Bayesian Analysis (2016) **11**, Number 2, pp. 545–571

Posterior Propriety for Hierarchical Models with Log-Likelihoods That Have Norm Bounds*

Sarah E. Michalak† and Carl N. Morris‡

Abstract. Statisticians often use improper priors to express ignorance or to provide good frequency properties, requiring that posterior propriety be verified. This paper addresses generalized linear mixed models, GLMMs, when Level I parameters have Normal distributions, with many commonly-used hyperpriors. It provides easy-to-verify sufficient posterior propriety conditions based on dimensions, matrix ranks, and exponentiated norm bounds, ENBs, for the Level I likelihood. Since many familiar likelihoods have ENBs, which is often verifiable via log-concavity and MLE finiteness, our novel use of ENBs permits unification of posterior propriety results and posterior MGF/moment results for many useful Level I distributions, including those commonly used with multilevel generalized linear models, e.g., GLMMs and hierarchical generalized linear models, HGLMs. Those who need to verify existence of posterior distributions or of posterior MGFs/moments for a multilevel generalized linear model given a proper or improper multivariate F prior as in Section 1 should find the required results in Sections 1 and 2 and Theorem 3 (GLMMs), Theorem 4 (HGLMs), or Theorem 5 (posterior MGFs/moments).

Keywords: exponentiated norm bound, generalized linear mixed model, hierarchical generalized linear model, improper prior, multilevel objective Bayes.

1 Introduction

Bayesian approaches are used widely with multi-level models. The hierarchical models considered here have three levels: the likelihood of the observed data (Level I), the distributions of the unknown Level I parameters (Level II), and the prior distribution of the unknown Level II hyperparameters (Level III). Improper (not finitely integrable) Level III distributions on the Level II parameters are often used in practice by Bayesians and objective Bayesians to represent ignorance and by many statisticians to achieve frequency-calibrated procedures. The need to verify posterior propriety has been stressed (Gelman et al., 1995; Carlin and Louis, 1996; Hobert and Casella, 1996, 1998; Natarajan and Kass, 2000). Unfortunately, except in the simplest models, when improper priors are used it can be daunting and time-consuming to verify that the resulting posterior

*This work has been authored by an employee of Los Alamos National Security, LLC, operator of the Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting this work for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce this work, or allow others to do so for United States Government purposes. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; however, the Laboratory as an institution does not endorse the viewpoint of a publication or guarantee its technical correctness. This work is published under LA-UR-13-27701.

†Los Alamos National Laboratory, Los Alamos, NM 87544, michalak@lanl.gov
‡Harvard University, Cambridge, MA 02138, morris@stat.harvard.edu

 DOI: 10.1214/15-BA962

distribution is proper. This demonstration sometimes is skipped, occasionally at the cost of making inference from an improper posterior distribution. For example, Natarajan and Kass (2000) prove that for certain GLMMs, a commonly-used invariant prior for the variance of the random effects used in Tiao and Tan (1965), Wang et al. (1994), and Zeger and Karim (1991) leads to an improper joint posterior distribution when that prior for the random effects variance and any prior (proper or improper) for the fixed effects are *a priori* independent. See Hobert and Casella (1998) and Berger et al. (2005) for further discussion of posterior impropriety with this prior. Although the output from a Markov chain Monte Carlo method may not indicate posterior impropriety (Hobert and Casella, 1996, 1998; Natarajan and McCulloch, 1998), typical Markov chain Monte Carlo estimates of expected values may not be consistent and may converge to 0 with probability 1 if the posterior distribution is improper (see Athreya and Roy (2014) and Hobert and Casella (1998), with related results in Athreya and Roy (2015)). Recognizing these problems, Bayesian analysts sometimes use proper prior distributions to assure posterior propriety, perhaps at the risk of overstating their prior information.

The development of the posterior propriety conditions presented here was motivated by our collaboration with U.S. Department of Veterans Affairs (VA) researchers, which used multi-level models for VA hospital evaluations. It was important to use noninformative priors at Level III (flat priors on the Level II regression coefficients and the Level II covariance matrix), partly to assure good frequentist properties, which generally require noninformative priors, and partly to assure hospital administrators that the procedures let the data speak. The models used were hierarchical generalized linear models, HGLMs, *e.g.*, hierarchical logistic regression models. The results here generalize from HGLMs to the following model structure, which includes GLMMs and HGLMs as special cases.

Assume a hierarchical model in which n_h observations are modeled for group h at Level I, $h = 1, \ldots, k_0$, with y_{hi} the observation from the ith unit in group h. In particular, in Level I the data vector $\boldsymbol{y}_h = (y_{h1}, \ldots, y_{hn_h})'$ has density or mass function f_h with parameter $\boldsymbol{\eta}_h = (\eta_{h1}, \ldots, \eta_{hn_h})'$ which follows a mixed model

$$\boldsymbol{y}_h \sim f_h(\cdot|\boldsymbol{\eta}_h);\ \boldsymbol{\eta}_h = \boldsymbol{U}_h\boldsymbol{\gamma} + \boldsymbol{X}_h\boldsymbol{\varepsilon}_h \tag{1}$$

where the $\boldsymbol{y}_h$ are assumed independent unless otherwise specified, $\boldsymbol{U}_h$ is $n_h \times d_u$ with ith row equal to $\boldsymbol{u}'_{hi}$, the $d_u \times 1$ vector of known fixed effects covariates for unit i, $\boldsymbol{\gamma}$ is a $d_u \times 1$ vector of typically unknown fixed effects, $\boldsymbol{X}_h$ is $n_h \times d_x$ with ith row equal to $\boldsymbol{x}'_{hi}$, the $d_x \times 1$ vector of known random effects covariates for unit i, and $\boldsymbol{\varepsilon}_h$ is a $d_x \times 1$ vector of unknown random effects specific to group h. Although each unit has a parameter η_{hi} associated with it, the number of unknown model parameters is typically far less than $\sum_{h=1}^{k_0} n_h$, which is the total number of elements in $(\boldsymbol{\eta}'_1, \ldots, \boldsymbol{\eta}'_{k_0})'$, because each η_{hi} is a linear function of $\boldsymbol{\gamma}$ and its $\boldsymbol{\varepsilon}_h$.

More general notation for $k \leq k_0$ groups, where k is defined immediately above (13), incorporates the observed data vector $\boldsymbol{y} = (\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_k)'$ and parameters $\boldsymbol{\eta} = (\boldsymbol{\eta}'_1, \ldots, \boldsymbol{\eta}'_k)'$, both with dimension $N = \sum_{h=1}^{k} n_h$, in a likelihood $L(\boldsymbol{\eta}|\boldsymbol{y})$ that does not necessarily assume independent observations, so that

$$\boldsymbol{\eta} = \boldsymbol{U}\boldsymbol{\gamma} + \boldsymbol{X}\boldsymbol{\varepsilon}. \tag{2}$$

With $K = k \cdot d_x$, $\boldsymbol{U}$ $(N \times d_u)$ is $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_k$ stacked one on top of the other, $\boldsymbol{X}$ $(N \times K)$ is block-diagonal with $\boldsymbol{X}_h$ in the hth position, and

$$\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1, \ldots, \boldsymbol{\varepsilon}'_k)' \sim N_K(\mathbf{0}, \boldsymbol{\Lambda}). \tag{3}$$

Here $\boldsymbol{\Lambda}$ $(K \times K)$ may be fully unspecified and the prior density $p(\boldsymbol{\gamma}, \boldsymbol{\Lambda})$ is assumed to satisfy

$$p(\boldsymbol{\gamma}, \boldsymbol{\Lambda}) d\boldsymbol{\gamma} d\boldsymbol{\Lambda} \leq M \pi(\boldsymbol{\Lambda}) d\boldsymbol{\gamma} d\boldsymbol{\Lambda} \tag{4}$$

where the finite constant M permits prior densities on $\boldsymbol{\gamma}$ that are bounded with respect to Lebesgue measure.

Special cases include GLMMs and HGLMs. For the GLMMs addressed here, the $d_x \times 1$ random effects vector $\boldsymbol{\varepsilon}_h$ is assumed to follow a multivariate Normal distribution in Level II with mean $\mathbf{0}$ and unknown variance matrix $\boldsymbol{A}$ that is fully unspecified so that $\boldsymbol{\Lambda}$ from (3) is equal to $\boldsymbol{I}_k \otimes \boldsymbol{A}$ and

$$\boldsymbol{\varepsilon}_h = (\varepsilon_{h1}, \ldots, \varepsilon_{hd_x})' \overset{\text{ind.}}{\sim} N_{d_x}(\mathbf{0}, \boldsymbol{A}). \tag{5}$$

The prior density for $\boldsymbol{\gamma}$ and $\boldsymbol{A}$ in Level III is assumed to have the following bound:

$$p(\boldsymbol{\gamma}, \boldsymbol{A}) d\boldsymbol{\gamma} d\boldsymbol{A} \leq M \pi(\boldsymbol{A}) d\boldsymbol{\gamma} d\boldsymbol{A} \tag{6}$$

for some finite constant M.

The HGLMs considered here have Levels I, II, and III specified as follows:

$$\boldsymbol{y}_h \overset{\text{ind.}}{\sim} f_h(\cdot|\boldsymbol{\eta}_h); \;\; \boldsymbol{\eta}_h = \boldsymbol{Z}_h \boldsymbol{\beta}_h, \tag{7}$$

$$\boldsymbol{\beta}_h \overset{\text{ind.}}{\sim} N_{d_z}((\boldsymbol{I}_{d_z} \otimes \boldsymbol{w}'_h)\boldsymbol{\gamma}, \boldsymbol{A}), \tag{8}$$

$$p(\boldsymbol{\gamma}, \boldsymbol{A}) d\boldsymbol{\gamma} d\boldsymbol{A} \leq M \pi(\boldsymbol{A}) d\boldsymbol{\gamma} d\boldsymbol{A} \tag{9}$$

where $\boldsymbol{Z}_h$ is $n_h \times d_z$ with ith row equal to $\boldsymbol{z}'_{hi}$, the $d_z \times 1$ vector of known Level I covariates for unit i in group h, $\boldsymbol{\beta}_h$ is a $d_z \times 1$ vector of unknown regression parameters for group h, $\boldsymbol{w}_h$ is a $d_w \times 1$ vector of known group-level covariates for group h, the vector $\boldsymbol{\gamma}$ has length $d_z \cdot d_w$, and the parameters $\boldsymbol{\gamma}$ and $\boldsymbol{A}$ are unknown and assigned the prior distribution in (6) for some constant M.

The HGLM in (7)–(9) is a special case of the GLMM because for the HGLM

$$\boldsymbol{\beta}_h = (\boldsymbol{I}_{d_x} \otimes \boldsymbol{w}'_h)\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_h; \;\; \boldsymbol{\varepsilon}_h \overset{\text{ind.}}{\sim} N_{d_z}(\mathbf{0}, \boldsymbol{A}). \tag{10}$$

With this,

$$\boldsymbol{\eta}_h = \boldsymbol{Z}_h[(\boldsymbol{I}_{d_x} \otimes \boldsymbol{w}'_h)\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_h] = \boldsymbol{Z}_h(\boldsymbol{I}_{d_x} \otimes \boldsymbol{w}'_h)\boldsymbol{\gamma} + \boldsymbol{Z}_h \boldsymbol{\varepsilon}_h \tag{11}$$

so the HGLM can be written in GLMM notation with

$$\boldsymbol{U}_h = \boldsymbol{Z}_h(\boldsymbol{I}_{d_x} \otimes \boldsymbol{w}'_h) \text{ and } \boldsymbol{X}_h = \boldsymbol{Z}_h. \tag{12}$$

In the case in which the Level II mean is equal to $\boldsymbol{\gamma}$, $\boldsymbol{U}_h = \boldsymbol{X}_h = \boldsymbol{Z}_h$. A slightly different representation of the HGLM facilitates the proof of Theorem 4 in Section 6.

For GLMMs and HGLMs, the results here address models for which an *exponentiated norm bound* (ENB) holds for the Level I likelihood for $k \leq k_0$ groups

$$L(\boldsymbol{\eta}|\boldsymbol{y}) \leq c_0 \exp(-c_1 \|\boldsymbol{\eta}\|) \tag{13}$$

for constants $c_0, c_1 > 0$. Our results address priors with the general prior structure in (6) or (9) and with the following bound

$$p(\boldsymbol{\gamma}, \boldsymbol{A}) d\boldsymbol{\gamma} d\boldsymbol{A} \equiv \pi(\boldsymbol{A}) d\boldsymbol{\gamma} d\boldsymbol{A} \leq M \frac{|\boldsymbol{A}|^a}{|\boldsymbol{I}_{d_x} + \boldsymbol{A}|^{a_0}} d\boldsymbol{\gamma} d\boldsymbol{A} \tag{14}$$

where the constant $M < \infty$. As detailed in Section 2, an ENB exists for a log-concave likelihood with all MLEs in a bounded set. Examples of log-concave likelihoods include exponential family models with natural links and some links that are not the natural link. The priors on $\boldsymbol{A}$ in (14) are proper if and only if $a > -1$ and $a_0 - a > d_x$ and are then called multivariate F distributions. Otherwise, when they are improper, we refer to them as improper multivariate F distributions. These priors are special cases of Condition 1 in Berger et al. (2005) when their parameter $\ell = 1$.

The difficulty of assuring posterior propriety for these types of models inspired our development of the main theoretical results. In particular, there were no easy-to-check conditions for posterior propriety of hierarchical logistic regression models. For convenience, our results for exponential family models, including GLMs, HGLMs, and GLMMs, are restricted to known convolution parameters ("dispersion" parameter in the language of McCullagh and Nelder (1996)), with unknown convolution parameters discussed in Section 9 Remark 1. Table 1 summarizes our main results, illustrating

Model	Prior Structure	Parameter that Must Have ENB	Required # of Groups with ENBs	Matrices Required to Be of Full Rank	Theorem
GLMM	Equation (14)	$\boldsymbol{\eta}$	$k > 2d_x + d_u + 2(a - a_0)$	Fixed Effects $\boldsymbol{U}$, Random Effects $\boldsymbol{X}$	3
HGLM	Equation (14)	$\boldsymbol{\eta}$, equivalently $\boldsymbol{\beta}$ or $\boldsymbol{\beta}^*$ (Equation (20))	$k > 2d_z + d_w + 2(a - a_0)$	Level I Predictors $\boldsymbol{Z}$, Level II Predictors $\boldsymbol{W}$	4
GLM	Bounded Prior on Regression Coefficients	$\boldsymbol{\eta}$	–	Predictors $\boldsymbol{U}$	1

Table 1: Summary of Conditions for Posterior Propriety.

GLMMs ($k_{min} = 2d_x + d_u + 1$)						
Citation	Model	link	d_x	d_u	k_{min}	k_0
Zeger and Karim (1991)	Bernoulli random intercept	logit	1	8	11	250
Diggle et al. (1994)	Poisson GLMM	log	2	4	9	30
Natarajan and Kass (2000)	Bernoulli GLMM	logit	2	4	9	30 or 50
Sinharay and Stern (2002)	Bernoulli random intercept	probit	1	2	5	31
Sinha (2004)	Bernoulli random intercept	logit	1	6	9	38
HGLMs ($k_{min} = 2d_z + d_w + 1$)						
Citation	Model	link	d_z	d_w	k_{min}	k_0
Wong and Mason (1985)	hierarchical logistic regression	logit	3	4	11	15
Aitchison and Ho (1989)	multivariate Poisson–Log Normal	log	2	1	6	100
Everson and Morris (2000)	bivariate Normal hierarchical model	identity	2	2	7	27
Michalak (2001)	hierarchical logistic regression	logit	4	1	10	26
Bronskill et al. (2002)	hierarchical logistic regression	probit	3	1	8	22

Table 2: Number of Groups That Must Have an ENB, k_{min}, When $a = a_0 = 0$ for GLMMs and HGLMs (Theorems 3 and 4).

their ease of application to a wide variety of models, and Table 2 applies them to several models from the literature, with details of their application provided in Section 8. The results in Table 2 demonstrate that the required number of groups with an ENB, k_{min}, is modest and typically much less than the number of groups being jointly modeled, k_0.

Our main contributions are as follows. An exponentiated norm bound for the Level I likelihood (13) provides an easy-to-work-with unified approach to posterior propriety and to finiteness of posterior moments and MGFs. The ENBs emphasized here enable a unified treatment for many likelihoods, including those with the structure specified in (2)–(4), i.e., GLMs, GLMMs, and HGLMs. Hence, the results here encompass certain results from Natarajan and Kass (2000), in which the random effects are a subset of the fixed effects, and certain results from Chen et al. (2002) that require independence of the fixed effects and the random effects.

The structure of this paper is as follows. Section 2 offers conditions under which the existence of an MLE implies that an ENB holds for GLMs, GLMMs, HGLMs, and other models. Section 3 reviews related work. Section 4 presents results for GLMs, other one-level models, and a general mixed model, all with ENBs. With the ENB satisfied and using a prior density bounded by an improper multivariate F prior density in (14), Sections 5 and 6 provide easily-verified sufficient conditions for posterior propriety for

GLMMs and HGLMs that are based on dimension counts and matrix ranks. Section 7 contains results for posterior moments and MGFs. Referencing Table 2, Section 8 discusses the application of our results to certain models from the literature. Section 9 presents remarks pertaining to and extensions and generalizations of this work, and Section 10 offers conclusions. Appendix A presents Theorem 6, which describes conditions that lead to ENBs for a variety of Level I likelihoods, and related discussion. Appendix B provides terminology, notation, and results that are used to develop the results here, and Appendix C contains proofs of the theorems in this paper.

2 Exponentiated Norm Bounds

All of our results involve verifying that an ENB exists for the likelihood of $\boldsymbol{\eta}$, where an ENB is defined to hold if constants $c_0, c_1 > 0$ exist such that

$$L(\boldsymbol{\eta}|\boldsymbol{y}) \leq c_0 \exp(-c_1\|\boldsymbol{\eta}\|). \tag{15}$$

The constants c_0 and c_1 can be chosen independently of the L_p norm, $p \geq 1$, because of *norm equivalence*, where two norms L_p and L_q on $\mathbb{R}^r$ are said to be norm-equivalent if and only if there exist constants $0 < c_2, c_3$ such that $c_2\|\boldsymbol{v}\|_p \leq \|\boldsymbol{v}\|_q \leq c_3\|\boldsymbol{v}\|_p$ for any vector $\boldsymbol{v}$ (Appendix B). While c_0 and c_1 in (15) cannot depend on $\boldsymbol{\eta}$, they may depend on any known values including $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{U}$, and the L_p norm, etc. Note that independence in (1) is unnecessary if an ENB holds in (15). Of course, (15) holds if the following equation (16), which can be easier to verify, holds

$$L(\boldsymbol{\eta}_h|\boldsymbol{y}_h) \leq c_{0h} \exp(-c_{1h}\|\boldsymbol{\eta}_h\|),\ h = 1, \ldots, k. \tag{16}$$

The bound in (15) or (16) may not hold for all k_0 groups being modeled jointly; thus the groups with an ENB are indexed from 1 to $k \leq k_0$, i.e., k is the number of groups for which (15) holds. Any additional groups without an ENB can only add information. If posterior propriety holds for the k groups with an ENB, then assuming no new parameters are added to the model it also must hold for all $k_0 \geq k$ groups. Remark 7 in Section 9 provides details.

Use of the ENB for the Level I likelihood as defined in (15) enabled the development of the results here, which cover a broad class of models. These results leverage ENBs for Level I likelihoods to provide relatively simple sufficient conditions for verifying the propriety of posterior distributions and of certain posterior moments in multi-level generalized linear models, including HGLMs and GLMMs.

The ENB condition is quite general and may hold for bounded likelihoods that are non-differentiable, discontinuous, non-concave, or that have other irregular behavior, but have tails that are bounded by an ENB.

More specifically, Theorem 6 in Appendix A proves that if a likelihood as a function of $\boldsymbol{\eta}$ is log-concave and the MLE of $\boldsymbol{\eta}$ exists and is unique (or more broadly if $\boldsymbol{\eta}$ has multiple MLEs, all MLEs lie a bounded set), then the likelihood has an ENB as a function of $\boldsymbol{\eta}$.

GLMs with natural links are log-concave. Thus, the likelihood function $L(\boldsymbol{\eta}|\boldsymbol{y})$ for a GLM with natural link and a finite MLE has an ENB as a function of $\boldsymbol{\eta}$. More generally, given a GLMM (or other model) with a log-concave likelihood, it has an ENB if the MLE of $\boldsymbol{\eta}$ exists. When the observed data follow a Bernoulli distribution, Pratt (1981), using a latent-variable response function approach, provides examples of link distributions, i.e., response functions, that lead to log-concavity of the likelihood. These include the Normal; Logistic; Sine; Extreme-Value; the Gamma, Weibull, and Pareto, all when the shape parameter is at least 1; the Beta when both parameters are at least 1; and the t-distribution. With binary data, finiteness of the MLE may also be verified using a separating hyperplane condition (Albert and Anderson, 1984; Natarajan and Kass, 2000; Chen and Shao, 2001; Roy and Hobert, 2007; Lin et al., 2013), but that requires the concatenated $[\boldsymbol{U}, \boldsymbol{X}]$ matrix to be of full rank, a condition that is not required for the posterior propriety results here.

In addition to Bernoulli data, our results apply to other exponential family models, including the Normal, Gamma, Poisson, and Binomial. These exponential family models can also have log-concave likelihoods when a link function other than the natural link is used. When the observed data follow a Gamma distribution, the natural link is the reciprocal, but the more commonly-used log link also leads to a log-concave likelihood. Wedderburn (1976) details that the Poisson likelihood is log-concave with certain power links and with the identity link given certain conditions and that the Binomial likelihood is log-concave when the arcsin, complementary log-log, or probit link is used. Thus, both the natural link functions and a number of other commonly-used link functions lead to log-concave likelihoods. Other examples of log-concavity include Laplace's double exponential distribution, which is not an exponential family (Spiegelhalter et al., 1995).

Wedderburn (1976) details when MLEs of the regression coefficients are finite, unique, and/or exist in the interior of the parameter space for Binomial, Poisson, Normal, and Gamma GLMs. The results in Wedderburn (1976) can be applied to the GLMM Level I model in (1) or (2) by using them to verify whether the MLE of $(\boldsymbol{\gamma}', \boldsymbol{\varepsilon}_h')'$ in (1) for a particular group h is finite or whether the MLE of $(\boldsymbol{\gamma}', \boldsymbol{\varepsilon}')'$ in (2) for a set of groups is finite. Statistical software that fits GLMs may also be helpful in determining when $\boldsymbol{\eta}$ has a finite MLE.

The existence of an ENB may depend on the data. For example, suppose a binary logistic regression is used to assess hospital patient outcomes. If all male patients have one outcome and all female patients have the other outcome and an indicator for gender is included in the model, then an ENB will not exist. Similarly, if all of the patients have the same outcome, then an ENB will not exist.

The results in Wedderburn (1976) and those involving the separating hyperplane condition for existence of an MLE with binary data (Albert and Anderson, 1984; Natarajan and Kass, 2000; Chen and Shao, 2001; Roy and Hobert, 2007; Lin et al., 2013) assume that the design matrix is of full rank. In the case of a GLMM this condition means that the concatenated $[\boldsymbol{U}, \boldsymbol{X}]$ matrix would need to be of full rank. The main posterior propriety results for GLMMs in Chen et al. (2002), with the exception of a result addressing a proper inverse Wishart prior on the Level II variance matrix, also assume that a subset of the rows in the concatenated $[\boldsymbol{U}, \boldsymbol{X}]$ matrix is of full rank. In contrast, our results do not require $[\boldsymbol{U}, \boldsymbol{X}]$ to be of full rank because $\boldsymbol{\eta}$ can have a finite MLE even when $[\boldsymbol{U}, \boldsymbol{X}]$ is not of full rank. As an example, both $\boldsymbol{U}$ and $\boldsymbol{X}$ may include an intercept term. In this case, without prior information on the components in $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ that correspond to the intercept terms, the intercept term parameters are non-identifiable. However, the sum of the intercept term parameters in $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ is estimable and hence $\boldsymbol{\eta}$ is estimable and can have a finite MLE. Similarly, an HGLM with no Level II predictors other than an intercept has $\boldsymbol{Z}_h = \boldsymbol{U}_h = \boldsymbol{X}_h$ so that $[\boldsymbol{U}, \boldsymbol{X}]$ does not have full rank, a case which our posterior propriety results also permit.

3 Related Work

Other work has considered posterior propriety for these and related models. Berger et al. (2005) lists various priors that provide posterior propriety for Normally-distributed data, assuming certain HGLM models. While our results for Normal and for non-Normal Level I data include many often-used improper hyperpriors, our priors are less general in the Normal case than those of Berger et al. (2005). However, our Theorem 4 provides posterior propriety results for multivariate Normal data and for certain non-Normal data when HGLMs have likelihoods with ENBs and when familiar improper multivariate F distributions, also used in Berger et al. (2005)'s Corollary 2.10(a), are specified. In so doing, our work provides results for non-Normal data that Berger et al. (2005) does not address. It is plausible that prior distributions that yield posterior propriety for Normal data also will enjoy posterior propriety when transported to other HGLMs that have ENBs. In providing necessary and sufficient conditions for the Normal case, Berger et al. (2005) shows that our sufficient conditions (Theorem 4) also are necessary, at least when applied to Normal observations at Level I.

Chen et al. (2002) presents broad conditions for posterior propriety of GLMMs in which, as here, the data follow an exponential family with a flat prior on the fixed effects and Normal distributions for the random effects. That work provides sufficient conditions and also provides necessary conditions for posterior propriety with certain improper priors on the random effects covariance matrix. For certain binary GLMMs, their Theorem 3.4 provides necessary and sufficient conditions broader than those here. However, their results for HGLMs, like those needed for the VA data, and for GLMMs with collinearity between the random effects matrix ($\boldsymbol{X}$) and the fixed effects matrix ($\boldsymbol{U}$), are limited to proper Wishart priors for the inverse of the random effects covariance matrix.

Natarajan and Kass (2000) presents a sufficient condition for posterior propriety for GLMMs in which the random effects are a subset of the fixed effects, a case the results here cover, as well as a necessary condition and a sufficient condition for posterior propriety for Bernoulli data. Their conditions sometimes require complicated integral evaluations.

Our use of an ENB permits unification of certain results from Chen et al. (2002) and Natarajan and Kass (2000). Specifically, an ENB approach provides a single theory and resulting posterior propriety conditions that apply both when the random effects are a subset of the fixed effects (Natarajan and Kass, 2000) and when the random effects are linearly independent of the fixed effects (Chen et al., 2002).

Sun et al. (2001) provides posterior propriety conditions for a class of GLMMs with a prior specification designed for spatial modeling. Kim et al. (2008) and Roy and Dey (2014) discuss posterior propriety for certain models with unknown link function parameters.

For Bayesian spline estimators in logistic regression and logistic GLMMs, Raghavan and Cox (1998) discusses equivalence of the existence of the MLE and posterior propriety, as we do in Section 4, using a bound for the fixed effects akin to an ENB. Speckman et al. (2009) presents related results about the equivalence of MLE existence and posterior propriety for multinomial logistic and probit choice models.

Results about posterior propriety for GLMs include those in Ibrahim and Laud (1991) (GLMs with Jeffreys priors); Dey et al. (1997) (overdispersed GLMs); Gelfand and Sahu (1999) (identifiability, posterior propriety, and Gibbs sampling for GLMs); Ghosh et al. (1999) (spatial GLMs); Garvan and Ghosh (1999) (two parameter dispersion models including some GLMs); and Chen et al. (2004) (regression models with missing covariates).

Related posterior moment conditions for non-hierarchical models include those of Ibrahim and Laud (1991) (GLMs), Kim and Ibrahim (2000) (Weibull and extreme value regression models), and Chen and Shao (2001) (dichotomous quantal response models). In the hierarchical setting, Yang and Chen (1995) establish conditions for the existence of certain posterior moments in multivariate Normal hierarchical models.

Finally, Dey et al. (1997) uses an intersecting-hyperplanes approach to bound log-concave likelihoods.

4 Posterior Propriety of One-Level Models with ENB Likelihoods, GLMs, and a General Class of Mixed Models

This section presents results for one-level models with ENB likelihoods, such as GLMs, and for a general class of hierarchical models. Theorem 1 addresses certain single-level models with ENBs, e.g., GLMs, that have flat priors on the regression coefficients. This theorem implies posterior propriety for any one-level model with an ENB for the unknown parameters when they are assigned a bounded prior distribution.

Theorem 1. *For fixed $\boldsymbol{y}$ assume a likelihood $L(\boldsymbol{\eta}|\boldsymbol{y})$ with an ENB for $\boldsymbol{\eta}$ as in (15), where $\boldsymbol{\eta} = \boldsymbol{U}\boldsymbol{\gamma}$, $\boldsymbol{U}$ $(N \times d_u)$ has full rank d_u, and the prior density for $\boldsymbol{\gamma}$ is bounded, i.e., $p(\boldsymbol{\gamma}) \leq M < \infty$. Then, the posterior distributions of $\boldsymbol{\gamma}$ and of $\boldsymbol{\eta}$ are proper and $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ have proper posterior MGFs.*

For the models described in Theorem 1, an ENB implies both finite MLE(s) and posterior propriety, e.g. for GLMs with natural links and known convolution parameters.

Theorem 1 holds for GLMs when $\boldsymbol{\eta}$ is the natural parameter if the MLE is unique (or more generally if multiple MLEs are uniformly bounded) because the likelihood is log-concave and therefore with a unique MLE will have an ENB. However, neither the natural link nor exponential families are required for log-concavity or for an ENB to hold; see Section 2.

Theorem 2 provides a result for a general class of models in which the likelihood, as a function of $\boldsymbol{\eta} \equiv \boldsymbol{U}\boldsymbol{\gamma} + \boldsymbol{X}\boldsymbol{\varepsilon}$, has an ENB as in (15), $\boldsymbol{\varepsilon} \sim N_K(\mathbf{0}, \boldsymbol{\Lambda})$ with $\boldsymbol{\Lambda}$ fixed and known, and $\boldsymbol{\gamma}$ is Lebesgue distributed on $\mathbb{R}^r$, $r \equiv d_u$.

This theorem builds on Theorem 1 by showing that if $\boldsymbol{\Lambda}$ is known, or if it has a proper prior distribution, then the posterior distribution is proper even when $[\boldsymbol{U}, \boldsymbol{X}]$ is not of full rank. Result 3 in Theorem 2 is crucial to the proofs of posterior propriety results for GLMMs (Theorem 3).

Theorem 2. *Let $\boldsymbol{\eta} \equiv \boldsymbol{U}\boldsymbol{\gamma} + \boldsymbol{X}\boldsymbol{\varepsilon}$ in N-dimensions have a likelihood function $L(\boldsymbol{\eta}|\boldsymbol{y})$ with an ENB, i.e., for some $c_0, c > 0$, $L(\boldsymbol{\eta}|\boldsymbol{y}) \leq c_0(\exp(-c\|\boldsymbol{\eta}\|))$. Matrices $\boldsymbol{U}$ $(N \times d_u)$ of full rank $r \equiv d_u \leq N$ and $\boldsymbol{X}$ $(N \times K)$ not necessarily of full rank are known. Assume $\boldsymbol{\gamma}$ has an (improper) Lebesgue distribution on $\mathbb{R}^r$ and that, given $\boldsymbol{\Lambda}$ (not necessarily of full rank), $\boldsymbol{\varepsilon} \sim N_K(\mathbf{0}, \boldsymbol{\Lambda})$.*

1. *Let $\boldsymbol{Q} \equiv \boldsymbol{I}_N - \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'$, the projection matrix orthogonal to $\boldsymbol{U}$. Then for any given $\boldsymbol{\Lambda}$ there exist constants $c_2, c_3 > 0$ (not depending on c or $\boldsymbol{\Lambda}$) such that*
$$E(L(\boldsymbol{\eta}|\boldsymbol{y})) \leq c_0 E(\exp(-c\|\boldsymbol{\eta}\|)) \leq (c_2/c^r) \times |\boldsymbol{I}_K + c_3 c^2 \boldsymbol{X}'\boldsymbol{Q}\boldsymbol{X}\boldsymbol{\Lambda}|^{-1/2}. \qquad (17)$$

2. *If $\boldsymbol{\Lambda}$ is known or has a proper prior distribution, then the posterior distribution of $(\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\Lambda})$ is proper.*

3. *If the prior distribution on $\boldsymbol{\Lambda}$ is improper, then the posterior distribution of $(\boldsymbol{\eta}, \boldsymbol{\gamma}, \boldsymbol{\Lambda})$ is proper if the expectation of the right side of (17) is finite.*

The dependency of the terms c_2/c^r and $c_3 c^2$ on c is included to address cases in which a convolution parameter exists and is unknown, with Remark 1 in Section 9 providing details. The constants c_2 and c_3 (and c_4 in Theorem 3) can depend on all known values including c_0 (but not c because c can absorb an unknown convolution parameter), all dimension counts, and the data $\boldsymbol{y}$, $\boldsymbol{X}$, and $\boldsymbol{U}$.

Theorem 2 applies in two special cases: (i) when $\boldsymbol{\gamma} = \mathbf{0}$ and $r = 0$ so that $\boldsymbol{Q} = \boldsymbol{I}_N$ and (ii) when $\boldsymbol{U}$ is square so that $\boldsymbol{Q}$ is a matrix of 0s.

Since $\boldsymbol{\Lambda}$ is arbitrary in Theorem 2, the outcomes for different groups need not be independent, allowing for a more general variance structure in Theorem 2 than for the typical GLMM, in which $\boldsymbol{\Lambda} = \boldsymbol{I}_k \otimes \boldsymbol{A}$.

5 Posterior Propriety of GLMMs

This section addresses widely-used GLMMs with exchangeability in Level II in which $\boldsymbol{\Lambda} = \boldsymbol{I}_k \otimes \boldsymbol{A}$ and $K \equiv k \cdot d_x$.

Theorem 3 applies to the GLMMs in (1), (5), and (6) when the data in (1) follow an exponential family with known convolution parameter, an ENB (13) for the Level I likelihood exists, and $\boldsymbol{U}$ is of full rank. More generally, it also applies when the Level 1 likelihood for a non-exponential family distribution has an ENB. Result 3 provides easy-to-check results for the prior in (14).

Theorem 3. *Assume the model in (1), (5), and (6) with $\boldsymbol{\Lambda} = \boldsymbol{I}_k \otimes \boldsymbol{A}$ ($\boldsymbol{A}$ is $d_x \times d_x$), that an ENB for the likelihood of $\boldsymbol{\eta}$ for k groups exists so that (17) holds with $c > 0$ (given $\boldsymbol{A}$), for Results 1–3 that $\boldsymbol{U}$ in (2) has full column rank (which implies $k \geq d_u$), and for Results 2 and 3 that $\boldsymbol{X}$ in (2) also has full column rank.*

1. *If $\pi(\boldsymbol{A})$ is proper, then the posterior distribution is proper.*

2. *Let $\pi(\boldsymbol{A})$ be improper and $k > d_u$. If there exists a finite $c_4 > 0$ (not depending on c) such that*

$$\int (|\boldsymbol{I}_{d_x} + c_4 c^2 \boldsymbol{A}|)^{-\frac{k-d_u}{2}} \pi(\boldsymbol{A}) d\boldsymbol{A} < \infty, \tag{18}$$

   then the posterior distribution on $(\boldsymbol{\gamma}, \boldsymbol{A})$ is proper.

3. *Let $k > d_u$ with the joint prior distribution on $\boldsymbol{\gamma}$ and $\boldsymbol{A}$ specified as in (14), i.e., $p(\boldsymbol{\gamma}, \boldsymbol{A}) \leq M \frac{|\boldsymbol{A}|^a}{|\boldsymbol{I}_{d_z} + \boldsymbol{A}|^{a_0}}$, and the convolution parameter known. If the prior distribution on $\boldsymbol{A}$ is proper, the posterior distribution is proper by Result 1. If the prior distribution is improper as a function of $\boldsymbol{A}$, then the posterior distribution is proper if $a > -1$ and $k > 2d_x + d_u + 2(a - a_0)$, i.e., the integer k must be at least k_{min}:*

$$k_{min} \equiv \begin{cases} 2d_x + d_u + 2(a - a_0) + 1 & \text{if } (2d_x + d_u + 2(a - a_0)) \in \mathbb{Z}, \\ \lceil 2d_x + d_u + 2(a - a_0) \rceil & \text{otherwise} \end{cases} \tag{19}$$

   where $\lceil x \rceil$ is the ceiling of x.

With Result 3 and given a and a_0, the k_{min} for posterior propriety requires one additional ENB group for each additional fixed effect that is added to the model (augmenting $\boldsymbol{U}_h$, equivalently $\boldsymbol{U}$, by one column) and two additional ENB groups for each random effect that is added to the model (augmenting $\boldsymbol{X}_h$ by one column).

Theorem 3 also applies when $\boldsymbol{U}$ is partially or fully collinear with $\boldsymbol{X}$, with the proof of the special case $\boldsymbol{U} = \boldsymbol{X}\boldsymbol{M}$ with $\boldsymbol{\eta} = \boldsymbol{X}(\boldsymbol{M}\boldsymbol{\gamma} + \boldsymbol{\varepsilon})$ instructive.

Sharper bounds may be obtained for patterned $\boldsymbol{A}$ matrices with fewer unknown parameters, e.g., if $\boldsymbol{A}$ were an equi-correlation matrix. Specifically, Lemma 5, referenced in the proof of Theorem 3, provides a lower bound (generally not sharp) for certain determinants that include projections. See Remark 5 in Section 9.

6 Posterior Propriety of HGLMs

This section presents posterior propriety results for HGLMs with the structure in (7)–(9).

In (7) $\boldsymbol{\eta}_h = \boldsymbol{Z}_h \boldsymbol{\beta}_h$, leading to $\boldsymbol{\eta} = \boldsymbol{Z}\boldsymbol{\beta}$, where $\boldsymbol{Z}$ is the block diagonal matrix with $\boldsymbol{Z}_h$ in the hth position and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_k')'$. As in the proof of Theorem 1, if $\boldsymbol{Z}$ is of full rank, it is equivalent to prove finiteness of the posterior distribution as a function of $\boldsymbol{\eta}$ or as a function of $\boldsymbol{\beta}$. This holds because with $\boldsymbol{Z}$ of full rank, $\boldsymbol{\eta}$ and $\boldsymbol{\beta}$ are norm equivalent (Section 2 and Appendix B define norm equivalence).

It facilitates the proof of Theorem 4 to reorder the elements of $\boldsymbol{\beta}$ so that $\boldsymbol{\beta}^* \equiv (\boldsymbol{\beta}_{11}, \boldsymbol{\beta}_{21}, \ldots, \boldsymbol{\beta}_{k1}, \ldots, \boldsymbol{\beta}_{1d_z}, \boldsymbol{\beta}_{2d_z}, \ldots, \boldsymbol{\beta}_{kd_z})'$, where $\boldsymbol{\beta}_{hj}$ is the jth element of $\boldsymbol{\beta}_h$. The distribution on $\boldsymbol{\beta}^*$ is given by

$$\boldsymbol{\beta}^* = (\boldsymbol{I}_{d_z} \otimes \boldsymbol{W})\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^* \tag{20}$$

where $\boldsymbol{\gamma}$ follows Lebesgue measure in $d_z \cdot d_w$ dimensions,

$$\boldsymbol{\varepsilon}^* = (\boldsymbol{\varepsilon}_{11}, \boldsymbol{\varepsilon}_{21}, \ldots, \boldsymbol{\varepsilon}_{k1}, \ldots, \boldsymbol{\varepsilon}_{1d_z}, \boldsymbol{\varepsilon}_{2d_z}, \ldots, \boldsymbol{\varepsilon}_{kd_z})' \sim N_{k\cdot d_z}(\boldsymbol{0}, \boldsymbol{A} \otimes \boldsymbol{I}_k), \tag{21}$$

$\boldsymbol{W}$ is the $k \times d_w$ matrix with hth row $\boldsymbol{w}_h'$, and $\boldsymbol{A} \otimes \boldsymbol{I}_k$ is different from $\boldsymbol{I}_k \otimes \boldsymbol{A}$ from the GLMM case when $d_z > 1$. With the notation in (20), $\boldsymbol{I}_{d_z} \otimes \boldsymbol{W}$ and $\boldsymbol{I}_{k\cdot d_z}$ are analogous to $\boldsymbol{U}$ and $\boldsymbol{X}$ in the GLMM, respectively.

The result for GLMMs in Theorem 2 is applied to HGLMs since HGLMs are special cases of GLMMs. However, for the HGLM, a sharper result is obtained because of its special form with $\boldsymbol{U} = \boldsymbol{I}_{d_z} \otimes \boldsymbol{W}$ and $\boldsymbol{X} = \boldsymbol{I}_{k\cdot d_z}$ in (20). It is assumed that the likelihood of $\boldsymbol{\beta}$, equivalently $\boldsymbol{\eta}$, has an ENB (13).

Theorem 4. *Assume the model in (7)–(9), that k groups have an ENB for the likelihood of $\boldsymbol{\beta}$ at some $c > 0$, that $\boldsymbol{Z}$ has full rank, that the joint prior distribution on $\boldsymbol{\gamma}$ and $\boldsymbol{A}$ is proportional to $\pi(\boldsymbol{A})d\boldsymbol{\gamma}d\boldsymbol{A}$, and that $\boldsymbol{W}$ in (20) has full rank $d_w \le k$.*

1. *If $\pi(\boldsymbol{A})$ is proper, then the posterior distribution is proper.*

2. *Let $\pi(\boldsymbol{A})$ be improper and $k > d_w$. If there exists a constant $c_3 > 0$ such that*

$$\int \left(|\boldsymbol{I}_{d_z} + c_3 c^2 \boldsymbol{A}|^{\frac{-(k-d_w)}{2}} \right) \pi(\boldsymbol{A}) d\boldsymbol{A} < \infty, \tag{22}$$

then the posterior distribution on $(\boldsymbol{\gamma}, \boldsymbol{A})$ is proper.

3. *Let $k > d_w$ with the joint prior distribution on $\boldsymbol{\gamma}$ and $\boldsymbol{A}$ specified as in (14), i.e., $p(\boldsymbol{\gamma}, \boldsymbol{A}) \le M \frac{|\boldsymbol{A}|^a}{|\boldsymbol{I}_{d_z} + \boldsymbol{A}|^{a_0}}$, and the convolution parameter known. If the prior distribution on $\boldsymbol{A}$ is proper, the posterior distribution is proper by Result 1. If the prior distribution is improper as a function of $\boldsymbol{A}$, posterior propriety requires $a > -1$ and $k > 2d_z + d_w + 2(a - a_0)$, i.e., the integer k must be at least k_{min}:*

$$k_{min} \equiv \begin{cases} 2d_z + d_w + 2(a - a_0) + 1 & \text{if } (2dz + d_w + 2(a - a_0)) \in \mathbb{Z}, \\ \lceil 2d_z + d_w + 2(a - a_0) \rceil & \text{otherwise.} \end{cases} \tag{23}$$

Similar to Theorem 3 for GLMMs, with Result 3 and given a and a_0, the k_{min} for HGLM posterior propriety requires two additional ENB groups for each Level I covariate that is added to the model (augmenting $\boldsymbol{Z}_h$ by one column) and one additional ENB group for each Level II covariate that is added to the model (augmenting $\boldsymbol{W}$ by one column).

7 Existence of Posterior Moments and MGFs

Theorem 5 provides conditions under which the posterior MGFs of $\boldsymbol{\eta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\varepsilon}$ exist. Level II variances and covariances also are addressed. This theorem includes HGLMs and GLMMs with non-Normal observations that are not covered by the references discussed in Section 3.

Theorem 5. *Assume the model in (1)–(4) with an ENB for some $c > 0$ and with $E(\exp(-c_*||\boldsymbol{\eta}||)) < \infty$ for some $c_* < c$ for at least k_{min} groups.*

1. *Then $\boldsymbol{\eta}$ has a proper posterior distribution and the posterior MGF of $\boldsymbol{\eta}$ exists.*
2. *Assume further that the concatenated $[\boldsymbol{U}, \boldsymbol{X}] \equiv \boldsymbol{C}$ matrix is of full rank. Then the posterior MGF of the corresponding vector $(\boldsymbol{\gamma}', \boldsymbol{\varepsilon}')'$ exists. Also, the trace of $\boldsymbol{\Lambda}$ $(K \times K)$, its eigenvalues, all of its elements, any norm of $\boldsymbol{\Lambda}$, and the generalized variance $|\boldsymbol{\Lambda}|^{1/K}$ have finite posterior MGFs. This means the determinant $|\boldsymbol{\Lambda}|$ has infinitely many moments.*
3. *Under the above conditions including the full rank condition in Result 2, when $\boldsymbol{\Lambda} = \boldsymbol{I}_k \otimes \boldsymbol{A}$, as in the GLMM case, the trace of $\boldsymbol{A}$, its eigenvalues, all of its elements, any norm of $\boldsymbol{A}$, and the generalized variance have proper posterior MGFs. This means the determinant $|\boldsymbol{A}|$ has infinitely many moments.*

In summary, under the conditions for posterior propriety, the posterior MGF of $\boldsymbol{\eta}$, $E(\exp(\boldsymbol{t}'\boldsymbol{\eta})|\boldsymbol{y})$, also exists (Result 1) if there is a value $c_* < c$ for which the expectation of the ENB is finite, whether or not $\boldsymbol{U}$ and $\boldsymbol{X}$ in the GLMM or HGLM case are collinear. Posterior MGFs of $\boldsymbol{\gamma}$, $\boldsymbol{\varepsilon}$, and of all elements of $\boldsymbol{\Lambda}$ or $\boldsymbol{A}$ are also proved to exist (Results 2 and 3) if in addition the concatenation $[\boldsymbol{U}, \boldsymbol{X}]$ has full rank.

8 Examples

Table 1 in Section 1 summarizes our results for posterior propriety. lt is straightforward to use these results to verify propriety of GLMMs discussed in Zeger and Karim (1991), Diggle et al. (1994), Natarajan and Kass (2000), Sinharay and Stern (2002), and Sinha (2004) and of HGLMs presented in Wong and Mason (1985), Aitchison and Ho (1989), Everson and Morris (2000), Michalak (2001), and Bronskill et al. (2002), with Table 2 in Section 1 presenting the value of k_{min} for each of these studies. The results in Table 2 assume that full rank conditions for the matrices of the covariates are met and the prior structure in (14) with $a = a_0 = 0$, although not all of the works in Table 2 use this prior distribution or a prior distribution at all. In all of the studies in Table 2, the value of k_0, the total number of groups being jointly modeled, substantially exceeds k_{min}.

The models in Aitchison and Ho (1989) and Everson and Morris (2000) have a multivariate outcome at Level I. In this case, the scalar entries in the multivariate outcome take the role of units within a group. As discussed in Remark 3 in Section 9, multivariate GLMMs, multivariate HGLMs, and other models with a vector of observations for each unit within a group are also covered by the theorems presented herein.

Everson and Morris (2000) use a bivariate Normal HGLM to model average patient outcomes in $k_0 = 27$ hospitals. In their model, the Level I variance is known, there are two outcomes per hospital and no covariates in Level I so that $\boldsymbol{Z}_h = \boldsymbol{I}_2$, and the Level II matrix of covariates $\boldsymbol{W}$ includes a constant term and a variable capturing within-hospital patient severity so that $d_w = 2$. The prior specified for the 4-dimensional $\boldsymbol{\gamma}$ vector and for $\boldsymbol{A}$ is equivalent to the prior in (14) with $a = a_0 = 0$. With this model, posterior propriety requires $k \geq k_{min} = 2d_z + d_w + 1 = 2 \times 2 + 2 + 1 = 7$ hospitals to have an ENB. For Normal data, the ENB for the Level I likelihood is automatic, but the matrix $\boldsymbol{W}$ must be verified to have full rank and the convolution parameter must be known. Theorem 5 Part 1 then provides that the posterior MGF of $\boldsymbol{\eta}$ exists.

9 Remarks and Extensions

The following remarks underscore extensions and implications of our results.

Remark 1. Unknown Convolution (Dispersion) Parameter.

In certain familiar GLMs, GLMMs, and HGLMs, the convolution parameter ν ("dispersion" parameter in McCullagh and Nelder (1996)) from the Level I density emerges in the ENB in the middle term of (17) as a multiplier of $c||\boldsymbol{\eta}||$. In such cases there may exist $0 < \nu_0 < \nu_1$ (ν_1 possibly infinite) such that for all ν in the interval $[\nu_0, \nu_1)$,

$$L(\boldsymbol{\eta}, \nu|\boldsymbol{y}) \leq c_0 \exp(-c\nu||\boldsymbol{\eta}||). \tag{24}$$

Then by monotonicity and by (17),

$$\begin{aligned} c_0 E\left(\exp(-c\nu||\boldsymbol{\eta}||)\right) &\leq c_0 E\left(\exp(-c\nu_0||\boldsymbol{\eta}||)\right) &(25)\\ &\leq (c_2/(c\nu_0)^r) \times |\boldsymbol{I}_N + c_3(c\nu_0)^2\boldsymbol{Q}\boldsymbol{X}\boldsymbol{\Lambda}\boldsymbol{X}'|^{-1/2}. &(26)\end{aligned}$$

To permit posterior propriety for a given proper prior distribution which is constrained to $[\nu_0, \nu_1)$ and which is independent of the remaining parameters, ν_0 must be large enough for posterior propriety to hold.

Remark 2. ENBs Include Non-Exponential Families and More General Models.

The results here merely require an ENB for the Level I likelihood of the unknown parameters $\boldsymbol{\eta}$ and not necessarily an exponential family in Level I or independence in Level I, as discussed in the context of GLMMs and HGLMs. ENBs can exist for likelihoods that are discontinuous, non-differentiable, or that have other irregularities. Other complicated examples might include cases in which different units or groups follow different Level I distributions and cases with missing data or other complications. In both cases, the required ENB condition must hold and the reasonableness of exchangeability in Level II must be assessed. As discussed in Section A, the likelihood itself need not be log-concave in order to have an ENB. Thus, the results here apply more broadly than to GLMs, GLMMs, and HGLMs.

Remark 3. Generalization to Multivariate Outcomes.

Various multivariate outcomes can be addressed using the models considered here and hence the results herein pertain in certain multivariate cases, e.g., as in Aitchison and Ho (1989) and Everson and Morris (2000). In these two works, which are referenced in Table 2 and Section 8, the observations for different outcomes within a group have an analogous role to the observations for different units within a group.

Further, a multivariate observation may be modeled for each unit, e.g., patient, in a group, e.g., hospital. For example, the vector observation $\boldsymbol{y}_{hi}$ for unit i in group h might be 2×1, where the observations for each unit and the observations for different units within a group may be correlated. In this case, $\boldsymbol{\eta}_{hi} = (\eta_{hi1}, \eta_{hi2})'$ is a 2×1 vector, where it is assumed that $\boldsymbol{\eta}_{hi} = \boldsymbol{\eta}_h$ for all i in each group h. The model might then be specified so that $\boldsymbol{\eta}_h = \boldsymbol{\gamma} + \boldsymbol{\varepsilon}_h$ for all h. Concatenating the $\boldsymbol{\eta}_h$ for k groups to form a vector $\boldsymbol{\eta}$ yields $\boldsymbol{\eta} = (\mathbb{1}_k \otimes \boldsymbol{I}_2)\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N_{k \cdot 2}(\mathbf{0}, \boldsymbol{I}_k \otimes \boldsymbol{A})$ and $\boldsymbol{A}$ is a 2×2 matrix, which is the structure already considered.

Remark 4. GLMMs when $\boldsymbol{U}$ and $\boldsymbol{X}$ Are Fully or Partially Collinear.

Theorems 2, 3, and 5 (Result 1) hold when there is full or partial collinearity between the columns of $\boldsymbol{U}$ and $\boldsymbol{X}$. For example, these results hold in the special case $\boldsymbol{U} = \boldsymbol{X}\boldsymbol{M}$ for $\boldsymbol{M}$ a full-rank matrix with dimension $K \times d_u$. Applied to GLMMs, these results extend several results in Chen et al. (2002) for which non-collinearity between $\boldsymbol{U}$ and $\boldsymbol{X}$ is required in the absence of a proper prior on $\boldsymbol{A}$ (which would assure identifiability).

Remark 5. Structured $\boldsymbol{A}$ Matrix.

Chen et al. (2002) includes conditions that hold when the random effects covariance matrix is structured, e.g., the intra-class correlation model. Theorem 2 and Results 1 and 2 of Theorems 3 and 4 apply to structured covariance matrices when the integral in (18) or (22) is reinterpreted to refer to an integral over the lower-dimensional parameterization of $\boldsymbol{A}$.

Remark 6. Proper Prior for Elements of $\boldsymbol{\gamma}$.

When $\boldsymbol{\gamma}$ is known and equals $\mathbf{0}$ ($\mathbf{0}$, without essential loss of generality), conditions for posterior propriety follow if the conditions of Theorem 2 are met with $r \equiv d_u = 0$. In the GLMM case if $\boldsymbol{\gamma}$ is entirely or partially known, let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2')'$ so that $\boldsymbol{\eta} - \boldsymbol{U}_2\boldsymbol{\gamma}_2 = \boldsymbol{U}_1\boldsymbol{\gamma}_1 + \boldsymbol{X}\boldsymbol{\varepsilon}$ with $\boldsymbol{U} = [\boldsymbol{U}_1, \boldsymbol{U}_2]$ partitioned accordingly, $\boldsymbol{\gamma}_2$ known, and $\boldsymbol{\gamma}_1$ with Lebesgue measure. This model has already been addressed with d_u replaced by the number of columns in $\boldsymbol{U}_1$. Because posterior propriety holds for all fixed $\boldsymbol{\gamma}_2$, it also holds when $\boldsymbol{\gamma}_2$ has a proper prior distribution. A similar result holds for HGLMs.

Remark 7. Additional Groups.

The previous theorems establish conditions for k_{min} groups that guarantee posterior propriety. With posterior propriety established for k groups, it continues to hold with the inclusion of additional groups that may or may not have ENBs or other deficiencies. However, propriety is not guaranteed if the additional groups introduce additional parameters and if the additional parameters do not have a proper prior.

Remark 8. Relationship with Stein Estimation when $a = a_0 = 0$.

Theorems 3 for GLMMs and 4 for HGLMs provide conditions under which using the class of multivariate F priors in (14) leads to proper posterior distributions. Lacking proper prior information and having observed independent Normal data, $\boldsymbol{Y} \sim N_k(\boldsymbol{\beta}, \boldsymbol{I}_k)$, i.e., the celebrated James–Stein setting (James and Stein, 1961), the prior density $1/||\boldsymbol{\beta}||^{(k-2)}$ was demonstrated by Stein to be a good choice (Stein, 1981) for estimating $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ with quadratic loss. This is Stein's "super-harmonic" prior density (SHP), which produces a posterior mean on $\boldsymbol{\beta}$, given $\boldsymbol{Y}$, which is a formal Bayes estimator that is spherically invariant, minimax (uniformly dominating $\boldsymbol{Y}$ as an estimator), and is a closely-related admissible alternative to the James–Stein estimator in this equal variance case. Letting $\boldsymbol{\beta} \sim N_k(\boldsymbol{0}, A\boldsymbol{I}_k)$ with the scalar $A \sim \mathit{Unif}(0, \infty)$, corresponding to $a = a_0 = 0$ when $d_x = 1$, yields the SHP upon marginalization (Morris, 1983a). More generally, choosing $a = a_0 = 0$ for various two-level models for Normal data and computing Bayesian 95% credible intervals for this prior thus far has produced confidence intervals that usually exceed and at worst nearly meet their nominal 95% rates, e.g., Morris (1983b), Everson and Morris (2000), Morris and Tang (2011), and Morris and Lysy (2012).

10 Conclusion

The major results of this work are easy-to-verify conditions for the existence of exponentiated norm bounds for likelihoods and for demonstrating posterior propriety for a broad class of hierarchical models that have ENBs with Normal Level II priors. First, the likelihoods of many common models have ENBs when MLEs exist (Section 2 and Appendix A). ENBs can thus be leveraged to provide posterior propriety, posterior MGFs, and possibly other results for a broad class of Level I models. Second, Theorems 2 and 3 provide conditions for posterior propriety for a broad class of models with ENBs and a partially or fully improper prior specification on the first two moments of the Level II Normal distribution. Third, Theorems 3 for GLMMs and 4 for HGLMs provide easy-to-check dimension-count and matrix rank conditions that suffice for posterior propriety for this flexible and widely-used class of models when a prior density bounded by some multiple of a multivariate F prior in (14), proper or improper, is used. Fourth, Theorem 5 shows that with posterior propriety, posterior MGFs or moments of key model parameters often exist.

The conditions herein are especially useful for data analysis because they merely require determining the ranks of matrices and counting dimensions, such as the number of groups and dimensions of covariate matrices, as needed for determining k_{min}. The only complicated check involves verifying the existence of an ENB for the Level I likelihood. For GLMs, GLMMs, and HGLMs with log-concave likelihoods (which are automatic for natural links) the bound may be verified by ascertaining finiteness of Level I MLEs, although MLE existence is not required for an ENB to hold.

The theorems here emphasize the use of non-informative priors for the parameters of the Level II Normal distribution. Such improper priors are used to provide procedures with good frequency properties.

Appendix A: Results on Exponentiated Norm Bounds (ENBs)

Theorem 6 provides conditions under which ENBs as in (15) exist for many models.

Theorem 6. *For fixed* $\boldsymbol{y}$*, let* $\ell(\boldsymbol{\eta}|\boldsymbol{y})$ *be a log-likelihood that is a function of* $\boldsymbol{\eta}$ *on a convex subset* $C \subseteq \mathbb{R}^N$*, with concave upper bound* $g(\boldsymbol{\eta})$*, where* g *is bounded and all maximizing values of* g *are contained in a bounded set.*

1. *For* $g(\boldsymbol{\eta})$ *given as above and for any* L_p *norm* $||\boldsymbol{\eta}|| \equiv ||\boldsymbol{\eta}||_p$*,* $1 \leq p \leq \infty$*, on* $\mathbb{R}^N$*, constants* c_0 *and* $c_1 > 0$ *exist such that*

$$L(\boldsymbol{\eta}|\boldsymbol{y}) \equiv \exp(\ell(\boldsymbol{\eta}|\boldsymbol{y})) \leq \exp(g(\boldsymbol{\eta})) \leq c_0 \ \exp(-c_1||\boldsymbol{\eta}||). \tag{27}$$

2. *When (27) holds, then for any* $\boldsymbol{t}$ *with* $||\boldsymbol{t}||_2 < c_1/2 \equiv c_2$

$$\exp(\boldsymbol{t}'\boldsymbol{\eta})L(\boldsymbol{\eta}|\boldsymbol{y}) \leq c_0 \ \exp(-c_2||\boldsymbol{\eta}||_2). \tag{28}$$

Proof. Result 1 holds because the negative of the concave function g is convex so that the following Lemma 1 applies with convex $f = -g$. The result then follows because $\ell(\boldsymbol{\eta}|\boldsymbol{y}) \leq g(\boldsymbol{\eta})$.

The proof of (28) uses (27) so that for $||\boldsymbol{t}||_2 \leq c_2 = c_1/2$ and because $|\boldsymbol{t}'\boldsymbol{\eta}| \leq ||\boldsymbol{t}||_2||\boldsymbol{\eta}||_2$,

$$\exp(\boldsymbol{t}'\boldsymbol{\eta})L(\boldsymbol{\eta}|\boldsymbol{y}) \leq \exp(c_2||\boldsymbol{\eta}||_2)c_0 \exp(-c_1||\boldsymbol{\eta}||_2) = c_0 \exp(-c_2||\boldsymbol{\eta}||_2). \tag{29}$$

□

The results above provide ENBs for non-differentiable or discontinuous likelihoods and those with non-concave logarithms. As a corollary, if each of several likelihoods has an ENB, then so does their product. The result in (28) prepares for proving finiteness of MGFs in Sections 4 and 7. Finally, if the right inequality in (27) holds for c_0 and c_1, of course, it also holds for any constant larger than c_0 and any positive constant less than c_1.

Lemma 1 below is used in the proof of Theorem 6.

Lemma 1. *Let* f *be a convex function defined on a convex subset* $C \subseteq \mathbb{R}^N$*, with all of its minima contained in a bounded set. Then* f *is bounded below by a linear function of any* L_p *norm of its argument,* $1 \leq p \leq \infty$*. That is, constants* $c_1 > 0$ *and* c_0^* *exist such that* $f(\boldsymbol{x}) \geq c_0^* + c_1\|\boldsymbol{x}\|$ *for all* $\boldsymbol{x}$*. Further, this result extends to convex proper subsets of* $\mathbb{R}^N$*.*

Proof. For simplicity, but without essential loss of generality, assume that $\mathbf{0}$ minimizes f with $f(\mathbf{0}) = 0$ and that all minimizers $\boldsymbol{x}$ of f have $\|\boldsymbol{x}\| < 1$. By continuity of f (convexity implies continuity) and by the compactness of the set $\{\boldsymbol{x} : \|\boldsymbol{x}\| = 1\}$, $f(\boldsymbol{x})$ has a minimum over the set $\|\boldsymbol{x}\| = 1$. Letting $\boldsymbol{x}_1$ with $\|\boldsymbol{x}_1\| = 1$ be such a minimizer, $c_1 \equiv f(\boldsymbol{x}_1) > 0$. Now consider any $\boldsymbol{x} \in C$ with $\|\boldsymbol{x}\| \geq 1$. Define $q \equiv 1/\|\boldsymbol{x}\|$ so $0 < q \leq 1$, and let $\boldsymbol{x}_2 \equiv q\boldsymbol{x}$. Then $\|\boldsymbol{x}_2\| = 1$ and so $f(\boldsymbol{x}_2) \geq c_1$. By convexity,

$$f(\boldsymbol{x}) = (qf(\boldsymbol{x}) + (1-q)f(\mathbf{0}))/q \geq f(q\boldsymbol{x} + (1-q)\mathbf{0})/q = f(\boldsymbol{x}_2)/q \geq c_1/q = c_1\|\boldsymbol{x}\|. \quad (30)$$

Therefore, with c_1 positive, $f(\boldsymbol{x}) \geq c_1\|\boldsymbol{x}\| \geq c_1\|\boldsymbol{x}\| - c_1$ when $\|\boldsymbol{x}\| \geq 1$. When $\|\boldsymbol{x}\| < 1$, $f(\boldsymbol{x}) \geq 0 \geq c_1\|\boldsymbol{x}\| - c_1$. Taking $c_0^* = -c_1$ proves the result. □

The constants c_0^* and c_1 can be chosen independently of p by letting c_0^* and $c_1 > 0$ be values that work for L_1 since $\|\boldsymbol{\eta}\|_1 \geq \|\boldsymbol{\eta}\|_p$ for all p.

Appendix B: Preliminaries for Posterior Propriety

The following terminology, notation, and results are used repeatedly.

Two norms L_p and L_q on $\mathbb{R}^r$ will be termed "norm-equivalent" if and only if there exist constants $0 < c_2, c_3$ such that $c_2\|\boldsymbol{v}\|_p \leq \|\boldsymbol{v}\|_q \leq c_3\|\boldsymbol{v}\|_p$ for any vector $\boldsymbol{v}$, and then we write $\|\boldsymbol{v}\|_p \cong \|\boldsymbol{v}\|_q$.

The easy-to-prove results in Lemmas 2 and 3 are stated here to streamline language and various proofs. Additional lemmas follow.

Lemma 2.

1. *If $\boldsymbol{a}$ and $\boldsymbol{b}$ are orthogonal, i.e., $\boldsymbol{a}'\boldsymbol{b} = 0$, then for any three norms L_p, L_q, and L_s with $1 \leq p,\ q,\ s \leq \infty$, $\|\boldsymbol{a} + \boldsymbol{b}\|_s \cong \|\boldsymbol{a}\|_p + \|\boldsymbol{b}\|_q$. More generally, $\|\boldsymbol{a} + \boldsymbol{b}\|_s \cong \|\boldsymbol{\Gamma}_1\boldsymbol{a}\|_p + \|\boldsymbol{\Gamma}_2\boldsymbol{b}\|_q$ where $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are orthogonal matrices that need not be mutually orthogonal.*

2. *Let $\boldsymbol{U}$ be an $N \times d_u$ matrix of rank m, $m \leq d_u \equiv r$. Let $\boldsymbol{\gamma} \sim$ Lebesgue, i.e., uniformly, on $\mathbb{R}^r$. Then for any fixed vector $\boldsymbol{v} \in \mathbb{R}^r$, $\boldsymbol{\eta} \equiv \boldsymbol{U}(\boldsymbol{\gamma} + \boldsymbol{v})$ has Lebesgue measure on the m-dimensional range space of $\boldsymbol{U}$.*

Lemma 3. *The MGF of a random vector $\boldsymbol{X} = (X_1, \ldots, X_m)'$ exists iff the MGF of $||\boldsymbol{X}||_p$ exists (any L_p-norm, $p \geq 1$) iff the marginal MGF exists for every X_j ($1 \leq j \leq m$) iff the MGF of $|X_j|$ exists for every X_j.*

Lemma 4.

1. *Let $\{Z_j\}_{j=1}^N$ be iid $N(0,1)$ and $t_1, \ldots, t_N > 0$. Then*

$$E(\exp(-\sum_{j=1}^N t_j|Z_j|)) \leq 2^N E(\exp(-\sum_{j=1}^N t_j^2 Z_j^2)) = 2^N / \prod_{j=1}^N (1 + 2t_j^2)^{1/2}. \quad (31)$$

2. *More generally, if* $\boldsymbol{Y} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ *in* N*-dimensions (*$\boldsymbol{\Sigma}$ *not necessarily of full rank), then for any* $c > 0$

$$E(\exp(-c||\boldsymbol{Y}||_1)) \leq E(\exp(-c||\boldsymbol{Y}||_2)) \leq 2^N/\sqrt{|\boldsymbol{I}_N + (2c^2/N)\boldsymbol{\Sigma}|}. \quad (32)$$

Proof. For Result 1, first assume $N = 1$ and denote Mill's ratio as $M(z) \equiv \Phi(-z)/\phi(z)$, the $N(0, 1)$ CDF divided by its density. As is well-known, $M(z) < 1/z$ for $z > 0$. The identity $E(\exp(-t_1|Z_1|)) = \sqrt{2/\pi}M(t_1)$ follows easily by completing the square in the defining integral and simplifying. Because $E(\exp(-t_1|Z_1|)) \leq 1$ and $M(t_1) < 1/t_1$,

$$E(\exp(-t_1|Z_1|)) \leq \min\left\{(2/\pi)^{1/2}/t_1, 1\right\} \leq 2/\sqrt{1 + 2t_1^2} = 2E(\exp(-t_1^2 Z_1^2)). \quad (33)$$

When $N > 1$, (31) holds since

$$E(\exp(-\sum_{j=1}^{N} t_j|Z_j|)) \leq 2^N \prod_{j=1}^{N} E(\exp(-t_j^2 Z_j^2)) = 2^N / \prod_{j=1}^{N}(1 + 2t_j^2)^{1/2}. \quad (34)$$

To prove Result 2, let $\boldsymbol{D}_t^2$ be the diagonal matrix of eigenvalues of $\boldsymbol{\Sigma}$ ($\boldsymbol{D}_t \equiv \mathrm{diag}(t_1, \ldots, t_N)$), so $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{D}_t^2\boldsymbol{\Gamma}'$ for $\boldsymbol{\Gamma}$ an orthogonal matrix. Then $\boldsymbol{\Gamma}'\boldsymbol{Y} \sim \boldsymbol{D}_t\boldsymbol{Z}$, with $\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{I}_N)$. Since $\|\boldsymbol{Y}\|_2 \leq \|\boldsymbol{Y}\|_1$,

$$\begin{aligned} E(\exp(-\|\boldsymbol{Y}\|_1)) &\leq E(\exp(-\|\boldsymbol{Y}\|_2)) = E(\exp(-\|\boldsymbol{D}_t\boldsymbol{Z}\|_2)) & (35)\\ &\leq E(\exp(-\sum_{j=1}^{N}|Z_j|t_j/\sqrt{N})) \leq 2^N/|\boldsymbol{I}_N + (2/N)\boldsymbol{\Sigma}|^{1/2} & (36) \end{aligned}$$

using (34). This proves (32) for $c = 1$, with other values of $c > 0$ given by replacing $\boldsymbol{Y}$ by $c\boldsymbol{Y}$. □

Lemma 5. *Let* $\boldsymbol{T} \geq 0$ *be a symmetric* $N \times N$ *matrix and* $\boldsymbol{Q}$ *be an* $N \times N$ *projection matrix of rank* s*. Denote* $d_u \equiv N - s$*. Then for all such* $\boldsymbol{Q}$

1. $|\boldsymbol{I}_N + \boldsymbol{Q}\boldsymbol{T}| \geq \prod_{j=1}^{s}(1 + \delta_j)$*, where the* δ_j *are the* s *smallest eigenvalues of* $\boldsymbol{T}$*.*
2. *In the special case when* $\boldsymbol{T}$ *is block diagonal with* $\boldsymbol{A}$ *appearing* k *times (*$\boldsymbol{A} \geq 0$*,* $d_x \times d_x$*) and with zeroes elsewhere,* $|\boldsymbol{I}_N + \boldsymbol{Q}\boldsymbol{T}| \geq |\boldsymbol{I}_{d_x} + \boldsymbol{A}|^{k-d_u}$*.*

Proof. It suffices to prove this for $\boldsymbol{T}$ a diagonal matrix because more generally if $\boldsymbol{T} = \boldsymbol{\Gamma}\boldsymbol{D}\boldsymbol{\Gamma}'$ where $\boldsymbol{\Gamma}$ is orthogonal and $\boldsymbol{D}$ is diagonal then

$$|\boldsymbol{I}_N + \boldsymbol{Q}\boldsymbol{\Gamma}\boldsymbol{D}\boldsymbol{\Gamma}'| = |\boldsymbol{I}_N + \boldsymbol{\Gamma}'\boldsymbol{Q}\boldsymbol{\Gamma}\boldsymbol{D}| = |\boldsymbol{I}_N + \boldsymbol{Q}^*\boldsymbol{D}| \quad (37)$$

where $\boldsymbol{Q}^* \equiv \boldsymbol{\Gamma}'\boldsymbol{Q}\boldsymbol{\Gamma}$ is a projection matrix with the same rank s as $\boldsymbol{Q}$.

Assume without loss of generality that the diagonal elements of $\boldsymbol{D}$ are increasing, i.e., $0 \leq \delta_1 \leq \delta_2 \leq \cdots \leq \delta_N$. For $j = 1, \ldots, N$ define $(m_j)^2 \equiv \max(\delta_s - \delta_j, 0)/(1 + \delta_s)$

and $\boldsymbol{M} \equiv \text{diag}(m_1, \ldots, m_N) \geq 0$. Define $\boldsymbol{D}_0 \leq \boldsymbol{D}$ as $\boldsymbol{D}_0 \equiv \delta_s \boldsymbol{I}_N - (1+\delta_s)\boldsymbol{M}^2 \leq \delta_s \boldsymbol{I}_N$. Write $\boldsymbol{Q}^* = \boldsymbol{G}\boldsymbol{G}'$, where $\boldsymbol{G}$ is orthogonal such that $\boldsymbol{G}'\boldsymbol{G} = \boldsymbol{I}_s$. Then

$$\begin{aligned}
|\boldsymbol{I}_s + \boldsymbol{G}'\boldsymbol{D}\boldsymbol{G}| &\geq |\boldsymbol{I}_s + \boldsymbol{G}'\boldsymbol{D}_0\boldsymbol{G}| && (38)\\
&= |(1+\delta_s)\boldsymbol{I}_s - (1+\delta_s)\boldsymbol{G}'\boldsymbol{M}^2\boldsymbol{G}| && (39)\\
&= (1+\delta_s)^s|\boldsymbol{I}_s - \boldsymbol{G}'\boldsymbol{M}^2\boldsymbol{G}| && (40)\\
&= (1+\delta_s)^s|\boldsymbol{I}_N - \boldsymbol{M}\boldsymbol{G}\boldsymbol{G}'\boldsymbol{M}| && (41)\\
&\geq (1+\delta_s)^s|\boldsymbol{I}_N - \boldsymbol{M}^2| \text{ (because } \boldsymbol{G}\boldsymbol{G}' = \boldsymbol{Q}^* \leq \boldsymbol{I}_N) && (42)\\
&= (1+\delta_s)^s \prod_{j=1}^{N}(1-m_j^2) = \prod_{j=1}^{s}(1+\delta_j), && (43)
\end{aligned}$$

proving Result 1.

Result 2 is trivial (and not useful) if $k \leq d_u$. With $k > d_u$, let $\boldsymbol{T}^*$ $(N \times N)$ have $\boldsymbol{I}_{k-d_u} \otimes \boldsymbol{A}$ ($\boldsymbol{A} \geq 0$, $d_x \times d_x$) in its upper left block and zeroes elsewhere so that $\boldsymbol{A}$ appears d_u fewer times in $\boldsymbol{T}^*$ than it does in $\boldsymbol{T}$. With this, the ordered eigenvalues of $\boldsymbol{T}^*$ are dominated by the ordered eigenvalues of $\boldsymbol{Q}\boldsymbol{T}$ since $\boldsymbol{T}^*$ has zeroes in place of at least the d_u largest eigenvalues of $\boldsymbol{T}$. Then

$$|\boldsymbol{I}_N + \boldsymbol{Q}\boldsymbol{T}| \geq |\boldsymbol{I}_N + \boldsymbol{T}^*| = |\boldsymbol{I}_{d_x} + \boldsymbol{A}|^{k-d_u}, \tag{44}$$

proving Result 2. □

Lemma 6. *Let $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ be diagonalizable $m \times m$ matrices that commute ($\boldsymbol{M}_1\boldsymbol{M}_2 = \boldsymbol{M}_2\boldsymbol{M}_1$), and let $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ be complementary $r \times r$ symmetric orthogonal projection matrices with ranks r_1 and r_2, where $r = r_1 + r_2$. Then*

$$|\boldsymbol{M}_1 \otimes \boldsymbol{P}_1 + \boldsymbol{M}_2 \otimes \boldsymbol{P}_2| = |\boldsymbol{P}_1 \otimes \boldsymbol{M}_1 + \boldsymbol{P}_2 \otimes \boldsymbol{M}_2| = |\boldsymbol{M}_1|^{r_1}|\boldsymbol{M}_2|^{r_2}. \tag{45}$$

Proof. Since $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ can be diagonalized simultaneously (guaranteed by their commutativity; see Horn and Johnson (1985, p. 50)), write $\boldsymbol{M}_j = \boldsymbol{H}\boldsymbol{D}_j\boldsymbol{H}^{-1}$, with $\boldsymbol{D}_j$ diagonal and $\boldsymbol{H}$ non-singular. Then

$$\boldsymbol{M}_j \otimes \boldsymbol{P}_j = (\boldsymbol{H} \otimes \boldsymbol{I}_m) \times (\boldsymbol{D}_j \otimes \boldsymbol{P}_j) \times (\boldsymbol{H}^{-1} \otimes \boldsymbol{I}_m). \tag{46}$$

Insertion into $|\boldsymbol{M}_1 \otimes \boldsymbol{P}_1 + \boldsymbol{M}_2 \otimes \boldsymbol{P}_2|$ and simplification gives

$$|\boldsymbol{M}_1 \otimes \boldsymbol{P}_1 + \boldsymbol{M}_2 \otimes \boldsymbol{P}_2| = |\boldsymbol{D}_1 \otimes \boldsymbol{P}_1 + \boldsymbol{D}_2 \otimes \boldsymbol{P}_2|. \tag{47}$$

Use these steps again, now for $\boldsymbol{P}_j = \boldsymbol{G}\boldsymbol{I}_j^*\boldsymbol{G}'$, $\boldsymbol{G}$ orthogonal, $\boldsymbol{I}_j^*$ a diagonal projection matrix with r_j unit diagonal entries and zeroes elsewhere so that $\boldsymbol{I}_1^* + \boldsymbol{I}_2^* = \boldsymbol{I}_r$. Simplifying further gives

$$|\boldsymbol{M}_1 \otimes \boldsymbol{P}_1 + \boldsymbol{M}_2 \otimes \boldsymbol{P}_2| = |\boldsymbol{D}_1 \otimes \boldsymbol{I}_1^* + \boldsymbol{D}_2 \otimes \boldsymbol{I}_2^*| = |\boldsymbol{D}_1|^{r_1}|\boldsymbol{D}_2|^{r_2} = |\boldsymbol{M}_1|^{r_1}|\boldsymbol{M}_2|^{r_2}. \tag{48}$$

The above approach, reapplied to the reversed determinant, gives

$$|\boldsymbol{P}_1 \otimes \boldsymbol{M}_1 + \boldsymbol{P}_2 \otimes \boldsymbol{M}_2| = |\boldsymbol{I}_1^* \otimes \boldsymbol{D}_1 + \boldsymbol{I}_2^* \otimes \boldsymbol{D}_2| = |\boldsymbol{M}_1|^{r_1}|\boldsymbol{M}_2|^{r_2}. \tag{49}$$

□

Appendix C: Proofs of Theorems

This appendix contains proofs of the theorems that appear in the body of this paper.

Proof of Theorem 1. We prove finiteness of the posterior MGF of $\boldsymbol{\eta}$ (or equivalently of $\boldsymbol{\gamma}$), from which posterior propriety follows. From Lemma 3 it suffices to prove finiteness of the MGF of $||\boldsymbol{\eta}||$, or alternatively of $||\boldsymbol{\gamma}||$, because, with $\boldsymbol{U}$ of full rank, $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ are norm-equivalent. Theorem 6 and its proof assure that $c_0, c_3 > 0$ exist for which

$$\exp(\boldsymbol{t}'\boldsymbol{\eta})L(\boldsymbol{\eta}|\boldsymbol{y}) \leq \exp(||\boldsymbol{t}||_2||\boldsymbol{\eta}||_2)L(\boldsymbol{\eta}|\boldsymbol{y}) \leq c_0\exp(-c_2||\boldsymbol{\eta}||_2) \leq c_0\exp(-c_3||\boldsymbol{\gamma}||_1) \quad (50)$$

if $||\boldsymbol{t}||_2$ is near 0 and with c_3 adjusting c_2 (Theorem 6) to account for the norm equivalence of $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ and norm equivalence of the L_2-norm and the L_1-norm. With this

$$\int \exp(\boldsymbol{t}'\boldsymbol{\eta})L(\boldsymbol{\eta}|\boldsymbol{y})d\boldsymbol{\gamma} \leq c_0 \int \cdots \int \exp(-c_3\sum_{i=1}^{d_u}|\gamma_i|)d\gamma_1\cdots d\gamma_{d_u}. \quad (51)$$

The right term is finite because each individual integral in the product of integrals is that of a double exponential distribution. □

Proof of Theorem 2. Because of norm equivalence, it suffices to prove Result 1 for the L_1-norm of $\boldsymbol{\eta}$, $\|\boldsymbol{\eta}\|_1$. Defining $\boldsymbol{P} \equiv \boldsymbol{I}_N - \boldsymbol{Q} = \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'$, write $\boldsymbol{\eta}$ as the sum of two orthogonal components $\boldsymbol{\eta} = \boldsymbol{U}\boldsymbol{\gamma}_\varepsilon + \boldsymbol{QX\varepsilon}$ where $\boldsymbol{\gamma}_\varepsilon \equiv \boldsymbol{\gamma} + (\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{X\varepsilon}$. Since $\boldsymbol{QU} = \boldsymbol{0}$, then, by Lemma 2, $\|\boldsymbol{\eta}\|_1 \geq c_4(\|\boldsymbol{U}\boldsymbol{\gamma}_\varepsilon\|_1 + \|\boldsymbol{QX\varepsilon}\|_2)$, c_4 is determined by norm equivalence, so that for the middle term in (17)

$$E(\exp(-c\|\boldsymbol{\eta}\|_1)) \leq E(\exp(-c\,c_4||\boldsymbol{U}\boldsymbol{\gamma}_\varepsilon||_1))E(\exp(-c\,c_4||\boldsymbol{QX\varepsilon}||_2)). \quad (52)$$

Conditioning on $\boldsymbol{\varepsilon}$ and using Lemma 2 Result 2, $\boldsymbol{\alpha} \equiv \boldsymbol{U}\boldsymbol{\gamma}_\varepsilon$ follows Lebesgue measure in the r-dimensional range space of $\boldsymbol{U}$. Hence, $\alpha_j \sim$ Lebesgue in one dimension, $\|\boldsymbol{\alpha}\|_1 = \sum_{j=1}^r |\alpha_j|$, and $\int_{-\infty}^{\infty}\exp(-c\,c_4|\alpha_j|)d\alpha_j = \frac{2}{c\,c_4}$ for each $j = 1,\ldots,r$. Thus, $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp(-c\,c_4\|\boldsymbol{U}\boldsymbol{\gamma}_\varepsilon\|_1)d\boldsymbol{\gamma}_\varepsilon = (\frac{2}{c\,c_4})^r < \infty$ where $d\boldsymbol{\gamma}_\varepsilon$ refers to Lebesgue measure (by Lemma 2). With this, in the right side of (52) $E(\exp(-c\,c_4||\boldsymbol{U}\boldsymbol{\gamma}_\varepsilon||)) = (\frac{2}{c\,c_4})^r$.

It remains, with $\boldsymbol{\varepsilon} \sim N_K(\boldsymbol{0}, \boldsymbol{\Lambda})$, to show that a positive constant multiple of the determinant term in (17) exceeds $E(\exp(-c\,c_4\|\boldsymbol{QX\varepsilon}\|_2))$. This follows by applying Lemma 4 Result 2 to $\boldsymbol{QX\varepsilon}$ with $\boldsymbol{\Sigma} \equiv \boldsymbol{QX\Lambda X}'\boldsymbol{Q}$ and because $|\boldsymbol{I}_N + (2c_4^2c^2/N)\boldsymbol{QX\Lambda X}'\boldsymbol{Q}| = |\boldsymbol{I}_N + (2c_4^2c^2/N)\boldsymbol{QX\Lambda X}'| = |\boldsymbol{I}_K + (2c_4^2c^2/N)\boldsymbol{X}'\boldsymbol{QX\Lambda}|$. This proves Result 1, upon setting $c_3 = 2c_4^2/N$ and choosing $c_2 = 2^N(2/c_4)^r$.

The posterior distribution is proper if the right side of (17) integrates finitely with respect to the prior distribution on $\boldsymbol{\Lambda}$, proving Result 3. Because the determinant in (17) is bounded above by unity, the right side of (17) always is bounded above by c_2/c^r and the expectation is finite if Λ has a proper prior, yielding Result 2. □

Proof of Theorem 3. When $\pi(\boldsymbol{A})$ is proper, the prior distribution on $\boldsymbol{\Lambda}$ is proper and Result 1 follows from Theorem 2 Result 2.

To prove that (18) in Result 2 yields posterior propriety, assume $k > d_u$ and recall that $\boldsymbol{X}$ is block diagonal $(\boldsymbol{X}_h)$, $h = 1,\ldots,k$, so that $\boldsymbol{X\Lambda X}'$ is block diagonal $(\boldsymbol{X}_h\boldsymbol{A}\boldsymbol{X}_h')$

and the eigenvalues of $\boldsymbol{X\Lambda X'}$ are those separately of the $\boldsymbol{X}_h\boldsymbol{A}\boldsymbol{X}'_h$. Since every $\boldsymbol{X}_h$ is of full column rank d_x, $\boldsymbol{X}'_h\boldsymbol{X}_h \geq c_5\boldsymbol{I}_{d_x}$ for some $c_5 > 0$, for all $h = 1, \ldots, k$. With this, for all h, the non-zero eigenvalues of $\boldsymbol{X}_h\boldsymbol{A}\boldsymbol{X}'_h$ are those of $\boldsymbol{A}^{1/2}\boldsymbol{X}'_h\boldsymbol{X}_h\boldsymbol{A}^{1/2}$, which is at least $c_5\boldsymbol{A}$. Define $\boldsymbol{T}$ as the $N \times N$ block diagonal matrix which has $\boldsymbol{I}_k \otimes \boldsymbol{A}$ in its upper left block and zeroes elsewhere (so that there are $N - K$ zeroes on the main diagonal where $K = k \cdot d_x$). With this, $\boldsymbol{X\Lambda X'}$ and $\boldsymbol{T}$ have the same number of non-zero eigenvalues and the non-zero ordered eigenvalues of $\boldsymbol{X\Lambda X'}$ are greater than or equal to the non-zero ordered eigenvalues of $c_5\boldsymbol{T}$.

The projection $\boldsymbol{Q} = \boldsymbol{I}_N - \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'$ has rank $s \equiv N - d_u > 0$ so that for any $b > 0$,

$$|\boldsymbol{I}_K + b(\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{X})\boldsymbol{\Lambda}| = |\boldsymbol{I}_N + b\boldsymbol{Q}(\boldsymbol{X\Lambda X}')| \geq |\boldsymbol{I}_N + bc_5\boldsymbol{Q}\boldsymbol{T}| \geq \min_{\boldsymbol{Q}^*} |\boldsymbol{I}_N + bc_5\boldsymbol{Q}^*\boldsymbol{T}|, \quad (53)$$

where $\boldsymbol{Q}^*$ is an arbitrary projection with the same rank as $\boldsymbol{Q}$.

Result 2 of Lemma 5 provides a lower bound for the right side of (53). With this,

$$|\boldsymbol{I}_K + b(\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{X})\boldsymbol{\Lambda}| \geq |\boldsymbol{I}_{d_x} + bc_5\boldsymbol{A}|^{k-d_u}. \quad (54)$$

Thus from (17) and (54) and with $b = c_3c^2$ and $c_4 = c_3c_5 > 0$,

$$E(\exp(-c\|\boldsymbol{\eta}\|)) \leq (c_2/c^r)\int(|\boldsymbol{I}_{d_x} + c_4c^2\boldsymbol{A}|^{-\frac{k-d_u}{2}})\pi(\boldsymbol{A})d\boldsymbol{A}. \quad (55)$$

Since an ENB exists for the likelihood for at least one $c > 0$ (sufficiently small), $c_2/c^r > 0$ is finite and (18) holds. This proves Result 2.

Under the conditions of Theorem 3 Result 3 the following integral must be finite:

$$\int \frac{|\boldsymbol{A}|^a}{|\boldsymbol{I}_{d_x} + \boldsymbol{A}|^{\frac{2a_o+k-d_u}{2}}}d\boldsymbol{A}. \quad (56)$$

The form in (56) is that of a multivariate F_{b_1,b_2} distribution,

$$p(\boldsymbol{F})d\boldsymbol{F} \propto \frac{|\boldsymbol{F}|^{\frac{b_1}{2}}}{|\boldsymbol{I}_m + \boldsymbol{F}|^{\frac{b_1+b_2}{2}}}\frac{d\boldsymbol{F}}{|\boldsymbol{F}|^{\frac{1+m}{2}}}, \quad (57)$$

where $\boldsymbol{F}$ is an $m \times m$ symmetric positive definite matrix. This density integrates finitely if and only if $b_1, b_2 > m - 1$. Thus, by comparing lines (56) and (57), $m = d_x$, $b_1 = 2a + d_x + 1$, and $b_2 = 2a_0 + k - d_u - 2a - d_x - 1$ so that posterior propriety requires $a > -1$ and $k - d_u > 2(a - a_0) + 2d_x$. □

Proof of Theorem 4. With $\boldsymbol{Z}$ of full rank and by norm equivalence, it is equivalent to prove finiteness of the posterior distribution as a function of $\boldsymbol{\eta}$ or as a function of $\boldsymbol{\beta}^*$. With this, Result 1 follows from Result 2 of Theorem 2.

For Result 2, apply Theorem 2 to (20) with $\boldsymbol{\eta} = \boldsymbol{\beta}^*$ to find that for some $c > 0$ there exists $c_2, c_3 > 0$ (not depending on c and $\boldsymbol{A}$) such that

$$E(\exp(-c\|\boldsymbol{\beta}^*\|_1)) \leq (c_2/c^r) \times |\boldsymbol{I}_{k\cdot d_z} + c_3c^2(\boldsymbol{I}_{d_z} \otimes \boldsymbol{Q}_W)(\boldsymbol{A} \otimes \boldsymbol{I}_k)|^{-1/2} \quad (58)$$

where $\boldsymbol{Q}_W = \boldsymbol{I}_k - \boldsymbol{P}_W$ and $\boldsymbol{P}_W$ is the projection onto the column space of $\boldsymbol{W}$. Now, with $b = c_3c^2$,

$$\begin{aligned}|\boldsymbol{I}_{d_z\cdot k} + b(\boldsymbol{I}_{d_z}\otimes \boldsymbol{Q}_W)(\boldsymbol{A}\otimes \boldsymbol{I}_k)| &= |\boldsymbol{I}_{d_z\cdot k} + b\boldsymbol{A}\otimes \boldsymbol{Q}_W| && (59)\\ &= |\boldsymbol{I}_{d_z}\otimes(\boldsymbol{P}_W+\boldsymbol{Q}_W) + b\boldsymbol{A}\otimes \boldsymbol{Q}_W| && (60)\\ &= |\boldsymbol{I}_{d_z}\otimes \boldsymbol{P}_W + (\boldsymbol{I}_{d_z}+b\boldsymbol{A})\otimes \boldsymbol{Q}_W| && (61)\\ &= |\boldsymbol{I}_{d_z}+b\boldsymbol{A}|^{k-d_w} && (62)\end{aligned}$$

where the last line holds by Lemma 6. Equation (62) provides the general case upon integrating over $\pi(\boldsymbol{A})$.

For Result 3, the proof of GLMM Theorem 3 provides posterior propriety for the prior in (14) if $k > 2d_z + d_w + 2(a - a_0)$ and if $a > -1$. □

Proof of Theorem 5. Since the ENB holds at c, it holds for all values less than c. Finiteness of the expectation holds for all values greater than c_*. Hence, both hold for all values between c_* and c inclusive.

For Result 1, by Lemma 3 the posterior MGF of $\boldsymbol{\eta}$ exists iff $E(\exp(t||\boldsymbol{\eta}||)) < \infty$ for some $t > 0$ ($t \leq 0$ need not be addressed, by monotonicity). With $p(\boldsymbol{\varepsilon}, \boldsymbol{\gamma}, \boldsymbol{\Lambda})$ the prior density on $\boldsymbol{\varepsilon}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\Lambda}$, the following holds up to a constant multiple

$$\begin{aligned}E(\exp(t||\boldsymbol{\eta}||)) &\leq \int \exp(t||\boldsymbol{\eta}||)\exp(-c||\boldsymbol{\eta}||)p(\boldsymbol{\varepsilon},\boldsymbol{\gamma},\boldsymbol{\Lambda})d\boldsymbol{\varepsilon}d\boldsymbol{\gamma}d\boldsymbol{\Lambda} && (63)\\ &= E(\exp(-(c-t)||\boldsymbol{\eta}||)). && (64)\end{aligned}$$

This is finite if $t > 0$ is chosen so close to 0 that $c - t > c_*$. Therefore, $\boldsymbol{\eta}$ has an MGF.

To prove Result 2, with $\boldsymbol{C} \equiv [\boldsymbol{U}, \boldsymbol{X}]$ of full rank and $(\boldsymbol{\gamma}', \boldsymbol{\varepsilon}')' \equiv \boldsymbol{\phi}$, $||\boldsymbol{\eta}|| = ||\boldsymbol{C}\boldsymbol{\phi}|| \geq b||\boldsymbol{\phi}||$, where b^2 is the smallest eigenvalue of $\boldsymbol{C}'\boldsymbol{C}$. With this and $t > 0$ sufficiently small, $E(\exp(bt||\boldsymbol{\phi}||)) \leq E(\exp(t||\boldsymbol{\eta}||)) < \infty$ so that $\boldsymbol{\phi}$ has an MGF by (63) and (64). Therefore, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\gamma}$ both have MGFs.

For Results 2 and 3 about $\boldsymbol{\Lambda}$ and $\boldsymbol{A}$, since $\boldsymbol{\varepsilon}$ has an MGF, for each $j = 1, \ldots, K$ then $E(\exp(t\varepsilon_j)) < \infty$ for small t near 0. Thus with $\varepsilon_j \sim N(0, \lambda_{jj})$, $E(\exp(t\varepsilon_j)) = E(\exp(t^2\lambda_{jj}/2)) < \infty$, so that λ_{jj} has an MGF for all j. Hence, MGFs exist for $\mathrm{tr}(\boldsymbol{\Lambda}) = \sum_j(\lambda_{jj})$, for all eigenvalues of $\boldsymbol{\Lambda}$, and therefore for $\sum_{i,j}|\lambda_{ij}|$. The generalized variance of $\boldsymbol{\Lambda}$ has an MGF because the geometric mean of $\boldsymbol{\Lambda}$'s eigenvalues is dominated by the arithmetic mean, $\mathrm{tr}(\boldsymbol{\Lambda})/K$. In the special case $\boldsymbol{\Lambda} = \boldsymbol{I}_k \otimes \boldsymbol{A}$ (GLMMs), these moment results are inherited by $\boldsymbol{A}$. □

References

Aitchison, J. and Ho, C. (1989). "The Multivariate Poisson–Log Normal Distribution." *Biometrika*, 76(4): 643–653. MR1041409. doi: http://dx.doi.org/10.1093/biomet/76.4.643. 549, 557, 558, 559

Albert, A. and Anderson, J. (1984). "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models." *Biometrika*, 71: 1–10. MR0738319. doi: http://dx.doi.org/10.1093/biomet/71.1.1. 551

Athreya, K. B. and Roy, V. (2014). "Monte Carlo Methods for Improper Target Distributions." *Electronic Journal of Statistics*, 8(2): 2664–2692. MR3292953. doi: http://dx.doi.org/10.1214/14-EJS969. 546

Athreya, K. B. and Roy, V. (2015). "Estimation of integrals with respect to infinite measures using regenerative sequences." *Journal of Applied Probability*, 52. 546

Berger, J., Strawderman, W., and Tang, D. (2005). "Posterior Propriety and Admissibility of Hyperpriors in Normal Hierarchical Models." *The Annals of Statistics*, 33(2): 606–646. MR2163154. doi: http://dx.doi.org/10.1214/009053605000000075. 546, 548, 552

Bronskill, S. E., Normand, S.-L. T., Beth Landrum, M., and Rosenheck, R. A. (2002). "Longitudinal profiles of health care providers." *Statistics in Medicine*, 21(8): 1067–1088. doi: http://dx.doi.org/10.1002/sim.1060. 549, 557

Carlin, B. P. and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman and Hall, first edition. MR1427749. 545

Chen, M.-H., Ibrahim, J. G., and Shao, Q.-M. (2004). "Propriety of the Posterior Distribution and Existence of the MLE for Regression Models with Covariates Missing at Random." *Journal of the American Statistical Association*, 99(466): 421–438. MR2062828. doi: http://dx.doi.org/10.1198/016214504000000368. 553

Chen, M.-H. and Shao, Q.-M. (2001). "Propriety of Posterior Distribution for Dichotomous Quantal Response Models." *Proceedings of the American Mathematical Society*, 129(1): 293–302. MR1694452. doi: http://dx.doi.org/10.1090/S0002-9939-00-05513-1. 551, 553

Chen, M.-H., Shao, Q.-M., and Xu, D. (2002). "Necessary and Sufficient Conditions on the Propriety of Posterior Distributions for Generalized Linear Mixed Models." *Sankhya*, 64 (Series A, Pt. 1): 57–85. MR1968375. 549, 551, 552, 553, 559

Dey, D. K., Gelfand, A. E., and Peng, F. (1997). "Generalized Linear Models for Small-Area Estimation." *Journal of Statistical Planning and Inference*, 64: 93–107. doi: http://dx.doi.org/10.1016/S0378-3758(96)00207-8. 553

Diggle, P., Liang, K.-Y., and Zeger, S. (1994). *Longitudinal Data Analysis*. New York: Oxford University Press. 549, 557

Everson, P. J. and Morris, C. N. (2000). "Inference for Multivariate Normal Hierarchical Models." *Journal of the Royal Statistical Society, Series B*, 62(2): 399–412. MR1749547. doi: http://dx.doi.org/10.1111/1467-9868.00239. 549, 557, 558, 559, 560

Garvan, C. W. and Ghosh, M. (1999). "On the Propriety of Posteriors for Dispersion Models." *Journal of Statistical Planning and Inference*, 78: 229–241. MR1705550. doi: http://dx.doi.org/10.1016/S0378-3758(98)00216-X. 553

Gelfand, A. E. and Sahu, S. K. (1999). "Identifiability, Improper Priors and Gibbs Sampling for Generalized Linear Models." *Journal of the American Statistical Association*, 94(445): 247–253. MR1689229. doi: http://dx.doi.org/10.2307/2669699. 553

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. London: Chapman and Hall, first edition. MR1385925. 545

Ghosh, M., Natarajan, K., Waller, L. A., and Kim, D. (1999). "Hierarchical Bayes GLMs for the Analysis of Spatial Data: An Application to Disease Mapping." *Journal of Statistical Planning and Inference*, 75: 305–318. MR1678978. doi: http://dx.doi.org/10.1016/S0378-3758(98)00150-5. 553

Hobert, J. P. and Casella, G. (1996). "The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models." *Journal of the American Statistical Association*, 91(436): 1461–1473. MR1439086. doi: http://dx.doi.org/10.2307/2291572. 545, 546

Hobert, J. P. and Casella, G. (1998). "Functional Compatibility, Markov Chains, and Gibbs Sampling with Improper Posteriors." *Journal of Computational and Graphical Statistics*, 7(1): 42–60. MR1628268. doi: http://dx.doi.org/10.2307/1390768. 545, 546

Horn, R. and Johnson, C. (1985). *Matrix Analysis*. Cambridge, UK: Cambridge University Press. MR0832183. doi: http://dx.doi.org/10.1017/CBO9780511810817. 564

Ibrahim, J. G. and Laud, P. W. (1991). "On Bayesian Analysis of Generalized Linear Models Using Jeffreys's Prior." *Journal of the American Statistical Association*, 86(416): 981–986. MR1146346. doi: http://dx.doi.org/10.1080/01621459.1991.10475141. 553

James, W. and Stein, C. (1961). "Estimation with Quadratic Loss Function." In: Neyman, J. (ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematic Statistics and Probability, Volume 1*, 361–379. Berkeley, CA: University of California Press. MR0133191. 560

Kim, S., Chen, M.-H., and Dey, D. K. (2008). "Flexible Generalized t-Link Models for Binary Response Data." *Biometrika*, 95(1): 93–106. MR2409717. doi: http://dx.doi.org/10.1093/biomet/asm079. 553

Kim, S. W. and Ibrahim, J. G. (2000). "On Bayesian Inference for Proportional Hazards Models Using Noninformative Priors." *Lifetime Data Analysis*, 6: 331–341. MR1818901. doi: http://dx.doi.org/10.1023/A:1026505331236. 553

Lin, X., Sun, D., Speckman, P. L., and Rouder, J. N. (2013). "Existence of MLE and Posteriors for a Recognition-Memory Model." *Statistics and Probability Letters*, 83: 2415–2421. MR3093833. doi: http://dx.doi.org/10.1016/j.spl.2013.06.033. 551

McCullagh, P. and Nelder, J. (1996). *Generalized Linear Models*. Cambridge, UK: Chapman and Hall. MR0727836. doi: http://dx.doi.org/10.1007/978-1-4899-3244-0. 548, 558

Michalak, S. E. (2001). "Multilevel Bernoulli Models for Evaluating Medical Departments in VA Hospitals." Ph.D. thesis, Harvard University Department of Statistics. MR2702055. 549, 557

Morris, C. N. (1983a). "Parametric Empirical Bayes Confidence Intervals." In: Box, G., Leonard, T., and Wu, C.-F. (eds.), *Statistical Inference, Data Analysis, and Robustness: Proceedings of a Conference Conducted by the Mathematics Research Center at the University of Wisconsin—Madison, November 4–6, 1981*, 25–50. New York: Academic Press, Inc. MR0772762. 560

Morris, C. N. (1983b). "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association*, 78(381): 47–55. MR0696849. doi: http://dx.doi.org/10.1080/01621459.1983.10477920. 560

Morris, C. N. and Lysy, M. (2012). "Shrinkage Estimation in Multi-Level Normal Models." *Statistical Science*, 27(1): 115–134. MR2953499. doi: http://dx.doi.org/10.1214/11-STS363. 560

Morris, C. N. and Tang, R. (2011). "Estimating Random Effects Via Adjustment for Density Maximization." *Statistical Science*, 26(2): 271–287. MR2858514. doi: http://dx.doi.org/10.1214/10-STS349. 560

Natarajan, R. and Kass, R. E. (2000). "Reference Bayesian Methods for Generalized Linear Mixed Models." *Journal of the American Statistical Association*, 95(449): 227–237. MR1803151. doi: http://dx.doi.org/10.2307/2669540. 545, 546, 549, 551, 552, 553, 557

Natarajan, R. and McCulloch, C. E. (1998). "Gibbs Sampling with Diffuse Proper Priors: A Valid Approach to Data-Driven Inference?" *Journal of Computational and Graphical Statistics*, 7(3): 267–277. 546

Pratt, J. W. (1981). "Concavity of the Log Likelihood." *Journal of the American Statistical Association*, 76(373): 103–106. MR0608179. doi: http://dx.doi.org/10.1080/01621459.1981.10477613. 551

Raghavan, N. and Cox, D. D. (1998). "Analysis of the Posterior for Spline Estimators in Logistic Regression." *Journal of Statistical Planning and Inference*, 71: 117–136. MR1651792. doi: http://dx.doi.org/10.1016/S0378-3758(98)00085-8. 553

Roy, V. and Dey, D. K. (2014). "Propriety of Posterior Distributions Arising in Categorical and Survival Models under Generalized Extreme Value Distribution." *Statistica Sinica*, 24: 699–722. MR3235395. 553

Roy, V. and Hobert, J. P. (2007). "Convergence Rates and Asymptotic Standard Errors for Markov Chain Monte Carlo Algorithms for Bayesian Probit Regression." *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 69(4): 607–623. MR2370071. doi: http://dx.doi.org/10.1111/j.1467-9868.2007.00602.x. 551

Sinha, S. K. (2004). "Robust Analysis of Generalized Linear Mixed Models." *Journal of the American Statistical Association*, 99(466): 451–460. MR2062830. doi: http://dx.doi.org/10.1198/016214504000000340. 549, 557

Sinharay, S. and Stern, H. S. (2002). "On the Sensitivity of Bayes Factors to the Prior Distributions." *The American Statistician*, 56(3): 196–201. MR1940207. doi: http://dx.doi.org/10.1198/000313002137. 549, 557

Speckman, P. L., Lee, J., and Sun, D. (2009). "Existence of the MLE and Propriety of Posteriors for a General Multinomial Choice Model." *Statistica Sinica*, 19: 731–748. MR2514185. 553

Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1995). *BUGS Version 0.5 Manual*. MRC Biostatistics Unit, Cambridge, UK. 551

Stein, C. M. (1981). "Estimation of the Mean of a Multivariate Normal Distribution." *The Annals of Statistics*, 9(9): 1135–1151. MR0630098. doi: http://dx.doi.org/10.1214/aos/1176345632. 560

Sun, D., Tsutakawa, R. K., and He, Z. (2001). "Propriety of Posteriors with Improper Priors in Hierarchical Linear Mixed Models." *Statistica Sinica*, 11: 77–95. MR1820002. 553

Tiao, G. and Tan, W. (1965). "Bayesian Analysis of Random-Effects Models in the Analysis of Variance I. Posterior Distribution of Variance Components." *Biometrika*, 52: 37–53. MR0208765. doi: http://dx.doi.org/10.1093/biomet/52.1-2.37. 546

Wang, C., Rutledge, J., and Gianola, D. (1994). "Bayesian Analysis of Mixed Linear Models via Gibbs Sampling with an Application to Litter Size of Iberian Pigs." *Genetique, Selection, Evolution*, 26: 1–25. 546

Wedderburn, R. (1976). "On the Existence and Uniqueness of the Maximum Likelihood Estimates for Certain Generalized Linear Models." *Biometrika*, 63(1): 27–32. MR0408092. doi: http://dx.doi.org/10.1093/biomet/63.1.27. 551

Wong, G. Y. and Mason, W. M. (1985). "The Hierarchical Logistic Regression Model for Multilevel Analysis." *Journal of the American Statistical Association*, 80(391): 513–524. doi: http://dx.doi.org/10.1080/01621459.1985.10478148. 549, 557

Yang, R. and Chen, M.-H. (1995). "Bayesian Analysis for Random Coefficient Regression Models Using Noninformative Priors." *Journal of Multivariate Analysis*, 55: 283–311. MR1370405. doi: http://dx.doi.org/10.1006/jmva.1995.1080. 553

Zeger, S. L. and Karim, M. R. (1991). "Generalized Linear Models with Random Effects: A Gibbs Sampling Approach." *Journal of the American Statistical Association*, 86(413): 79–86. MR1137101. doi: http://dx.doi.org/10.1080/01621459.1991.10475006. 546, 549, 557

Acknowledgments

The authors thank Ted Stefos, Jim Burgess, and Cindy Christiansen for their collaboration on the VA hospital profiling work that motivated this research, Scott Vander Wiel for providing an alternative proof of Lemma 5, Joe Blitzstein for helpful discussions, Nick Hengartner for reading an early draft of this work, and the editor and a referee for their very useful comments. Dr. Michalak's research was partially supported by the Management Science Group, U.S. Department of Veterans Affairs, Bedford, MA. Dr. Morris' research was partially supported by the Management Science Group, U.S. Department of Veterans Affairs, Bedford, MA; The Center for Health Quality, Outcomes, and Economic Research, U.S. Department of Veterans Affairs, Bedford, MA; and NSF grant DMS-97-05156.