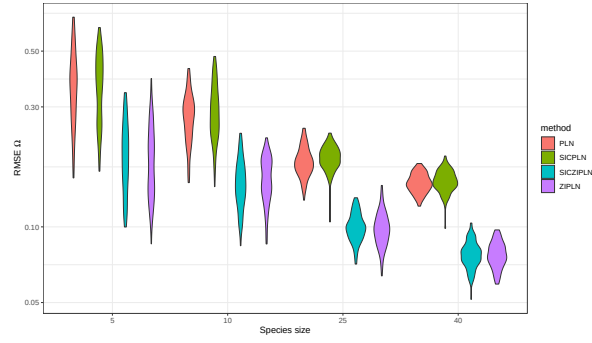
The figure 4 presents the RMSE of  $\Omega$  for each simulation experiment, with varying sample sizes, proportions of zeros in  $\mathbf{B}$ , and numbers of species.



(a) RMSE of  $\Omega$  with variation of sample size



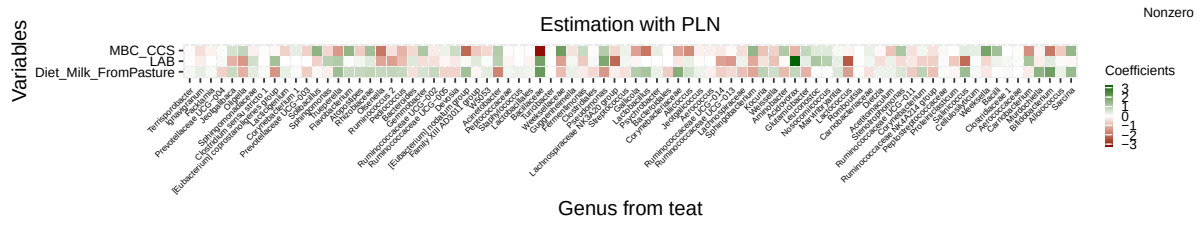
(b) RMSE of  $\Omega$  with variation of the proportion of zero in  $B$



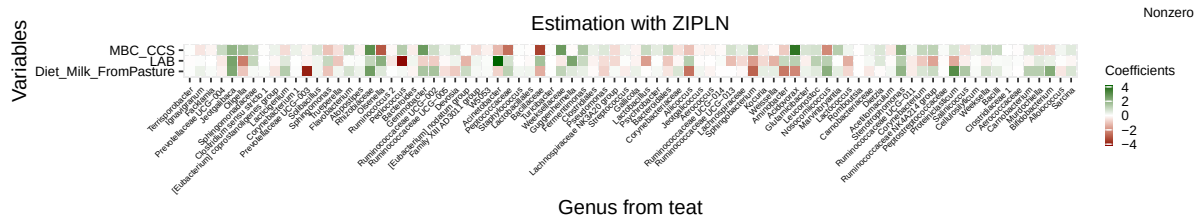
(c) RMSE of  $\Omega$  with variation of species size

**Fig. 4:** Violin plots depicting the RMSE values over 100 datasets, comparing the true values of  $\Omega$  with the corresponding estimates for varying sample sizes, proportions of zeros in  $B$ , and numbers of species

### 3 Applications: results for scenario 2

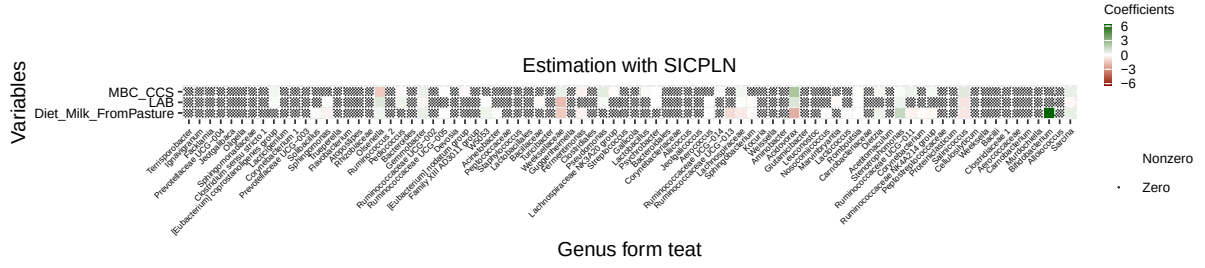


(a) Estimation with PLN

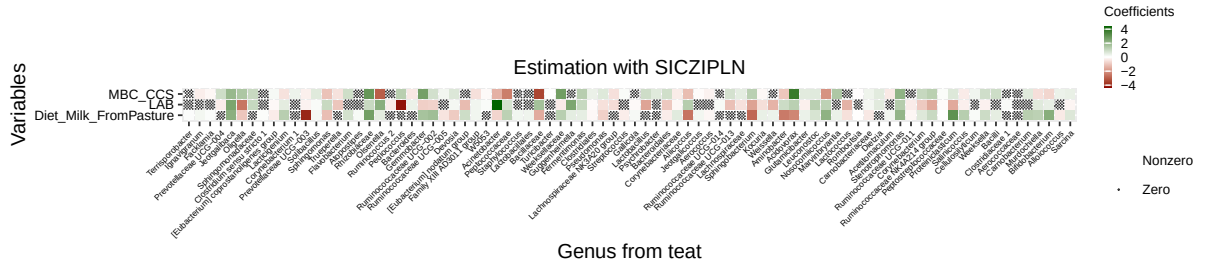


(b) Estimation with ZIPLN

**Fig. 5:** Scenario 2: coefficient estimates from unpenalized PLN and ZIPLN models on winter teat microbiota with 52% zeros in the count matrix



(a) Estimation with SICPLN



(b) Estimation with SICZIPLN

**Fig. 6:** Scenario 2: coefficient estimates from penalized PLN and ZIPLN models on winter teat microbiota with 52% zeros in the count matrix