Statistics and Its Interface Volume 13 (2020) 437–447

Bayesian meta-regression model using heavy-tailed random-effects with missing sample sizes for self-thinning meta-data

Zhihua Ma, Ming-Hui Chen*, and Yi Tang

Motivated by the self-thinning meta-data, a random-effects meta-analysis model with unknown precision parameters is proposed with a truncated Poisson regression model for missing sample sizes. The random effects are assumed to follow a heavy-tailed distribution to accommodate outlying aggregate values in the response variable. The logarithm of the pseudo-marginal likelihood (LPML) is used for model comparison. In addition, in order to determine which self-thinning law is more supported by the meta-data, a measure called "Plausibility Index (PI)" is developed. A simulation study is conducted to examine empirical performance of the proposed methodology. Finally, the proposed model and the PI measure are applied to analyze a self-thinning meta-data set in details.

Keywords and phrases: Outliers, Plausibility index, Self-thinning law, Truncated Poisson model.

1. INTRODUCTION

Self-thinning refers to the phenomenon that population density will decline with the increase in individual body size, which is very common in plant populations, from short shrubs to tall trees [10]. Ecologists have been seeking the law that governs population self-thinning for a long time [9]. Generally, the relationship between population density and individual biomass can be described using a power function, which is known as the self-thinning power law. Two kinds of power laws, $-3/2$ and $-4/3$ power laws, are proposed. Yoda et al. [24] found that in the fully crowded pure populations, the relationship between the average plant weight (W) and the population density (D) could be formulated by the equation $W = C \cdot D^{-3/2}$, named as the $-3/2$ power law of self-thinning. Enquist et al. [5] noted that the average plant size should be scaled as $-4/3$ power of maximum population density, based on the finding that rates of resource use in individual plants were scaled as approximately $3/4$ power of body mass [19]. Many researches, especially field investigations [7, 12, 20, 18], were conducted on determining the value of the power exponent. However, the value of the power exponent in the self-thinning law has not achieved a consensus yet.

In order to determine the most suitable self-thinning law, one effective approach is to conduct meta-analysis (MA) on the collected meta-data on self-thinning. MA analytically combines the results of individual studies to provide an overall estimate of the population treatment effects, which can lead to higher statistical power and more robust point estimates than single measures derived from individual studies. One of the research goals in this paper is to develop a new statistical methodology within the MA framework to determine which self-thinning law is more supported by the meta-data.

In the MA literature, it is common to assume that the variances of the random errors are known and set to be the squares of the SEs of the estimates from individual studies [21]. This assumption has been relaxed so that the variances are unknown and estimated using the information of the SEs along with sample sizes from individual studies [22, 23]. In some cases, such as the self-thinning data in our motivating example, the sample sizes may be missing. Under the models with known variances, missing sample sizes is not an issue. However, missing sample sizes cannot be ignored under the models with unknown variances since it will lead to a biased estimator if the missing data mechanism is not missing completely at random (MCAR). Researches have been done on missing data in MA, which primarily focus on dealing with missing outcomes, SEs, or covariates [8, 14, 15, 13]. However, the literature on missing sample sizes in analyzing aggregate MA data is still sparse. In this paper, we propose a truncated Poisson regression model for missing sample sizes.

In the MA literature, random effects in MA are traditionally assumed to be normally distributed, which is not necessarily appropriate in practice. The normality assumption for random effects may lead to inappropriate inferences under systematic departures from normality such as outliers or skewness [11]. Other distributions have been used for modeling random effects, including non-parametric distributions [1] and the t distribution [11]. In this paper, since there are outlying aggregate values in the self-thinning coefficient of the motivating meta-data, the random effects are assumed to follow a heavy-tailed distribution to flexibly deal with outliers.

*Corresponding author.

To carry out Bayesian inference, we develop an efficient Markov chain Monte Carlo (MCMC) sampling algorithm for sampling from the posterior distributions and a new Monte Carlo method for computing the Conditional Predictive Ordinate (CPO) for model comparison. Most importantly, in order to address the determination of self-thinning law supported by the meta-data, we propose a new index, called Plausibility Index (PI), to quantify the degree of plausibility of a self-thinning law. Furthermore, an efficient Monte Carlo method is developed to compute the PI.

The remainder of this paper is organized as follows. Section 2 gives a brief introduction of our motivating example. Section 3 presents the proposed heavy-tailed random-effects MA model as well as the missing sample size model. Bayesian inference procedure and a model selection criterion, LPML, is developed in Section 4. In Section 5, Plausibility Index, an index for measuring how a given value of the self-thinning coefficient is supported by the data, is proposed. A simulation study is conducted in Section 6. A detailed analysis of the self-thinning meta-data is carried out in Section 7. Finally we conclude the paper with a brief discussion in Section 8.

2. MOTIVATING EXAMPLE

In the plant community, the mass-density scaling relationship describing the variation in population density with body size can be written as

$$W = C \times D^{\mu}, \tag{1}$$

where C is a constant, and W and D represent the mean biomass of individuals and population density, respectively, both of which are positive real numbers. In Eq. (1), μ, a real number representing the exponent of the scaling relationship, is our study of interest. By applying a logarithmic transformation, Eq. (1) can be rewritten as $\log(W) = \log(C)+\mu\times\log(D)$, so the exponent self-thinning parameter μ becomes a regression coefficient or a slope, which is also called the "self-thinning coefficient" in the literature. The reduced major axis regression (RMA) is one of the most popular approaches for estimating the self-thinning coefficient μ, which minimizes the sum of the geometric means of the squared deviations from the fitted line in each dimension in the two-variable case [16]. RMA has some desirable properties such as handling measurement errors in both the dependent and independent variables and unit invariance in the corresponding fitted line.

In order to obtain a reliable estimate of the true underlying self-thinning coefficient, individual estimates from related studies are collected to form the meta-data. Journal articles published before April 2018 were searched using the Web of Science resource with terms "self-thinning" and "plants". The selection criteria include (i) field experiment data; (ii) using RMA for estimation; (iii) estimates of the slope in self-thinning line; (iv) the SEs or the 95% confidence intervals of the slope estimations; and (v) the sample size and the types of plant population. A total of $N = 100$ records were identified. Each record contains an estimate of the self-thinning coefficient, 95% confidence interval (CI) and/or standard error (SE), sample size, and plant type. Plant type is a categorical variable, which has three levels: forest, herb, and shrub. Some of the SEs, CIs, and sample sizes are missing in the meta-data. The missing SEs and CIs can be obtained by algebraically calculation from each other, but unfortunately, the missing sample sizes cannot be obtained in the same way. The complete meta-data with some remarks are given in Sections S.1 and S.2 of the Supplementary Materials http://www.intlpress.com/site/pub/files/_supp/sii/2020/0013/0004/SII-2020-0013-0004-s002.pdf.

For the self-thinning meta-data, there are a total of 100 records, and 4.00% of which have missing sample sizes. The average of the remaining sample sizes is 32 with a standard deviation of 45.09. Three plant types including forest, herb and shrub are involved. A summary of the self-thinning coefficients of different plant types is given in Table 1 and the violin-boxplots of the estimated self-thinning coefficients for these plant types, respectively, are shown in Figure 1. For the missing sample sizes, instead of omitting the missing records, a model for the missing sample sizes will be developed. In addition, Figure 1 shows that there are some outliers in the data, which motivates us to develop the meta-analysis model with heavy-tailed random-effects.

Table 1. Summary of the self-thinning coefficients by plant types

Plant types	Count	Self-thinning coefficients	
		Mean	(Minimum, Maximum)
Forest	66	-1.5230	$(-2.9152, -0.8696)$
Herb	28	-0.9835	$(-1.9401, 0.8212)$
Shrub	6	-1.1372	$(-1.3480, -0.8114)$

3. THE PROPOSED MODEL

Suppose that the meta-data are collected from a series of N independent studies with estimated self-thinning coefficients $Y_1, \ldots, Y_N$, standard errors $\mathrm{SE}_1, \cdots, \mathrm{SE}_N$ and sample sizes $n_1, \cdots, n_N$, with some sample sizes that are missing. According to the motivating example, we assume p plant types and each of them has the corresponding underlying self-thinning coefficient. Thus, we have $\boldsymbol{\mu} = (\mu_{\mathrm{type}_1}, \cdots, \mu_{\mathrm{type}_p})'$. To account for heterogeneity among studies, we assume the random-effects MA model for Y_i is given by

$$Y_i = \boldsymbol{z}_i'\boldsymbol{\mu} + \tau_i + \epsilon_i,$$
$$\tau_i \sim N(0, \psi^{-1}),\ \epsilon_i \sim N(0, \phi_i^{-1}),$$

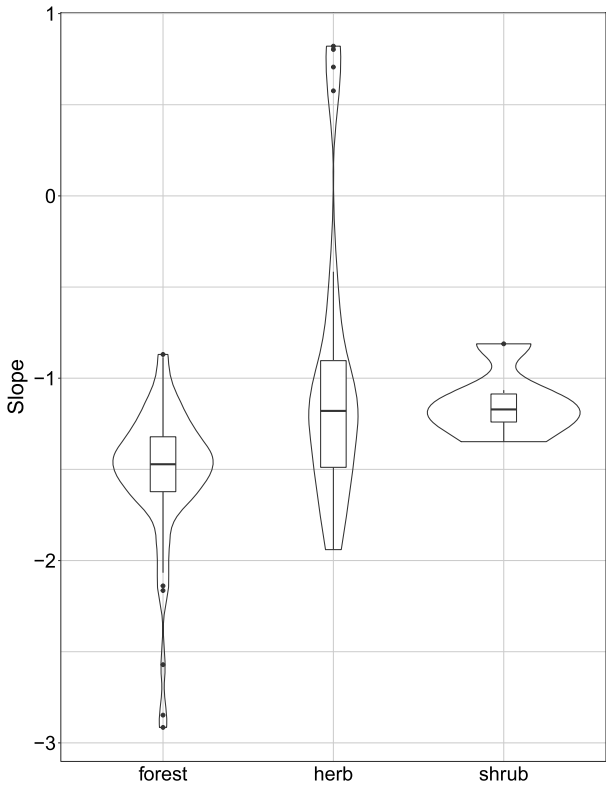


Figure 1. Violin-boxplot of the self-thinning meta-data by plant types.

where $\boldsymbol{z}_i$ denotes the vector indicating which plant type the ith study belongs to. For example, among the p plant types, the ith study belongs to type-1, then $\boldsymbol{z}_i = (1, 0, 0, \cdots, 0)$. ψ and ϕ_i are precisions of independent random variables τ_i and ϵ_i, respectively. The above model can be rewritten as

$$Y_i|\tau_i, \mu, \phi_i \sim N(\boldsymbol{z}_i'\boldsymbol{\mu} + \tau_i, \phi_i^{-1}), \tag{2}$$

where the random effects τ_i in the model capture heterogeneity among the studies. In the literature, ϕ_i^{-1} is assumed to be known and fixed to be the square of the corresponding standard error SE_i^2, see Whitehead [21] for example. In order to provide precise estimates especially when sample sizes are small, following Yao et al. [22] and Yao et al. [23], we take $\phi_i^{-1} = \sigma_i^2/n_i$ and further assume

$$n_i(n_i - 1)\mathrm{SE}_i^2 \sim \sigma_i^2\chi_{n_i-1}^2, \tag{3}$$

where $\chi_{n_i-1}^2$ denotes the χ^2-distribution with $(n_i - 1)$ degrees of freedom. In addition, according to Figure 1, a heavy-tailed distribution for τ_i is built to deal with outliers. We assume that

$$\tau_i|\lambda_i, \psi \sim N(0, (\lambda_i\psi)^{-1}),\ \lambda_i \sim \mathrm{Gamma}(\frac{v}{2}, \frac{2}{v}), \tag{4}$$

where v represents the degrees of freedom of the distribution, which is determined according to model selection criteria.

It should be noted that there are missing values in the sample sizes of the meta-data. One approach for dealing with this kind of data in many studies is to exclude the missing values and apply a complete case analysis (CC), which may lead to lower power. In this paper, we assume the missing data mechanism for the missing sample sizes is missing at random (MAR), and a missing sample size model is built as

$$n_i \sim \mathrm{Poisson}(\eta_i)1\{m_1, m_2\},\ \log(\eta_i) = \boldsymbol{X}_i'\boldsymbol{\beta}. \tag{5}$$

Here we assume the missing sample sizes follow a truncated Poisson distribution with a rate parameter η_i and lower and upper bounds m_1 and m_2. $1\{m_1, m_2\}$ is the indicator function, which equals to 1 when $n_i \in (m_1, m_2)$ and 0 otherwise. In (5), $\boldsymbol{X}_i$ is a p-dimensional aggregate covariate vector in the ith study and $\boldsymbol{\beta}$ is the vector of corresponding coefficients.

4. BAYESIAN INFERENCE AND MODEL SELECTION

4.1 Bayesian inference

In our model, the marginal density of random effects τ_i is given as

$$\begin{aligned} f(\tau_i|\psi) &= \int f(\tau_i, \lambda_i|\psi)d\lambda_i = \int f(\tau_i|\lambda_i, \psi)f(\lambda_i)d\lambda_i \\ &= \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})}\frac{\psi^{1/2}}{\sqrt{\pi v}}\left(1 + \frac{1}{v}\left(\frac{\tau_i}{\psi^{-1/2}}\right)^2\right)^{-\frac{v+1}{2}}, \end{aligned}$$

which is the density of a non-standardized Student's t-distribution $t_v(0, \psi^{-1})$ with v degrees of freedom and a scale parameter ψ.

Let $D = (N, \boldsymbol{Y}, \mathbf{SE}, \boldsymbol{n})$ denote the complete data, where $\boldsymbol{Y} = (Y_1, \cdots, Y_N)'$, $\mathbf{SE} = (\mathrm{SE}_1, \cdots, \mathrm{SE}_N)'$, $\boldsymbol{z} = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N)'$ and $\boldsymbol{n} = (n_1, \cdots, n_N)'$. The complete data likelihood function is given by

$$\begin{aligned} &L(\boldsymbol{\mu}, \psi, \boldsymbol{\sigma}, \boldsymbol{\beta}|D) = \\ &\prod_{i=1}^{N} f(Y_i|\boldsymbol{\mu}, \tau_i, \sigma_i, n_i)f(\mathrm{SE}_i|\sigma_i)f(\tau_i|\psi)f(n_i|\boldsymbol{\beta}), \end{aligned}$$

where $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_N)'$. Our main interest lies in the posterior inferences on $\boldsymbol{\mu}$ and ψ based on the observed data. Write $\boldsymbol{n} = (\boldsymbol{n}_{mis}, \boldsymbol{n}_{obs})$, where $\boldsymbol{n}_{mis}$ and $\boldsymbol{n}_{obs}$ denote the missing and observed components of sample sizes $\boldsymbol{n}$, respectively. Let $D_{obs} = (N, \boldsymbol{Y}, \mathbf{SE}, \boldsymbol{z}, \boldsymbol{n}_{obs})$ denote the observed data. We assume the missing data mechanism of $\boldsymbol{n}_{mis}$ is ignorable, then the likelihood function based on the observed data is given by

$$\begin{aligned} &L_o(\boldsymbol{\mu}, \psi, \boldsymbol{\sigma}, \boldsymbol{\beta}|D_{obs}) = \\ &\prod_{i=1}^{N}\sum_{\boldsymbol{n}_{mis}}\Big[\int f(Y_i|\boldsymbol{\mu}, \tau_i, \sigma_i, \boldsymbol{n}_{mis})f(\mathrm{SE}_i|\sigma_i)f(\tau_i|\psi)d\tau_i\Big] \\ &\times f(\boldsymbol{n}_{mis}|\boldsymbol{\beta}). \end{aligned}$$

Thus, the joint posterior density of the unknown parameters is given by

$$\pi(\boldsymbol{\mu},\psi,\boldsymbol{\sigma},\boldsymbol{\beta}|D_{obs}) \propto L_o(\boldsymbol{\mu},\psi,\boldsymbol{\sigma},\boldsymbol{\beta}|D_{obs})\pi(\boldsymbol{\mu},\psi,\boldsymbol{\sigma},\boldsymbol{\beta}),$$

where $\pi(\boldsymbol{\mu},\psi,\boldsymbol{\sigma},\boldsymbol{\beta})$ denotes the joint prior distribution of the unknown parameters. Here we assume that the prior distributions of these parameters are independent a priori, and further specify the following prior distributions for these parameters: $\boldsymbol{\mu} \sim \text{MVN}(\mathbf{0},\Sigma_\mu)$ with $\Sigma_\mu = \text{diag}(\psi_\mu^{-1},\ldots,\psi_\mu^{-1})$, $\pi(\psi) \propto \psi^{a_0-1}\exp(-b_0\psi)$, $\pi(\sigma_i^2) \propto \exp(-b_1/\sigma^2)/(\sigma^2)^{a_1+1}$, and $\beta_\ell \sim N(0,\psi_{\beta_\ell}^{-1}), \ell = 1,\cdots,p$. Note that ψ_μ, a_0, b_0, a_1, b_1, and ψ_{β_ℓ} are pre-specified hyperparameters. In this paper, we use $\psi_\mu = \psi_{\beta_\ell} = 0.001$, $a_0 = b_0 = 1$, and $a_1 = b_1 = 0.001$.

Markov chain Monte Carlo (MCMC) sampling algorithms are used to sample from the posterior distributions. It requires sampling the following parameters in turn from their respective full conditional distributions: (i) $\pi(\boldsymbol{n}_{mis}|\boldsymbol{\mu},\boldsymbol{\sigma},\psi,\boldsymbol{\beta},D_{obs})$, (ii) $\pi(\boldsymbol{\mu}|\boldsymbol{n}_{mis},\boldsymbol{\sigma},\psi,\boldsymbol{\beta},D_{obs})$, (iii) $\pi(\boldsymbol{\sigma}|\boldsymbol{\mu},\boldsymbol{n}_{mis},\psi,\ \boldsymbol{\beta},D_{obs})$, and (iv) $\pi(\boldsymbol{\beta}|\boldsymbol{\mu},\boldsymbol{n}_{mis},\boldsymbol{\sigma},\psi,D_{obs})$. Take $\boldsymbol{\mu}$ as an example. The full conditional distribution of $\boldsymbol{\mu}$ can be written as

$$\begin{aligned}\pi(\boldsymbol{\mu}|\boldsymbol{n},\boldsymbol{\sigma},\psi,\boldsymbol{\beta},D_{obs}) &\propto \prod_{i=1}^{N} p(Y_i|\boldsymbol{\mu},\tau_i,n_i,\sigma_i)\pi(\boldsymbol{\mu}) \\ &\propto \prod_{i=1}^{N}\exp\Big\{-\frac{(Y_i-\boldsymbol{z}_i'\boldsymbol{\mu}-\tau_i)^2}{2\sigma_i^2/n_i}\Big\}\exp\Big\{-\frac{1}{2}\boldsymbol{\mu}'\Sigma_\mu^{-1}\boldsymbol{\mu}\Big\} \\ &= \exp\Big\{-\sum_{i=1}^{N}\frac{(Y_i-\boldsymbol{z}_i'\boldsymbol{\mu}-\tau_i)^2}{2\sigma_i^2/n_i}-\frac{1}{2}\boldsymbol{\mu}'\Sigma_\mu^{-1}\boldsymbol{\mu}\Big\},\end{aligned}$$

which implies

$$\begin{aligned}(6)\quad \boldsymbol{\mu}|\boldsymbol{n},\boldsymbol{\sigma},\psi,\boldsymbol{\beta},D_{obs} \sim \\ \text{MVN}\big(\big(\sum_{i=1}^{N} n_i/\sigma_i^2+\Sigma_\mu^{-1}\big)^{-1}\sum_{i=1}^{N}\boldsymbol{z}_i'(Y_i-\tau_i)n_i/\sigma_i^2, \\ \big(\sum_{i=1}^{N} n_i/\sigma_i^2+\Sigma_\mu^{-1}\big)^{-1}\big).\end{aligned}$$

Thus, sampling $\boldsymbol{\mu}$ from $\pi(\boldsymbol{\mu}|\boldsymbol{n}_{mis},\boldsymbol{\sigma},\psi,\boldsymbol{\beta},D_{obs})$ is straightforward.

4.2 Model selection

In order to determine the degrees of freedom v in our model, we employ the Logarithm of the Pseudo-Marginal Likelihood (LPML) for model selection. LPML can be calculated via Conditional Predictive Ordinates (CPOs).

Let $D_{obs}^{(-i)} = \{Y_j : j = 1,\ldots,i-1,i+1,\ldots,N\}$ denote the observed data with the ith study deleted. The CPO for the ith study is defined as

$$(7)\quad \text{CPO}_i = \int f(Y_i|\boldsymbol{\mu},\psi,\sigma_i)\pi(\boldsymbol{\mu},\psi,\sigma_i|D_{obs}^{(-i)})d(\boldsymbol{\mu},\psi,\sigma_i),$$

where $f(Y_i|\boldsymbol{\mu},\psi,\sigma_i) = \int f(Y_i,\tau_i|\boldsymbol{\mu},\psi,\sigma_i)d\tau_i$ denotes the marginal distribution of Y_i, $f(Y_i,\tau_i|\boldsymbol{\mu},\psi,\sigma_i) = f(Y_i|\boldsymbol{\mu},\tau_i,\sigma_i)f(\tau_i|\psi)$ is the joint distribution of (Y_i,τ_i),

$$\pi(\boldsymbol{\mu}|D_{obs}^{(-i)}) = \frac{\prod_{j\neq i} f(Y_j|\boldsymbol{\mu},\psi,\sigma_j)\pi(\boldsymbol{\mu},\psi,\sigma_j)}{c(D_{obs}^{(-i)})},$$

and $c(D_{obs}^{(-i)})$ is the normalizing constant.

The calculation of CPO_i in (7) involves an integral over the random effect τ_i, which is computationally extensive. In order to circumvent this numerical integration issue, we employ the new CPO identity proposed by Zhang et al. [25]. Letting $w_i(\tau_i)$ be a normalized weight function such that $\int w_i(\tau_i)d\tau_i = 1$, we have

$$\begin{aligned}&\text{CPO}_i^{-1} = \\ &\int \frac{w_i(\tau_i)}{f(Y_i,\tau_i|\boldsymbol{\mu},\psi,\sigma_i)}\pi(\boldsymbol{\mu},\psi,\sigma_i,\tau_i|D_{obs})d\tau_i d(\boldsymbol{\mu},\psi,\sigma_i).\end{aligned}$$

Let $\{(\boldsymbol{\mu}_t,\psi_t,\sigma_{it},\tau_{it}) : i = 1,\ldots,n; t = 1,\ldots,T\}$ denote a Gibbs sample of $(\boldsymbol{\mu},\psi,\sigma_i,\tau_i)$ from $\pi(\boldsymbol{\mu},\psi,\sigma_i,\tau_i|D_{obs})$. Using the above new CPO identity, a Monte Carlo estimate of CPO_i^{-1} can be given as

$$\widehat{\text{CPO}}_i^{-1} = \frac{1}{T}\sum_{t=1}^{T}\frac{w_i(\tau_{it})}{f(Y_i,\tau_{it}|\boldsymbol{\mu}_t,\psi_t,\sigma_{it})},$$

where $f(Y_i,\tau_{it}|\boldsymbol{\mu}_t,\psi_t,\sigma_{it}) = f(Y_i|\tau_{it},\boldsymbol{\mu}_t,\sigma_{it})\times\ f(\tau_{it}|\psi_t)$, and $w_i(\tau_i)$ can be constructed as a normal distribution via the Laplace approximation to the conditional distribution of τ_i with the density proportional to $f(Y_i,\tau_i|\boldsymbol{\mu},\psi,\sigma_i)$. The approximated normal distribution of $w_i(\tau_i)$ can be written as

$$\begin{aligned}w_i(\tau_i) &\sim N(\hat{\tau}_i,\text{H}^{-1}), \\ \text{H} &= -\frac{\partial^2 \log f(Y_i,\tau_i|\boldsymbol{\mu},\psi,\sigma_i)}{\partial\tau_i^2}\Big|_{\tau_i=\hat{\tau}_i},\end{aligned}$$

where $\hat{\tau}_i$ is the mode of τ_i. Then LPML based on the new CPO identity can be obtained as

$$(8)\qquad \widehat{\text{LPML}} = \sum_{i=1}^{n}\log(\widehat{\text{CPO}}_i).$$

A greater value of LPML represents a better model.

5. PLAUSIBILITY INDEX

In order to determine which self-thinning law is preferred, in this section we propose an index to quantify the degree of plausibility of a given value of the self-thinning coefficient for a given dataset, which is called "Plausibility Index (PI)". For simplicity, we assume that there is only one plant type. Thus, the vector $\boldsymbol{\mu}$ can be simplified as a scalar quantity μ.

Suppose that a given value of μ is denoted by μ_0, and let $\pi_0 = \pi(\mu = \mu_0|D_{obs})$. The PI measure is defined as

$$\mathrm{PI}(\mu = \mu_0) = \int_R \int_0^{\pi(\mu|D_{obs})\wedge\pi_0} d\delta d\mu.$$

Visually, the PI measure is the area under the marginal posterior density curve $\delta = \pi(\mu|D_{obs})$ and below the horizontal line $\delta = \pi_0$, which is represented by the colored area in Figure 2. A larger value of the PI measure means a greater plausibility that the given value is supported by the data. See Section S.4 of the Supplementary Materials for a connection between the PI and the classical p-value under a very simple setting.

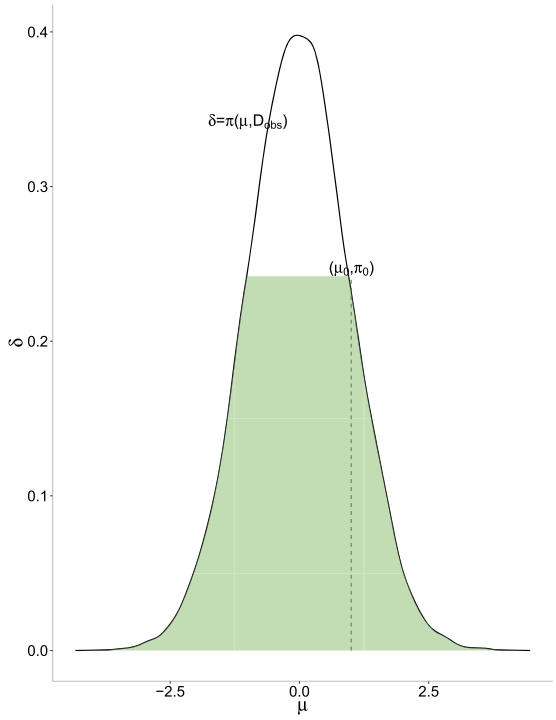


Figure 2. Visual plot of the PI measure.

Usually, the marginal posterior density of a parameter does not have a closed form and is difficult to be calculated directly. Here we use the conditional marginal density estimator (CMDE) proposed by Gelfand et al. [6] to estimate the marginal posterior density. Assuming that the analytical form of the conditional posterior density $\pi(\mu|\sigma_i, \psi, D_{obs})$ is available, the marginal posterior density can be written as

$$\pi(\mu_0|D_{obs}) = \int \pi(\mu_0|\psi, \boldsymbol{\sigma}, D_{obs})\pi(\mu, \psi, \boldsymbol{\sigma}|D_{obs})d(\mu, \psi, \boldsymbol{\sigma}).$$

Then CMDE can be obtained as

$$\hat{\pi}_{\mathrm{CMDE}}(\mu_0|D_{obs}) = \frac{1}{T}\sum_{t=1}^{T} \pi(\mu_0|\psi_t, \boldsymbol{\sigma}_t, D_{obs}).$$

It can be shown that under some mild regularity conditions, $\hat{\pi}_{\mathrm{CMDE}}(\mu_0|D_{obs})$ is an unbiased and consistent estimator of the marginal posterior density.

In our case, the conditional posterior density $\pi(\mu|\psi, \boldsymbol{\sigma}, D_{obs})$ is given by (6), so we can obtain the CMDE of μ as discussed above. With CMDE as an estimate of the marginal posterior density, the PI can also be computed as

$$\mathrm{PI}(\mu_0) = \frac{1}{T^*}\sum_{t^*=1}^{T^*}\left[1\{\pi(\mu_{t^*}) \le \pi(\mu_0)\} + \frac{\pi(\mu_0)}{\pi(\mu_{t^*})}1\{\pi(\mu_{t^*}) > \pi(\mu_0)\}\right],$$

where $1\{\pi(\mu_{t^*}) > \pi(\mu_0)\}$ is the indicator function, which equals to 1 when $\pi(\mu_{t^*}) > \pi(\mu_0)$ and 0 otherwise.

6. A SIMULATION STUDY

In this section, a simulation study is conducted to examine the empirical performance of LPML on model selection, as well as the comparison between our proposed model and some existing models under different simulation settings.

In order to examine the performance of LPML, we simulated a dataset with $N = 100$ individual records. For $i = 1, \cdots, N$, the response variable Y_i was randomly generated according to (2) with $\mu = -4/3$, σ_i^2 was generated from Uniform$(0, 1)$, the sample size n_i was generated according to (5) with $m_1 = 6, m_2 = 1000$, and the rate parameter η_i was generated according to $\log(\eta_i) = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i}$, where X_{1i} and X_{2i} were independently generated from the standard normal distribution, $\beta_1 = 2$, $\beta_2 = -2$ and $\beta_3 = -1$. We generated SEs according to (3).

Missing data for the sample size n_i were generated with an ignorable missing data mechanism. Specifically, let $R_i = 1$ if n_i is observed and $R_i = 0$ if n_i is missing. A logistic regression model was built for the missing indicator R_i as

$$f(R_i|X_{1i}, X_{2i}, \boldsymbol{\phi}) = \frac{\exp(\phi_1 + \phi_2 X_{1i} + \phi_3 X_{2i})}{1 + \exp(\phi_1 + \phi_2 X_{1i} + \phi_3 X_{2i})},$$

where $\phi_1 = 0$, $\phi_2 = -1$ and $\phi_3 = -3$. The average percentage of missing sample sizes in $T = 100$ simulations in this study is about 20%.

The true distribution of τ_i is a non-standardized Student's t-distribution $t_v(0, \psi^{-1})$, where $\psi = 1$ and $v = 2$. We built our proposed model (2)–(5) with different values of v, and applied LPML to conduct model selection. With the thinning interval of 5, 8000 samples are kept for posterior calculation after a burn-in of 2000 samples. Average LPML results of different values of v are shown in Table 2. Boxplots of the differences of LPML values between the selected model and the other models are also shown in Figure 3.

From Table 2 and Figure 3 we can see that when τ_i indeed follows a non-standardized Student's t-distribution

Table 2. *Average LPML values for different values of v*

	$v=2$	$v=4$	$v=6$
LPML	−194.95	−198.51	−201.68
	$v=8$	$v=10$	$v=\infty$
LPML	−203.94	−205.17	−211.53

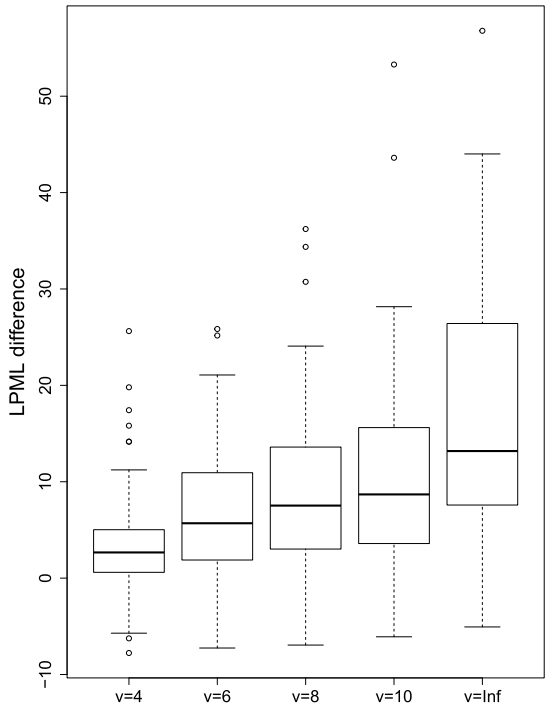


Figure 3. LPML differences with different values of v.

with $v=2$, the LPML value of a model with $v=2$ is the largest, and decreases as v increases. These results show that LPML performs well in selecting the best model under our settings.

The average PI values of $\mu=-4/3$ and $\mu=-3/2$ are also calculated, which turn out to be 0.99 and 0.49, respectively. The corresponding boxplots are shown in Figure 4. This result shows that the PI index can help quantify the degree of plausibility of a given value and choose the one that is more supported by the data.

For assessing the precision of the posterior estimates, we employ the following assessment measures. Take parameter μ as an example: Bias $=\frac{1}{T}\sum_{t=1}^{T}(\hat{\mu}_t-\mu^0)$, SD $=\frac{1}{T}\sum_{t=1}^{T}\mathrm{sd}(\hat{\mu}_t)$, MSE $=\frac{1}{T}\sum_{t=1}^{T}(\hat{\mu}_t-\mu^0)^2$, and CP $=\frac{1}{T}\sum_{t=1}^{T}1(\mu_0\in\mathrm{CI}(\hat{\mu}_t))$, where $\hat{\mu}_t$ denotes the posterior estimation of μ in the tth iteration, μ^0 denotes the true value of parameter μ, $\mathrm{sd}(\hat{\mu}_t)$ is the posterior standard deviation of the estimate, and $\mathrm{CI}(\hat{\mu}_t)$ is the estimated 95% credible interval of $\hat{\mu}$ in the tth iteration.

For the simulated datasets, we fit the data using three different models: the proposed model with $v=2$ and the normal random effects models with known and unknown

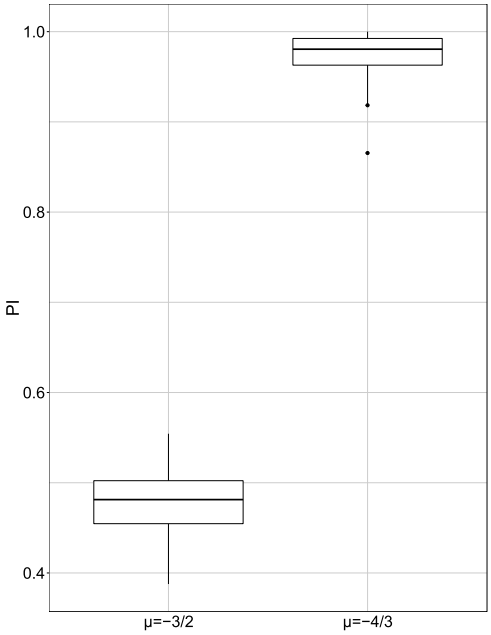


Figure 4. Boxplots of PI for the different values of μ in the simulation study.

variances. We also carried out the CC analysis. The simulation results under these three models and the CC analysis are shown in Table 3.

When the true distribution of τ_i is actually a non-standardized Student's t-distribution with $v=2$, from Table 3 we can see that (i) estimates in the proposed model have higher precision than the other alternative models; (ii) the posterior SDs of parameter ψ under the models with unknown variance are larger than those under the model with known variance; (iii) simulation results of the normal random-effects models with known and unknown variances are similar except for larger SDs under the model with unknown variance; and (iv) estimation biases under the CC analysis are larger compared to those under the proposed model due to a MAR missing data mechanism.

We also consider the situations with different values of the precision parameter of the random effects. When the true values of ψ are 0.1, 1 and 5, the simulation results under the proposed model and the normal random effects models with known and unknown parameters are shown in Table 4.

The results in Table 4 indicate that (i) generally, the posterior estimation biases under the proposed model are smaller than those under the other two alternative models, and the CPs of the parameters under the proposed model are both greater than 0.92 while the CPs of ψ under the normal random effects models are 0; (ii) as ψ gets larger, the bias of ψ becomes larger, and the posterior SD of the parameter μ becomes smaller while the posterior SD of ψ gets larger for all of the three models; (iii) under the proposed model, the posterior SD of μ is smaller than those

Table 3. Simulation results under different models: True model with $\tau_i \sim t_2(0, 1^{-1})$

	True value	The proposed model with $v = 2$				Normal random-effects model with unknown variance			
		Bias	SD	MSE	CP	Bias	SD	MSE	CP
μ	$-4/3$	-0.0012	0.1285	0.0198	0.95	0.0139	0.2288	0.0565	0.94
ψ	1	0.0725	0.3002	0.0573	0.96	-0.7843	0.0432	0.5923	0.00
β_1	2	-0.0302	0.0306	0.0024	0.93	-0.0297	0.0363	0.0067	0.92
β_2	-2	0.0172	0.0269	0.0014	0.92	0.0180	0.0270	0.0018	0.92
β_3	-1	0.0295	0.0319	0.0027	0.93	-0.0308	0.0275	0.0024	0.92
	True value	Normal random-effects model with known variance				CC analysis			
		Bias	SD	MSE	CP	Bias	SD	MSE	CP
μ	$-4/3$	-0.0131	0.2137	0.0471	0.93	0.0073	0.1475	0.0239	0.93
ψ	1	-0.8053	0.0279	0.6590	0.00	0.0803	0.2861	0.0828	0.94

Table 4. Simulation results of the proposed model and models with known and unknown variances under different true values of ψ

		The proposed model with $v = 2$		Normal random effects model with unknown variance		Normal random effects model with known variance	
		μ	ψ	μ	ψ	μ	ψ
$\psi = 0.1$	Bias	0.0079	0.0125	-0.0392	-0.0783	-0.0431	-0.0796
	SD	0.3861	0.0293	0.7366	0.0033	0.7112	0.0029
	MSE	0.1003	0.0010	0.6874	0.0061	0.9126	0.0068
	CP	0.93	0.94	0.90	0.00	0.88	0.00
$\psi = 1$	Bias	-0.0012	0.0725	0.0139	-0.7843	-0.0131	-0.8053
	SD	0.1285	0.3002	0.2288	0.0432	0.2137	0.0279
	MSE	0.0198	0.0573	0.0565	0.5923	0.0471	0.6590
	CP	0.95	0.96	0.94	0.00	0.93	0.00
$\psi = 5$	Bias	0.0017	-0.5933	-0.0121	-2.9532	-0.0119	-3.0182
	SD	0.0719	0.9110	0.1145	0.2273	0.1185	0.2140
	MSE	0.0041	0.9429	0.0113	12.7892	0.0122	13.0860
	CP	0.96	0.97	0.98	0.00	0.96	0.00

under each of the other two models, while the posterior SD of ψ is the largest among these models; and (iv) the estimation results of the two normal random effects models are similar, but the posterior SDs of the parameters under the model with unknown variance are larger than those under the model with known variance.

In order to demonstrate the impact of different percentages of missing sample sizes on the performance of the proposed model and the CC analysis, we simulated data with 20%, 30% and 40% missing sample sizes with the other simulation settings unchanged. The simulation results are shown in Table 5.

From Table 5, we can see that as the missing percentage increases, the biases and MSEs of the estimates under the CC analysis increase. The differences between the estimates under the proposed model and the CC analysis also become larger with a larger missing percentage. The results in Table 5 confirm that the CC analysis would yield biased estimates under MAR missing data settings, and the bias becomes larger with a higher missing percentage.

Table 5. Simulation results under the proposed model and the CC analysis with different missing percentages

		The proposed model		CC analysis	
		μ	ψ	μ	ψ
20% missing	Bias	-0.0012	0.0725	0.0073	0.0803
	SD	0.1285	0.3002	0.1475	0.2861
	MSE	0.0198	0.0573	0.0239	0.0828
	CP	0.95	0.96	0.93	0.94
30% missing	Bias	-0.0037	0.0807	-0.0150	0.1086
	SD	0.1325	0.2758	0.1752	0.3032
	MSE	0.0203	0.0510	0.0360	0.0902
	CP	0.95	0.94	0.92	0.93
40% missing	Bias	-0.0085	0.1023	0.0235	0.1203
	SD	0.1331	0.2903	0.1742	0.3303
	MSE	0.0156	0.0603	0.0437	0.1161
	CP	0.94	0.96	0.90	0.93

Table 6. LPML values for selecting v in the distribution of the random effects

v	1	2	3	4	5
LPML	−61.84	**-54.64**	−55.60	−57.01	−60.28
v	6	7	8	9	10
LPML	−61.45	−61.32	−62.16	−65.43	−65.76
v	20	100	∞		
LPML	−68.26	−71.24	−71.98		

7. EMPIRICAL ANALYSIS

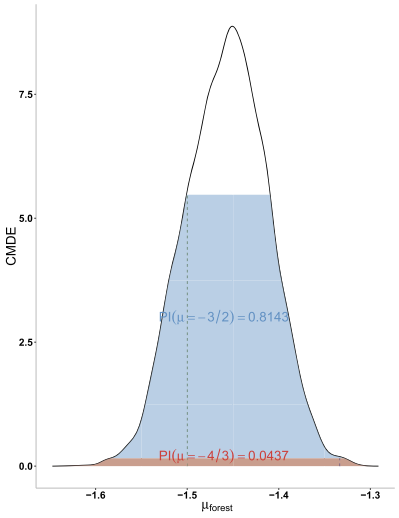
In this section, the proposed methodology is applied to analyze the self-thinning meta-data introduced in Section 2. For the missing sample sizes model (5), we set $m_1 = 3$ and $m_2 = 1000$. Plant type is used as the covariate, and the rate parameter η_i is built as $\log(\eta_i) = \beta_0 + \mathrm{I}(\text{herb})\beta_1 + \mathrm{I}(\text{shrub})\beta_2$. We fit the self-thinning meta-data using the heavy-tailed random effects model with unknown variance (2)–(4). In order to see whether the self-thinning coefficients would vary among different plant types, the model for Y_i is $Y_i|\boldsymbol{\mu}, \psi, \phi_i \sim N(\boldsymbol{z}_i'\boldsymbol{\mu}, \psi^{-1} + \phi_i^{-1})$ with $\boldsymbol{\mu} = (\mu_{\text{forest}}, \mu_{\text{herb}}, \mu_{\text{shrub}})$. The priors for the parameters are specified in Section 4.1. The MCMC sampling algorithms are implemented through an R package nimble [4] and the nimble code can be found in Section S.3 of the Supplementary Materials. With a thinning interval of 5, 8000 samples are kept for calculating the posterior estimates after a burn-in of 2000 samples. The convergence of the MCMC sampling algorithm is checked using several diagnostic procedure discussed in Cowles and Carlin [3] and Chen et al. [2].

The model selection criterion defined in (8) is applied to determine the value of v in the non-standardized Student's t-distribution of the random effects. In Table 6, the values of LPML corresponding to different values of v are shown. According to the results in Table 6, since the LPML value under $v = 2$ is the largest among these models, the heavy-tailed random effects model with the non-standardized t_2-distribution is selected as the best model for this meta-data set. These results are empirically appealing since a heavy-tailed non-standardized t_2-distribution can better accommodate the outliers shown in Figure 1 than those lighter tailed distributions including the normal distribution.

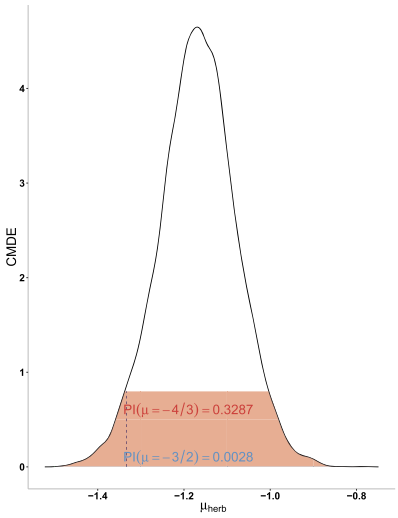
The posterior estimates of the parameters μ and ψ under the chosen model are shown in Table 7. In the same table, we also show the posterior estimates under the normal random effects models with known and unknown variance, and the CC analysis.

Table 7. Posterior estimates under the proposed model, the normal random effects models with known and unknown variances, and the CC analysis

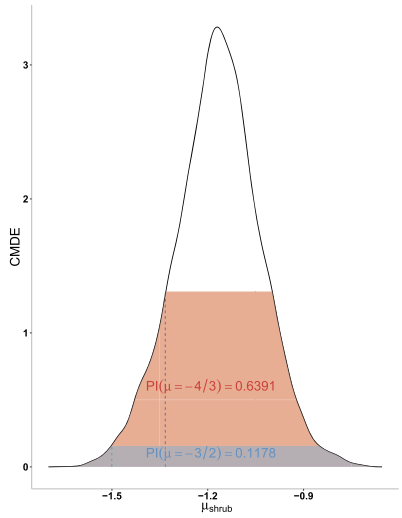
		t-distributed random effects with unknown variance	Normal random effects with unknown variance
μ_{forest}	Estimate	−1.4567	−1.4970
	SD	0.0444	0.0645
	95% HPD interval	(−1.5414, −1.3725)	(−1.6274, −1.3716)
μ_{herb}	Estimate	−1.1678	−0.9188
	SD	0.0885	0.1008
	95% HPD interval	(−1.3465, −0.9935)	(−1.1134, −0.7208)
μ_{shrub}	Estimate	−1.1677	−1.1358
	SD	0.1311	0.2053
	95% HPD interval	(−1.4307, −0.9104)	(−1.5251, −0.7551)
ψ	Estimate	12.4525	4.4971
	SD	2.5166	0.8179
	95% HPD interval	(8.1507, 17.8941)	(3.1116, 6.1799)
		Normal random effects with known variance	CC analysis
μ_{forest}	Estimate	−1.5015	−1.4612
	SD	0.0591	0.0463
	95% HPD interval	(−1.6170, −1.3869)	(−1.5536, −1.3741)
μ_{herb}	Estimate	−0.9182	−1.1690
	SD	0.0958	0.0889
	95% HPD interval	(−1.1091, −0.7310)	(−1.3413, −0.9910)
μ_{shrub}	Estimate	−1.1525	−1.1673
	SD	0.1992	0.1256
	95% HPD interval	(−1.5468, −0.7409)	(−1.4050, −0.9052)
ψ	Estimate	4.8461	11.7985
	SD	0.7578	2.3804
	95% HPD interval	(3.4243, 6.6061)	(7.7116, 16.9585)



(a) PI of μ_{forest}



(b) PI of μ_{herb}



(c) PI of μ_{shrub}

Figure 5. PI of μ for different plant types under the proposed model.

From Table 7, we can see that (i) under our proposed model, the posterior estimates of the underlying self-thinning coefficients vary among different plant types; (ii) estimates under the proposed model are different from those under the normal random effects models, especially for the precision parameter of the random effects; (iii) the posterior SDs of the estimates under the normal random effects model with unknown variance are larger than those under the model with known variance; and (iii) since the percentage of missing sample size is quite small, the results of the self-thinning coefficients under the CC analysis are not much different than those under our proposed model.

Apart from obtaining estimates of the underlying self-thinning coefficients for different plant types, we also want to explore which self-thinning law is more supported by the meta-data. The PI introduced in Section 5 is employed for determination by calculating the PI values under $\mu_0 = -4/3$ and $\mu_0 = -3/2$ for each plant types. We first calculate $\hat{\pi}_{\text{CMDE}}(-4/3|D_{obs})$ and $\hat{\pi}_{\text{CMDE}}(-3/2|D_{obs})$ as the respective estimates of the marginal posterior density of μ, and then obtain the PI measures according to (6). The PI values of $\mu_0 = -4/3$ and $\mu_0 = -3/2$ under our proposed model for different plant types are shown in Table 8. And visually, the PI values under our proposed model are shown in Figure 5.

From Table 8 and Figure 5, we can see that for forest, the PI values of $\mu_0 = -3/2$ are much larger than those of $\mu_0 = -4/3$, indicating that the meta-data support more on the "$-3/2$ self-thinning law" than the "$-4/3$ self-thinning law" for forest plants. However, for plant types herb and shrub,

Table 8. PI values of different plant types under the proposed model

	Forest		Herb		Shrub	
	$\mu_0 = -3/2$	$\mu_0 = -4/3$	$\mu_0 = -3/2$	$\mu_0 = -4/3$	$\mu_0 = -3/2$	$\mu_0 = -4/3$
PI value	0.8143	0.0437	0.0028	0.3287	0.1178	0.6391

the meta-data support more on the "$-4/3$ self-thinning law" instead since the degree of plausibility of the "$-4/3$ self-thinning law" is much higher than the other for herb and shrub in this case.

8. DISCUSSION

Motivated by the self-thinning meta-data with missing sample sizes and outliers, in this paper we develop a heavy-tailed random-effects model with a truncated Poisson missing sample size model to obtain estimates of the self-thinning coefficient. A heavy-tailed distribution is assumed for the random effects to account for outliers, and the variance of the response variable is set to be unknown. Additionally, an index measuring how a given value of the self-thinning coefficient is supported by the data is proposed, so we can employ the index to determine the degree of plausibility of the self-thinning laws and figure out which self-thinning law is more supported by the meta-data.

Some extensions can be considered for this study. In this paper, a heavy-tailed distribution is assigned for the random effects, which can be extended to skewed distributions or nonparametric distributions. Besides, the estimation method in this paper is restricted to RMA, which can be extended to multiple methods and take into account the impact of different methods. In addition, it is necessary to explore the properties of the PI measure and try to explore hypothesis testings using this measure to decide whether the differences among different PI values are significant or not. Furthermore, in this paper, the missing data mechanism of the sample sizes is assumed to be MAR and a truncated Poisson model is employed for modeling. It is a possibility to consider other models for the missing sample sizes and extend the assumption for the missing data mechanism to be missing not at random (MNAR).

ACKNOWLEDGEMENTS

We would like to thank the Guest Editor and the reviewer for their helpful suggestions and comments, which have led to a much improved version of the manuscript. Dr. Chen's research was partially supported by NIH grants #GM70335 and #P01CA142538, and Dr. Tang's research was supported by National Nature Science Foundation of China #41870709.

Received 25 January 2019

REFERENCES

[1] Burr, D. and Doss, H. (2005). A Bayesian semiparametric model for random-effects meta-analysis. *Journal of the American Statistical Association* **100** 242–251. MR2156834

[2] Chen, M-H., Shao, Q-M. and Ibrahim, J. G. (2000). *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media. MR1742311

[3] Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* **91** 883–904. MR1395755

[4] de Valpine, P., Turek, D., Paciorek C. J., Anderson-Bergman, C., Lang D. T. and Bodik, R. (2017). Programming with models: writing statistical algorithms for general model structures with nimble. *Journal of Computational and Graphical Statistics* **26** 403–413. MR3640196

[5] Enquist, B. J., Brown, J. H. and West, G. B. (1998). Allometric scaling of plant energetics and population density. *Nature* **395** 163.

[6] Gelfand, A. E., Smith, A. F. and Lee, T. M. (1992). Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association* **87** 523–532. MR1173816

[7] Harper, J. L. (1997). *Population Biology of Plants*. Academic Press.

[8] Higgins, J. P., White, I. R. and Wood, A. M. (2008). Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clinical Trials* **5** 225–239.

[9] Hutchings, M. J. (1983). Shoot performance and population structure in pure stands of mercurialis perennis l., a rhizomatous perennial herb. *Oecologia* **58** 260–264.

[10] Keeley, E. R. (2003). An experimental analysis of self-thinning in juvenile steelhead trout. *Oikos* **102** 543–550.

[11] Lee, K. J. and Thompson, S. G. (2008). Flexible parametric models for random-effects distributions. *Statistics in Medicine* **27** 418–434. MR2418453

[12] Miyanishi, K., Hoy, A. R. and Cavers, P. B. (1979). A generalized law of self-thinning in plant populations. *Journal of Theoretical Biology* **78** 439–442.

[13] Quartagno, M. and Carpenter, J. R. (2016). Multiple imputation for ipd meta-analysis: allowing for heterogeneity and studies with missing covariates. *Statistics in Medicine* **35** 2938–2954. MR3528235

[14] Robertson, C., Idris, N. R. and Boyle, P. (2004). Beyond classical meta-analysis: can inadequately reported studies be included? *Drug Discovery Today* **9** 924–931.

[15] Stevens, J. W. (2011). A note on dealing with missing standard errors in meta-analyses of continuous outcome measures in Winbugs. *Pharmaceutical Statistics* **10** 374–378,.

[16] Tofallis, C. (2002) Model fitting for multiple variables by minimising the geometric mean deviation. In *Total Least Squares and Errors-in-Variables Modeling* 261–267. Springer, 2002. MR1952953

[17] Turner, R. M., Omar, R. Z. and Thompson, S. G. (2001). Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine* **20** 453–472.

[18] Weller, D. E. (1987). A reevaluation of the-3/2 power rule of plant self-thinning. *Ecological Monographs* **57** 23–43.

[19] West, G. B., Brown, J. H. and Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. *Science* **276** 122–126.
[20] White, J. (1985). *The thinning rule and its application to mixtures of plant populations.* Academic Press.
[21] Whitehead, A. (2002). *Meta-analysis of controlled clinical trials*, volume 7. John Wiley & Sons.
[22] Yao, H., Chen, M-H. and Qiu, C. (2011). Bayesian modeling and inference for meta-data with applications in efficacy evaluation of an allergic rhinitis drug. *Journal of Biopharmaceutical Statistics* **21** 992–1005. MR2823362
[23] Yao, H., Kim, S., Chen, M-H., Ibrahim, J. G., Shah, A. K. and Lin, J. (2015). Bayesian inference for multivariate meta-regression with a partially observed within-study sample covariance matrix. *Journal of the American Statistical Association* **110** 528–544. MR3367245
[24] Yoda, K., Kira T., Ogawa, H. and Hozumi, K. (1963). Self-thinning in overcrowded pure stands under cultivated and natural conditions. *Journal of Biology, Osaka City University* **14** 107–129.
[25] Zhang, D., Chen, M-H., Ibrahim, J. G., Boye, M.E. and Shen, W. (2017). Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials. *Journal of Computational and Graphical Statistics* **26** 121–133. MR3610413

Zhihua Ma
Department of Statistics, School of Economics
Shenzhen University
Shenzhen
China
E-mail address: mazh1993@outlook.com

Ming-Hui Chen
Department of Statistics
University of Connecticut
Storrs, CT
USA
E-mail address: ming-hui.chen@uconn.edu

Yi Tang
School of Life Science
Liaoning University
Shenyang
China
E-mail address: tangyi@lnu.edu.cn