Biometrika (2018), **103**, 1, *pp*. 1–7
Printed in Great Britain Advance Access publication on 31 July 2018

Analysis of grouped data using conjugate generalized linear mixed models

BY JAROD Y. L. LEE, PETER J. GREEN AND LOUISE M. RYAN

School of Mathematical and Physical Sciences, University of Technology Sydney, Sydney 2007, Australia

yan.lee@uts.edu.au peter.green@uts.edu.au louise.m.ryan@uts.edu.au

SUMMARY

This article concerns a class of generalized linear mixed models for two-level grouped data, where the random effects are uniquely indexed by groups and are independent. We derive necessary and sufficient conditions for the marginal likelihood to be expressed in explicit form. These models are unified under the conjugate generalized linear mixed models framework, where conjugate refers to the fact that the marginal likelihood can be expressed in closed form, rather than implying inference via the Bayesian paradigm. The proposed framework allows simultaneous conjugacy for Gaussian, Poisson and gamma responses, and thus can accommodate both unit- and group-level covariates. Only group-level covariates can be incorporated for the binomial distribution. In a simulation of Poisson data, our framework outperformed its competitors in terms of computational time, and was competitive in terms of robustness against misspecification of the random effects distributions.

Some key words: Closed-form marginal likelihood; Longitudinal data; Multilevel model; Random effect; Unit-level model.

1. INTRODUCTION

Generalized linear mixed models can account for the dependence structure of multilevel and longitudinal data, where the responses of units within a group are correlated. The goal is to model the response as a function of unit and group-level covariates while accounting for group-to-group variability. For example, outcomes of patients within the same hospital are likely to be dependent due to similar risk profiles and a common clinical management practice. Generalized linear mixed models provide a natural framework for modelling dependencies by allowing for random group-specific effects. The random effects are typically assumed to be Gaussian and additive with respect to the linear predictors.

Despite their popularity, generalized linear mixed models are computationally intensive to fit. Estimation for generalized linear mixed models is typically likelihood-based, involving integrals which do not usually have analytic expressions. Common estimation procedures include numerical quadrature (Rabe-Hesketh et al., 2002), Monte Carlo methods, Laplace approximation (Tierney & Kadane, 1986), penalized quasi-likelihood (Breslow & Clayton, 1993), hierarchical likelihood (Lee & Nelder, 1996), and simulated maximum likelihood (Train, 2009, p.238–239). Some of these approaches apply an expectation-maximization algorithm that treats the random effects as missing data (McCulloch, 1997). Certain generalized linear mixed models can also be fitted using softwares for generalized additive models (Wood, 2017, p.288).

For large-scale applications, it is important that models can be fitted in a reasonable time. Several methods have been proposed in various specific settings. Zhang & Koren (2007) exploit the sparsity of predictors to achieve speedup for Bayesian hierarchical models. Luts et al. (2014) fit variational approximation Bayesian hierarchical models for streaming data. Perry (2017) proposes a non-iterative moment-based procedure that is well-suited for distributed computing.

This article proposes a new framework for analyzing multilevel and longitudinal data, where the marginal likelihood is tractable. Having an explicit marginal likelihood means that the models are more computationally convenient. We focus on two-level grouped data, where the random effects are uniquely indexed by groups and are independent.

2. EXPONENTIAL FAMILY AND CONJUGATE PRIOR

The likelihood of a one-parameter exponential family with dispersion can be written as

$$f_{Y|\theta}(y \mid \theta, \phi) = \exp\left[\{y\theta - b(\theta)\}/\phi + c(y, \phi)\right], \tag{1}$$

for some specified functions $b(\theta)$ and $c(y, \phi)$, where θ is the canonical parameter and can be expressed as a function of the mean $\theta(\mu)$, and ϕ is the dispersion parameter, assumed known.

For such an exponential family, there exists a family of prior distributions on θ such that the posterior density lies in the same family as the prior. Such a conjugate prior for θ is defined as

$$f_\Theta(\theta \mid \chi, \nu) = g(\chi, \nu) \exp\{\chi\theta - \nu b(\theta)\}, \tag{2}$$

where χ and ν are parameters and $g(\chi, \nu)$ denotes the normalizing factor. The posterior density is of the form

$$f_{\Theta|y}(\theta \mid y, \chi, \nu, \phi) \propto \exp\{c(y, \phi)\} g(\chi, \nu) \exp\{\theta(\chi + y/\phi) - b(\theta)(\nu + 1/\phi)\}, \tag{3}$$

which has the same kernel as the prior but with different parameters. The kernel of a probability density function is the form after the normalization factor is removed. The updated parameters, based on a single observation y, are $\tilde{\chi} = \chi + y/\phi$ and $\tilde{\nu} = \nu + 1/\phi$. For n independent and identically distributed observations $y_1, \ldots, y_n$, it is straightforward to show that conjugacy still holds and the updated parameters are $\tilde{\chi} = \chi + \sum_{j=1}^{n} y_j/\phi$ and $\tilde{\nu} = \nu + n/\phi$.

These are standard results for independent and identically distributed data in the Bayesian context. In this article, we aim to compute an explicit marginal likelihood for generalized linear mixed models in the frequentist setting. This is attained by connecting the posterior in the Bayesian paradigm and the marginal likelihood in the frequentist paradigm, and relaxing the assumption of identical distribution. The result is a class of models which we refer to as the conjugate generalized linear mixed models, where unit-level covariates can be conveniently incorporated while maintaining a closed-form marginal likelihood.

3. CONJUGATE GENERALIZED LINEAR MIXED MODELS

3·1. *From Bayesian formalism to frequentist inference*

We now move from the Bayesian paradigm, where θ is a parameter and its distribution is the prior, to the frequentist paradigm, where θ is a group-specific random effect and its distribution describes variation between groups. Consider the two-level setting where the responses $y_{ij} (j = 1, \ldots, n_i)$ are grouped within a higher-level structure indexed by $i = 1, \ldots, I$. The responses are assumed to come from the same exponential family. Random effects with a specified distribution are introduced at the group level. Within each group, the responses are conditionally independent given the group-specific random effects.

For this model setup, the marginal likelihood, obtained by integrating out the random effects, is

$$\prod_{i=1}^{I} \int \prod_{j=1}^{n_i} f_{Y|\theta_i}(y_{ij} \mid \theta_i, \phi) f_{\Theta_i}(\theta_i \mid \chi, \nu) d\theta_i,$$

where the integrand for a single observation is proportional to the posterior in (3). Imposing a conjugate prior distribution on the random effects would ensure that the integrand comes from a recognizable density function, which enable the marginal likelihood to be expressed in closed form. Here and for the rest of the paper, we use the word prior in the frequentist setting for brevity.

Solving for the integral, the marginal likelihood contribution for the entire data is

$$\prod_{i=1}^{I}\left[\frac{\exp\left\{\sum_{j=1}^{n_i} c(y_{ij},\phi)\right\} g(\chi,\nu)}{g\left(\chi+\sum_{j=1}^{n_i} y_{ij}/\phi, \nu+n_i/\phi\right)}\right].$$

This is the formulation for group-level models in the absence of of unit-level covariates. Although the random effects $\theta_i = \theta(\mu_i)$ are typically expressed in terms of a monotonic transformation of μ_i, interest usually lies in the distribution of μ_i. Consonni & Veronese (1992) and Gutiérrez-Pẽna & Smith (1995) showed that the conjugate distribution on μ_i coincides with the prior on μ_i induced by the conjugate distribution on θ_i if and only if the exponential family has a quadratic variance function. This holds for the Gaussian, Poisson, binomial, and gamma distributions (Morris, 1983), providing a convenient way to incorporate group-level variables, for example, via the mean μ_i using a monotonic link function.

3·2. *Relaxing the assumption of identical distribution: unit-level models*

Relaxing the assumption of identical distribution, we consider the regression setting where each observation y_{ij} may have a separate parameter $\theta_{ij} = \theta(x_{ij})$ that is a function of the covariates, while ϕ, if present, is constant across all observations. We want to explore the most generic formulation that leads to marginal likelihood simplification, so at this stage we leave open the functional dependence of θ_{ij} on x_{ij}. Denote $\theta_0 = \theta(x_0)$ as the baseline parameter, where x_0 is an arbitrary baseline covariate value. Technically, θ_0 is also indexed by i to reflect the group correlated data structure, but generally this can be suppressed without ambiguity. Likewise, for ease of notation, the i and j indexing are suppressed for most of the remaining article.

Remark 1. A common assumption is $x_0 = 0$, but users can take any baseline appropriate for the problem at hand. With this formulation, within a group, we can think of units with covariate configurations that deviate from the baseline characteristics as modifying θ_0. This is as opposed to the standard formulation of generalized linear mixed models, where for a given unit with a particular covariate configuration, it is the group membership that modifies the linear predictor.

The log of the integrand of the marginal likelihood for a single observation is $[y\theta(x) - b\{\theta(x)\}]/\phi + \chi\theta_0 - \nu b(\theta_0)$. A conjugate prior distribution is imposed on θ_0, rather than explicitly on $\theta_{ij} = \theta(x_{ij})$. The integrand of the marginal likelihood lies in the same family as (2) in its dependence on θ_0 if and only if both $\theta(x)$ and $b\{\theta(x)\}$ are affine functions of θ_0 and $b(\theta_0)$, i.e., if there exist functions p, q, r, s, t and u of x such that $\theta(x) = p(x)\theta_0 + q(x)b(\theta_0) + r(x)$ and $b\{\theta(x)\} = s(x)\theta_0 + t(x)b(\theta_0) + u(x)$. These conditions can be combined to obtain

$$b\left\{p(x)\theta_0 + q(x)b(\theta_0) + r(x)\right\} = s(x)\theta_0 + t(x)b(\theta_0) + u(x). \tag{4}$$

This is the key equation in deriving the functional solutions for p, q, r, s, t and u. We are interested in families where $\theta(x)$ has non-trivial dependence on x, that is, at least one of p, q or r must depend on x. When this occurs, the induced prior for $\theta(x)$ exhibits simultaneous conjugacy across all values of x, and the resulting model is capable of incorporating unit-level covariates while maintaining a closed-form marginal likelihood. This holds for Gaussian, Poisson, and gamma responses. For Gaussian responses, the functional form of the mean is $\mu(x)$ is $\zeta_1(x)\mu_0 + \zeta_2(x)$, where ζ_1 and ζ_2 are user-defined functions of x, subject to $\zeta_1(x_0) = 1$ and $\zeta_2(x_0) = 0$. For gamma responses, $\mu(x) = \mu_0/\zeta(x)$, where $\zeta(x)$ is a user-specified function of x, subject to $\zeta(x_0) = 1$. Section 4·1 discusses Poisson data. For Gaussian data, modelling under the conjugate generalized linear mixed model framework coincides with that of generalized linear mixed model framework. When the dependence of $\theta(x)$ on x is trivial, this formulation reduces to the group-level models. See the Supplementary Material.

Since $\theta_0 = \theta(x_0)$ and $b(\theta_0) = b\{\theta(x_0)\}$, it is clear that $p(x_0) = 1$, $q(x_0) = 0$, $r(x_0) = 0$, $s(x_0) = 0$, $t(x_0) = 1$ and $u(x_0) = 0$. These constraints must be satisfied when choosing functional solutions for p, q, r, s, t and u.

Under (4), the log of the integrand of the marginal likelihood for a single observation is $\theta_0[\chi + \{yp(x) - s(x)\}/\phi] - b(\theta_0)[\nu + \{t(x) - yq(x)\}/\phi]$. Solving for this integral, the marginal likelihood contribution for observations within a single group is

$$\frac{\exp\left\{\sum_{j=1}^{n_i} c(y_{ij}, \phi)\right\} g(\chi, \nu) \exp\left[\sum_{j=1}^{n_i} \{r(x_{ij})y_{ij} - u(x_{ij})\}/\phi\right]}{g\left[\chi + \sum_{j=1}^{n_i} \{y_{ij}p(x_{ij}) - s(x_{ij})\}/\phi, \nu + \sum_{j=1}^{n_i} \{t(x_{ij}) - y_{ij}q(x_{ij})\}/\phi\right]}. \tag{5}$$

For multiple groups, the marginal likelihood can be obtained by multiplying (5) across the group index i.

4. POISSON RESPONSES

4·1. *Derivation*

The Poisson density function can be written in the form $\exp(y \log \mu_0 - \mu_0 - \log y!)$, where $\mu_0 > 0$ is the rate parameter. This can be written in the form of (1) if we write $\theta_0 = \log \mu_0$, $b(\theta_0) = e^{\theta_0}$, $\phi = 1$ and $c(y, \phi) = -\log y!$. To determine the conjugate distribution for θ_0, we compute the normalization factor

$$g(\chi, \nu) = \left[\int \exp\{\chi\theta_0 - \nu \exp(\theta_0)\}\, d\theta_0\right]^{-1} = \frac{\nu^\chi}{\Gamma(\chi)},$$

where the integrand is the kernel of a log-gamma density function with shape $A = \chi > 0$ and scale $B = \nu^{-1} > 0$, $\Gamma(\cdot)$ is the gamma function. This implies that $\mu_0 = \exp(\theta_0) \sim$ Gamma(A, B).

Christiansen & Morris (1997) considered a similar model without covariates in the Bayesian setting. Group-level covariates can be incorporated via the mean of μ_0, by letting $E(\mu_0) = AB \equiv \exp(x_i^T\beta)$ for example. As a result, we replace B by $B_i = \exp(x_i^T\beta)/A$. To incorporate unit-level covariates, (4) requires $b\{\theta(x)\} = \exp\{p(x)\theta_0 + q(x)\exp(\theta_0) + r(x)\} \equiv s(x)\theta_0 + t(x)\exp(\theta_0) + u(x)$, which gives $p(x) = 1$, $q(x) = 0$, $r(x) = \zeta(x)$, $s(x) = 0$, $t(x) = e^{\zeta(x)}$, and $u(x) = 0$, where $\zeta(x)$ is a user-specified function of x, subject to $\zeta(x_0) = 0$. This implies $\theta(x) = \log \mu(x) = \theta_0 + \zeta(x)$, or equivalently, $\mu(x) = \mu_0 \exp \zeta(x)$.

Choosing $\zeta(x) = x^T\beta$ leads to $\mu(x) = \mu_0 \exp(x^T\beta)$, where x does not include the constant 1 so that $\zeta(x_0) = 0$, assuming $x_0 = 0$. This choice gives rise to a sensible model as $\mu(x)$ is guaranteed positive. Similar multiplicative models with unit-level covariates have been considered by Lee & Nelder (1996), and Lee et al. (2017a,b) under various settings. An alternative formulation is to include the intercept but with constraint $E(\mu_0) = 1$. Consequently, $B = 1/A$ and var(μ_0) = $1/A$. We used the latter formulation in our simulation as it allows a direct comparison with Poisson generalized linear mixed models, since both frameworks involve an intercept and a constraint on the random intercepts.

An estimator for var(μ_0) can be inferred from the maximum likelihood estimate for A, $\hat{A}$, via $1/\hat{A}$. Predictions for the random effects can be obtained by minimizing the overall mean squared error of prediction, resulting in the best predictor $\hat{\mu}_{0i} = (\sum_{j=1}^{n_i} y_{ij} + \hat{A})/\{\sum_{j=1}^{n_i} \exp(x_{ij}^T\hat{\beta}) + \hat{A}\}$, where $\hat{\beta}$ is the maximum likelihood estimate for β. Starting values for the fixed effects can be obtained by fitting a Poisson generalized linear model.

4·2. *Simulation study*

We conducted a limited simulation study to assess the performance of Poisson conjugate generalized linear mixed models: $y_{ij} \mid \mu_{0i} \sim$ Poisson$\{\mu_{0i}\exp(x_{ij}^T\beta)\}$, $\mu_{0i} \sim$ Gamma$(A, 1/A)$ versus generalized linear mixed models: $y_{ij} \mid b_i \sim$ Poisson$\{\exp(x_{ij}^T\beta + b_i)\}$, $b_i \sim$ Gaussian$(0, \sigma^2)$ in terms of computational speed and inferential accuracy. Due to the intractable marginal likelihood of the Poisson generalized linear mixed models, various approximation methods were used to estimate the marginal likelihood. These methods are implemented within the R package lme4 (Bates et al., 2017). Results for generalized linear models are included for comparison. We generated data from the two models, and fitted each dataset using both Poisson models with multiplicative gamma and additive Gaussian random effects. Each dataset consists of $50,000$ groups, two observations within each group, and a binary predictor. The true parameter

Table 1: Estimated fixed intercept $\hat{\beta}_0$, estimated fixed slope $\hat{\beta}_1$, estimated standard error of the random effects $\hat{\sigma}$, average deviance, and relative elapsed times for Poisson generalized linear models, conjugate generalized linear mixed models, and generalized linear mixed models. The numbers are averages across 1000 simulations.

(a) True distribution: Gamma multiplicative random effects

	GLM	CGLMM	GLMM			
			LAP	AGQ2	AGQ5	AGQ10
$\hat{\beta}_0$	0.50	0.50	0.03	0.03	0.03	0.03
$\hat{\beta}_1$	1.00	1.00	1.00	1.00	1.00	1.00
$\hat{\sigma}$	NA	1.00	1.02	1.01	1.02	1.03
Deviance	3.61	0.68	0.71	0.71	0.71	0.71
Relative time	0.02	1.00	1.95	2.56	2.35	2.94

(b) True distribution: Gaussian additive random effects

	GLM	CGLMM	GLMM			
			LAP	AGQ2	AGQ5	AGQ10
$\hat{\beta}_0$	1.00	1.00	0.50	0.50	0.50	0.50
$\hat{\beta}_1$	1.00	1.00	1.00	1.00	1.00	1.00
$\hat{\sigma}$	NA	0.99	0.99	0.99	1.00	1.00
Deviance	6.12	0.65	0.68	0.68	0.68	0.68
Relative time	0.02	1.00	1.56	1.60	1.98	2.53

GLM, generalized linear models; CGLMM, conjugate generalized linear mixed models; LAP, Laplace approximation; AGQ2, adaptive Gauss-Hermite quadrature with 2 quadrature points; AGQ5, adaptive Gauss-Hermite quadrature with 5 quadrature points; AGQ10, adaptive Gauss-Hermite quadrature with 10 quadrature points; NA, not applicable.

values are: $\beta_0 = 0.5$, $\beta_1 = 1$, and $\sigma = 1$. For models where the true random effects distribution is gamma, the true value of A is implied by σ, that is, $A = \sigma^{-2}$.

Table 1 presents the results averaged across 1000 simulations. The fixed intercept estimates are best interpreted in conjunction with the random intercepts, as a bias in the fixed intercept may be corrected by a consistent shift in the predicted random intercepts. The fixed slopes and the standard errors of the random effects estimates are similar for both models, regardless of the true random effects distributions. We used the average deviance to measure the overall inferential accuracy of the models. For Poisson responses, the average deviance is $2n^{-1}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\{y_{ij}\log(y_{ij}/\hat{\mu}_{ij}) - y_{ij} + \hat{\mu}_{ij}\}$, where $\hat{\mu}$ is the predicted mean based on the estimated model parameters. If the model fits the data well, the observed values will be close to their predicted values, resulting in a small average deviance. The average deviance was very similar across the models even when the true random effects distribution is misspecified. This suggests that the overall accuracy of the proposed Poisson conjugate generalized linear mixed models is comparable to that of the Poisson generalized linear mixed models when the random effects distribution is misspecified, at least within this simulation setting. The nature of the simulation study prevents us from looking at the distribution of the predicted random effects closely, but it would be interesting to investigate this in applications. The Poisson conjugate generalized linear mixed models took less time to run, even though the calculations were done without careful optimization of code as in the competing lme4 package.

The simulation was designed with big data in mind. In particular, we simulated data with a large number of groups and with few observations within each group. In practice, accuracy will depend on a number of factors, including the number of observations within groups, and standard deviation of the random effects. See the 2017 University of Technology Sydney PhD thesis by Jarod Y. L. Lee for a comprehensive simulation study.

5. REMARKS

Group-level conjugate models have long been used in the context of Bayesian small area estimation and disease mapping, the most common ones being the gamma-Poisson (Rao & Molina, 2015, p. 383) and the beta-binomial models (Rao & Molina, 2015, p. 389). The primary advantage of the proposed modelling framework is mathematical convenience, but the assumed conjugate random effect distribution may not accurately reflect the real variation between groups. Other applications of the proposed framework include privacy preservation in large-scale administrative databases (Lee et al., 2017a) and the fitting of discrete choice models (Lee et al., 2017b).

Except for Gaussian responses, our framework can only handle a single layer of random effects corresponding to two levels of grouping, with a unique random effect per group. For multilevel models, one may alleviate some computational complexity by imposing random effects via the conjugate generalized linear mixed models framework on the first layer, and then using Gaussian random effects for the remaining layers. Similar strategy can be applied to models with more than one random effect per group, where we can impose a distribution for the random intercepts via our framework, and then using Gaussian random effects for the random slopes. For more complicated structures, such as crossed designs, the proposed framework provides a computationally efficient way for obtaining sensible starting values.

Some of the models derived from our conjugate generalized linear mixed models framework are similar to those of the conjugate hierarchical generalized linear models framework proposed by Lee & Nelder (1996). While conjugate in our framework refers to the fact that the marginal likelihood can be made explicit, it has quite a different meaning in the hierarchical likelihood framework (Lee & Nelder, 1996, p. 621), where it refers to the fact that a Bayesian conjugate prior is imposed on the random effects distribution, though this may not yield a closed-form marginal likelihood.

Molenberghs et al. (2010) consider models that can simultaneously accommodate both overdispersion and correlation induced by grouping structures via two separate sets of random effects. Although they use the conjugate distribution for a set of random effects, the resulting marginal likelihood is generally not explicit.

ACKNOWLEDGEMENT

We thank the editors and referees for constructive criticism. Lee and Ryan are member of the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers. Ryan is also an Adjunct Professor of Biostatistics at Harvard University. Green is also a Professorial Research Fellow at University of Bristol.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes derivations for Gaussian, binomial and gamma responses, and an illustrative example. R code for simulation can be found at Github: `https://github.com/jarodleeyl/conjugate_glmm`.

REFERENCES

BATES, D., MÄCHLER, M., BOLKER, B., WALKER, S., CHRISTENSEN, R. H. B., SINGMANN, H., DAI, B., SCHEIPL, F., GROTHENDIECK, G. & GREEN, P. (2017). *lme4: Linear Mixed-Effects Models using 'Eigen' and S4*. R package version 1.1-15.

BRESLOW, N. E. & CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Am. Statist. Assoc.* **88**, 9–25.

CHRISTIANSEN, C. L. & MORRIS, C. N. (1997). Hierarchical Poisson regression modeling. *J. Am. Statist. Assoc.* **92**, 618–632.

CONSONNI, G. & VERONESE, P. (1992). Conjugate priors for exponential families having quadratic variance functions. *J. Am. Statist. Assoc.* **87**, 1123–1127.

GUTIÉRREZ-PĔNA, E. & SMITH, A. F. M. (1995). Conjugate parameterizations for natural exponential families. *J. Am. Statist. Assoc.* **90**, 1347–1356.

Lee, J. Y. L., Brown, J. J. & Ryan, L. M. (2017a). Sufficiency revisited: Rethinking statistical algorithms in the big data era. *Am. Stat.* **71**, 202–208.

Lee, J. Y. L., Green, P. J. & Ryan, L. M. (2017b). On the "Poisson trick" and its extensions for fitting multinomial regression models. *arXiv preprint arXiv:1706.09523* .

Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *J. R. Statist. Soc.* B **58**, 619–678.

Luts, J., Broderick, T. & Wand, M. (2014). Real-time semiparametric regression. *J Comput. Graph. Stat.* **23**, 589–615.

McCulloch, C. (1997). Maximum likelihood algorithms for generalized linear mixed models. *J. Am. Statist. Assoc.* **92**, 162–170.

Molenberghs, G., Verbeke, G., Demétrio, C. G. & Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Stat. Sci.* **25**, 325–347.

Morris, C. N. (1983). Natural exponential families with quadratic variance functions: Statistical theory. *Ann. Statist.* **11**, 515–529.

Perry, P. O. (2017). Fast moment-based estimation for hierarchical models. *J. R. Statist. Soc.* B **79**, 267–291.

Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal* **2**, 1–21.

Rao, J. N. K. & Molina, I. (2015). *Small Area Estimation*. Hoboken, NJ: Wiley.

Tierney, L. & Kadane, J. (1986). Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Assoc.* **81**, 82–86.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Boca Raton, FL: Chapman and Hall.

Zhang, Y. & Koren, J. (2007). Efficient Bayesian hierarchical user modeling for recommendation systems. In *Proc. 30th A. Int. Association for Computing Machinery Special Interest Group on Information Retrieval Conf. Research and Development in Information Retrieval*. New York: Association for Computing Machinery.

[*Received* 2 *January* 2018. *Editorial decision on* 1 *April* 2018]