Selection of Random Effects Distributions in Mixed Counts Models: A Quasi-Likelihood Approach

Selection Of Random Effects Distributions In Mixed Counts Models: A Quasi-Likelihood Approach

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Jamie D. Riggs

University Of Northern Colorado
Greeley, Colorado
The Graduate School

College of Education and Behavioral Science
Applied Statistics and Research Methods

December 2013

This Dissertation by: Jamie D. Riggs

Entitled: *Selection Of Random Effects Distributions In Mixed Counts Models: A Quasi-Likelihood Approach*

has been approved as meeting the requirement for the Degree of Doctor of Philosophy in College of Education and Behavioral Sciences, Program of Applied Statistics and Research Methods

Accepted by the Doctoral Committee

____

Jay Schaffer, Ph.D., Co-Chair

____

Trent Lalonde, Ph.D., Co-Chair

____

Dennis Akos, Ph.D., Committee Member

____

Robert Heiny, Ph.D., Faculty Representative

Date of Dissertation Defense ________________________________

Accepted by the Graduate School

____

Linda L. Black, Ed.D., LPC
Dean of the Graduate School and International Admissions

Abstract

Riggs, J.D. (2013). *Selection of Random Effects Distributions in Mixed Counts Models: A Quasii-Likelihood Approach.* PhD thesis, University of Northern Colorado, Greeley, CO.

Generalized linear models with responses comprised of counts data are clustered into homoscedastic groups by random effects that are considered to follow either a normal or a gamma distribution. There are data sets for which it is shown that the standard errors of the random effects estimates are only slightly adjusted from those of the normal and gamma distributions when the random effects use a power function, mean-variance relationship, or quasi distribution. The adjusted standard errors is demonstrated by subjecting the random effects of selected data sets to the power function quasi distribution, after first using these data to estimate a value for the power function exponent, a method not used in the literature. The efficacy of the quasi distribution is measured by comparing it with normal and gamma distributions' random effects standard errors, model overdispersion, and the standardized random effects deviance residuals diagnostic plots. Comparison results show that the power function quasi distribution model of random effects benefits is data set-dependent.

Dedication

This dissertation is dedicated to:

George, a jewel in the crown of statistics;
Kimberly, a jewel bound for the statistical crown;
Keyleigh, a ready source of encouragement, and a wonderful listener;
Louise, for whom seeking the truth is a constant endeavor;
Lauren, the light of my life;
Jordan, my inspiration.

Acknowledgments

I acknowledge Dr. Trent Lalonde for his patience and guidance. My thanks to Dennis Akos for his comments, recommendations, and advice. Jay Schafer offered suggestions and counsel that led to my successful graduate experience. My faculty representative, Robert Heiny, offered insights that strengthened my research. Thank you all for giving me a rewarding research experience.

I give a special thanks to Rodney Howe of the American Association of Variable Star Observers Solar Section for making the sunspot data available for this dissertation.

Preface

Discovering the unexpected in scientific data is often the path to fundamental breakthroughs in the physical sciences. The complexity of scientific inquiry and experimentation increases as more subtle discoveries are sought. These subtle discoveries often elude the capabilities of classical statistical techniques to detect, suggesting the use of modern statistical methods. For instance, least squares solutions may result in biased parameter estimates whereas likelihood estimators may provide optimal error structures.

The motivation for this research into the power function mean-variance relationship quasi distributions materialized from unsuccessful application of normal linear model methods to sunspot counts data. Generalized linear modeling also had difficulties when assuming that the random effects were normally distributed. However, the power function quasi distribution holds promise.

The results of this dissertation show that non-normal random effects is one of the subtle complexities that evades resolution by classical least squares solutions, and yet is well-suited to hierarchical generalized linear models quasi-likelihood solutions. While this research produced promising outcomes for the sunspot data, there is room for further refinement of the technique. This modern quasi-likelihood method may be artfully applied to many complex and subtle modeling problems.

Symbols

Bold character	vectors or matrices
$\doteq$	approximately equal to
$\oplus$	exclusive or
$\in$	an element of
$\prime$	indicates a derivative
$\propto$	proportional to
$\sim$	distributed as
$\supset$	contained by
$\hat{\cdot}$	an estimated parameter
$\boldsymbol{\beta}$	fixed effects parameter vector
$Beta(\cdot,\cdot)$	beta distribution
$\boldsymbol{\gamma}$	response or conditional response dispersion model parameter vector
Γ	diagonal matrix of overdispersion parameters
$\boldsymbol{\delta}$	random effects dispersion model parameter vector
ζ	random effects dispersion parameter
$\boldsymbol{\eta}$	systematic component matrix of the conditional response mean model
$\boldsymbol{\eta}_d$	systematic component of the conditional response dispersion model
$\boldsymbol{\eta}_R$	systematic component of the random effects mean model
$\boldsymbol{\eta}_{dR}$	systematic component of the random effects dispersion model
$\boldsymbol{\theta}$	exponential family distribution canonical parameter vector
λ	power normal distribution truncation parameter
μ	mean of a distribution
μ_i	response mean
μ_{ij}	conditional response mean
μ_{Ri}	random effects mean
$\boldsymbol{\xi}$	vector of pseudo-responses for the iterated weighted least squares
π	mathematical constant
$\boldsymbol{\Sigma}$	variance-covariance matrix
σ^2	variance

ϕ	dispersion parameter
$\Phi(\cdot,\cdot)$	standard normal cumulative distribution function
χ^2	chi-square distribution
ψ	power function mean-variance relationship quasi distribution exponent
ω	fixed, unknown parameter vector
$\mathcal{E}(\mathrm{cot})$	expected value
$D(\cdot,\cdot)$	deviance function
$\boldsymbol{d}$	conditional response mean model deviance residuals vector
$\boldsymbol{d}_R$	random effects mean model deviance residuals vector
e	mathematical constant
e^x	exponential function
$exp(\cdot)$	exponential function
$f(\cdot)$	probability density function
g	sunspot group count
$g(\cdot)$	conditional response mean model link function
$g_d(\cdot)$	conditional response dispersion link function
$g_R(\cdot)$	random effects mean model link function
$g_{dR}(\cdot)$	random effects dispersion model link function
$gamma(\cdot,\cdot)$	gamma distribution
$h(\cdot,\cdot)$	h-likelihood
$h^+(\cdot,\cdot)$	adjusted profile h-likelihood
$\boldsymbol{I}$	identity matrix
iid	identically independently distributed
$L(\cdot)$	likelihood function
$l(\cdot)$	log-likelihood function
$\ln(\cdot)$	natural logarithm
$\max(\cdot)$	maximum
$\min(\cdot)$	minimum
$\mathcal{N}(\cdot,\cdot)$	normal distribution
$\mathcal{N}(\cdot,\cdot)$	normal distribution
$Poi(\cdot)$	Poisson distribution
$Q(\cdot,\cdot)$	quasi-likelihood
$Q^+(\cdot,\cdot)$	extended quasi-likelihood
$Q^*(\cdot,\cdot)$	generalized extended quasi-likelihood
$q(\cdot,\cdot)$	total quasi-likelihood
$q^+(\cdot,\cdot)$	total extended quasi-likelihood
s	sunspot count
$s.e.(\cdot)$	standard error
$\boldsymbol{u}$	random effects parameter vector

$Var(\cdot)$	variance of a random variable
$\boldsymbol{v}$	transformed random effects parameter vector
$\boldsymbol{W}$	matrix of iterated least squares weights
w	Wolf number
X	random variable
X^2	Pearson's chi-square
$\boldsymbol{X}$	fixed effects design matrix
Y	response vector, particularly of counts data
$\boldsymbol{Z}$	random effects design matrix

Abbreviations

AAVSO	American Association of Variable Star Observers
BCT	Box-Cox transformation
CI	confidence interval
CLT	central limit theorem
CR	conditional response
DBC	deviance-based criterion
DEQL	double extended quasi-likelihood
DHGLM	double hierarchical generalized linear model
EF	exponential family
EQGLM	extended quasi-generalized linear model
EQL	extended quasi-likelihood
FE	fixed effect
GEQL	generalized extended quasi-likelihood
GEQGLM	generalized extended quasi-generalized linear model
GLM	generalized linear model
HGLM	hierarchical generalized linear model
IWLS	iterated weighted least-squares
JGLM	joint generalized linear model
LSE	least squares estimator
MHLE	maximum h-likelihood estimate
MLE	maximum likelihood estimator
MME	method of moments estimator
PL	profile likelihood
PN	power normal distribution
Q-Q	quantile-quantile (plot)
QGLM	quasi-generalized linear model
QHGLM	quasi-hierarchical generalized linear model
QL	quasi-likelihood
RE	random effect

REML	restricted maximum likelihood
TN	truncated normal distribution

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Discovering the unexpected in scientific data is often the path to fundamental breakthroughs in the physical sciences. The complexity of scientific inquiry and experimentation increases as more subtle discoveries are sought. These subtle discoveries often elude the capabilities of classical statistical techniques to detect, suggesting the use of modern statistical methods. For instance, least squares solutions may result in biased parameter estimates whereas likelihood estimators may provide optimal error structures.

The motivation for this research into the power function mean-variance relationship quasi distributions materialized from unsuccessful application of normal linear model methods to sunspot counts data. Generalized linear modeling also had difficulties when assuming that the random effects were normally distributed. However, the power function quasi distribution holds promise.

The results of this dissertation show that non-normal random effects is one of the subtle complexities that evades resolution by classical least squares solutions, and yet is well-suited to hierarchical generalized linear models quasi-likelihood solutions. While this research produced promising outcomes for the sunspot data, there is room for further refinement of the technique. This modern quasi-likelihood method may be artfully applied to many complex and subtle modeling problems.

Generalized linear models enjoy substantial treatment in the statistics literature. These treatments consider data with binary, multinomial, and counts responses modeled by techniques similar to linear regression and analysis of variance techniques in the forms of logit, probit, and log-linear analyses. Nelder & Wedderburn (1972) showed that these techniques and analyses share properties such as linearity in model parameters and common estimation of parameters methods. Nelder's work has been expanded from fixed effects models to models for random effects and dispersion. The generalized linear model, then, may be extended from one to four submodels to form a single generalized linear model; that is, a model for the mean of the fixed effects, a model for the fixed effects dispersion, a model of the mean of the random effects, and a model for the random effects dispersion.

This dissertation examines modeling the mean and variance of random effects in generalized linear models that fit what is considered to be overdispersed counts data. Overdispersion for Poisson-distributed counts data occurs when the estimated variance is greater than the estimated expected value. Suppose a data set is thought to follow a Poisson distribution with mean μ, but the variance of these data exceed μ, that is, for a random variable $Y_i,\ i = 1, 2, \ldots, n,\ Var(\hat{Y}_i) > \mathcal{E}(\hat{Y}_i)$. Then these data are considered to be overdispersed. Overdispersion occurs frequently in practice and can manifest in data clustered by random effects. For example, for a Poisson process, the observation interval may be a random length rather than a fixed length; or each observed event may contribute a random amount to the total. In these cases, a random effect added in the model may reduce the amount of overdispersion.

Overdispersion in data with a Poisson response conditional on random effects of unspecified distribution is the system examined. These overdispered data will be analyzed within the context of generalized linear models using maximum likelihood, double extended quasi-likelihood (Lee & Nelder, 1996), h-likelihood (Lee & Nelder, 2006), and the extension proposed in Chapter 3 of this work.

1.1 Generalized Linear Model Structures

Classical response-with-covariates models, that is, general linear mixed models, assume the response variable and the random effects (REs), including the residual error, follow a normal distribution, are linear in the model parameters, and have constant variance. This allows model parameter estimation with least squares methods. Many data sets have response variables and random effects that violate one or more of these assumptions, for example, responses that follow a Poisson distribution are unlikely to be linear in the model parameters. While remedial measures such as transformations on the response variable or the covariates may be applied, these measures may fall short of satisfying the assumptions. For data sets for which classical models are ill suited, the extended class of models, generalized linear models (GLMs), are used, with model parameters often estimated using likelihood methods.

Nelder & Wedderburn (1972) introduced a unification of models linear on the systematic component (model predictors) such as logistic and probit analysis for binomial variates, contingency tables for multinomialvariates, and regression for Poisson- and gamma-distributed variates, in the form of the GLM. GLMs for an individual random response variable, $Y_i,\ i = 1, 2, \ldots, n$, may be represented as

$$\begin{aligned} &Y_i \sim EF(\mu_i, a(\phi)V(\mu_i)),\\ &\eta_i = \boldsymbol{x}_i^T\boldsymbol{\beta},\\ &\eta_i = g(\mu_i). \end{aligned} \tag{1.1}$$

In equation 1.1, EF denotes a response variable distribution from the exponential family (EF), μ_i is the response variable mean, ϕ is the EF dispersion parameter in the dispersion function $a(\cdot)$, $V(\mu_i)$ is the response variable variance function, η_i is the systematic component, the $\boldsymbol{x}_i^T = \{x_{i1},\ x_{i2},\ \ldots,\ x_{ip}\}^T$ is the i^{th} row of the systematic component design matrix of covariates, $\boldsymbol{\beta} = \{\beta_1,\ \beta_2,\ \ldots,\ \beta_p\}$ is the vector of covariates parameters, and $g(\cdot)$ is the link function.

Wedderburn (1974) generalized the GLM response variable distribution from a specific probability distribution in the EF to an EF distribution with just a specification of the first and second moments. Specification of the just the first and second moments gives the components of a quasi-generalized linear model as

$$\begin{aligned} &Y_i \sim (\mu_i, a(\phi)V(\mu_i)),\\ &\eta_i = \boldsymbol{x}_i^T\boldsymbol{\beta},\\ &\eta_i = g(\mu_i), \end{aligned} \tag{1.2}$$

where the terms are as defined in Equations 1.1, except no EF probability distribution for the response variable is specified. Models of the form of Equations 1.1 and 1.2 are known as mean models as the mean is linked to the model parameters. The random response Y_i are assumed to be distributed according to an unknown EF, namely, $Y_i \sim EF(\cdot, \cdot)$; however, the generalized linear model literature drops the EF designator for GLMs that are not necessarily probability distribution function specific, that is, $Y_i \sim (\cdot, \cdot)$. This practice will be followed here as well.

Nelder & Pregibon (1987) expanded the quasi-generalized linear model from a solitary mean model to include a model of the variance components, which is known as the dispersion model. Wedderburn's quasi-generalized linear model assumes the dispersion parameter ϕ is constant. Nelder and Pregibon relax this requirement allowing the dispersion parameter ϕ_i to vary with each predictor. A varying dispersion parameter allows modeling of overdispersion. Nelder's and Pregibon's form of the generalized linear model is called an extended quasi-generalized linear model. Figure 1.1 column 1 is a representation of the relationships between the mean model and, in column 2, the dispersion model in the extended quasi-generalized linear model. The terms for the mean model are as in Equation 1.2 with the response variable variance function denoted $\phi_i V(\mu_i)$ to account for the varying dispersion parameter. The random response variable of the dispersion model is the deviance d_i that follows a distribution with the first and second moments specified by the EF dispersion parameter ϕ_i with the variance quadratic on the mean, the systematic component η_{di} is the i^{th} row of the dispersion covariates design matrix $\boldsymbol{g}_{di} = (g_{i1},\ g_{i2},\ \ldots,\ g_{ip})^T$, dispersion covariates parameters vector $\boldsymbol{\gamma}$, and the link function is $g_d(\cdot)$. The mean model and the dispersion model may be solved jointly (Nelder & Lee, 1991), and are then called joint GLMs (JGLMs).

Mean Model Components	Dispersion Model Components
$Y_i \sim (\mu_i, a(\phi_i)V(\mu_i))$	$d_i \sim (\phi_i, 2\phi_i^2)$
$\eta_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$	$\eta_{di} = \boldsymbol{g}_{di}^T \boldsymbol{\gamma}$
$\eta_i = g(\mu_i)$	$\eta_{di} = g_d(\phi_i)$

Figure 1.1 The GLM structures for the JGLMs include a mean model and a dispersion model.

The various forms of the mean and dispersion models presented above describe the mean model for covariates that consist of fixed effects. Fixed effects (FEs) are defined by Milliken & Johnson (1992) as effects for which all possible levels are used in the model. A significant number of data sets contain sources of variance in addition to those attributable to the dispersion models already discussed. This additional variation often originates with random effects (REs), defined by Milliken & Johnson (1992) as those effects for which an incomplete number of levels of covariates are used in the model, or the covariates are of random length. Recall as an example the random interval length of a Poisson process. Random effects in generalized linear models were introduced into the GLM to form the Hierarchical Generalized Linear Model (HGLM) by Lee & Nelder (1996) and Lee & Nelder (2001). The HGLM is the GLM analog to the general linear mixed model, which may contain both fixed effects and random effects.

There are three submodels in the hierarchical generalized linear model: the conditional response mean model which replaces the fixed effects mean models in GLMs and quasi-GLMs due to dependence on random effects, the conditional response dispersion model, and the random effects mean model. The terms for the first row of Figure 1.2 are the random, individual grouped responses, conditional (clustered) on random effects u_i, Y_{ij} from a generalized linear model family such that $\mathcal{E}(Y_{ij} \mid u_i) = \mu_{ij}$, and $Var(Y_{ij} \mid u_i) = a(\phi_{ij})V(\mu_{ij})$ with linear predictor

$\eta_{ij} = g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{v}$. The conditional response design matrix X, the parameter vector β and the $\eta_{ij} = g(\mu_{ij})$ is analogous to the GLM and quasi-GLM above. The random effects parameter vector $\boldsymbol{v} = (v_1,\ v_2,\ \ldots,\ v_n)^T$ is a monotonic function of the random effects u_i that transforms the u_i commensurate with the link function $g(\mu_{ij})$. The random effects design matrix is $\boldsymbol{Z} = (\boldsymbol{z}_{i1},\ \boldsymbol{z}_{i2},\ \ldots,\ \boldsymbol{z}_{iq})^T$.

The second submodel is the dispersion model analogous to the dispersion model in Figure 1.1. The difference in indices is due to the clustering resulting from the random effects.

The second row of Figure 1.2 is the mean model of the random effects. The variable u_i is a vector of random effects that may be based on a particular distribution or on the specification of first and second moments. The function $v_i = g_R(u_i)$ accounts for the use of a link function η_{ij}, and is a strictly monotonic function. This is the third submodel.

Effect	Mean Model Components	Dispersion Model Components
Fixed	$Y_{ij} \mid u_i \sim (\mu_{ij}, a(\phi_{ij})V(\mu_{ij}))$ $\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{v}$ $\eta_{ij} = g(\mu_{ij})$	$d_{ij} \sim (\phi_{ij}, 2\phi_{ij}^2)$ $\eta_{dij} = \boldsymbol{g}_i^T\boldsymbol{\gamma}$ $\eta_{dij} = g_d(\phi_{ij})$
Random	$u_i \sim (\mu_{Ri}, \zeta V_R(\mu_{Ri}))$ $v_i = g_R(u_i)$	

Figure 1.2 The GLM structures for the HGLMs include a mean model, a dispersion model, and a random effects model.

Generalized linear models with random effects were further extended by Lee & Nelder (2006) to include a model of the random effects dispersion in addition to the random effects mean model. This extension is known as the double hierarchical generalized linear model (DHGLM). Figure 1.3 is a representation of the relationships between the conditional response mean and dispersion models, and the random effects mean and dispersion models in the DHGLM. The first row of Figure 1.3 is as defined in Figure 1.2, as is the second row first column. The second column of the second row is the random effects dispersion model. The random effects dispersion model terms are the random effects deviance components d_{Ri} for which the first and second moments of a distribution are specified, the systematic component η_{dRi} represents the design matrix $\boldsymbol{g}_{Ri}^T$ that combines with the the dispersion model parameter vector $\boldsymbol{\delta}$ to be estimated, and the link function $g_{dR}(\cdot)$.

An aspect of HGLMs that is not treated in the generalized linear model lit-

Effect	Mean Model Components	Dispersion Model Components
Fixed	$Y_i \mid u_i \sim (\mu_{ij}), a(\phi_{ij})V(\mu_{ij}))$ $\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{v}$ $\eta_{ij} = g(\mu_{ij})$	$d_{ij} \sim (\phi_{ij}, 2\phi_{ij}^2)$ $\eta_{dij} = \boldsymbol{g}_{di}^T\boldsymbol{\gamma}$ $\eta_{dij} = g_d(\phi_{ij})$
Random	$u_i \sim (\mu_{Ri}, \zeta_i V_R(\mu_{Ri}))$ $v_i = g_R(u_i)$	$d_{R_i} \sim (\zeta_i, 2\zeta_i^2)$ $\eta_{dRi} = \boldsymbol{g}_{dRi}^T\boldsymbol{\delta}$ $\eta_{dRi} = g_{dR}(\zeta_i)$

Figure 1.3 The GLM structures for the DHGLMs include a mean model, a dispersion model, a random effects model, and a random effects dispersion model.

erature, and is the focus of this dissertation, is when the random effects have a power function mean-variance relationship, $V_R(\mu_{Ri}) = \zeta\mu_{Ri}^{\psi}$ to form a quasi distribution. Figure 1.4 is the random effects power mean function generalized linear models. The conditional response mean model is as defined in Figure 1.2. The conditional response dispersion model is simplified from that of Figure 1.2 by setting $\eta_{di} = \boldsymbol{g}_i^T\boldsymbol{\gamma} = \gamma_0$, for all $i = 1, 2, \ldots, n$. The random effects part of Figure 1.4 differs from that of the HGLM (Figure 1.2) by the power function relationship of the mean-variance relationship of the random effects mean model. The random effects dispersion model is simplified similarly to the conditional response dispersion model by setting $\eta_{dRi} = \boldsymbol{g}_{Ri}^T\boldsymbol{\delta} = \delta_0$, also for all $i = 1, 2, \ldots, n$. This GLM structure of the HGLM is studied in this dissertation.

Effect	Mean Model Components	Dispersion Model Components
Fixed	$Y_{ij} \mid u_i \sim (\mu_{ij}), a(\phi_{ij})V(\mu_{ij}))$ $\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{v}$ $\eta_{ij} = g(\mu_{ij})$	$d_{ij} \sim (\phi_{ij}, 2\phi_{ij}^2)$ $\eta_{dij} = \gamma_0$ $\eta_{dij} = g_d(\phi_{ij})$
Random	$u_i \sim (\mu_{Ri}, \zeta_i\mu_{Ri}^{\psi})$ $v_i = g_R(u_i)$	$d_{R_i} \sim (\zeta_i, 2\zeta_i^2)$ $\eta_{dRi} = \delta_0$ $\eta_{dRi} = g_{dR}(\zeta)$

Figure 1.4 The GLM structures for the DHGLMs include a mean model, a dispersion model, a random effects power function, mean-variance relationship model, and a random effects dispersion model.

A key characteristic for constructing the random effects power function mean-variance relationship is overdispersion. An incorrect specification of the random effects mean-variance relationship power function exponent, ψ, is known to lead to overdispersion. Recall that overdispersion is more variation than is expected from the GLM specifications, and now is described in more detail.

Overdispersion has two major consequences (Cox, 1983). The first is that summary statistics and parameter estimates have variances that are larger than anticipated under a simpler model. This may result in incorrect conclusions concerning inferences about the model. This problem has long been recognized, and is commonly allowed for by an empirical inflation factor, either assumed from prior experience or estimated. The second consequence is the possible loss of efficiency in using statistics that are appropriate for a single-parameter family. Overdispersion, as it is affected by the incorrect specification of random effects mean-variance power function exponent, ψ, is the focal point of this dissertation. The method for estimating ψ so overdispersion is minimized now is described.

1.2 Methodology

The generalized linear model literature has not addressed the situation when the random effects mean model is defined by a power function mean-variance relationship of the form $V_R(\mu_{Ri}) = \zeta\mu_{Ri}^{\psi},\ i = 1, 2, \ldots, n$, where ζ is the dispersion parameter, $V_R(\mu_i)$ is the variance function, μ_{Ri} is the mean of the random effects, and ψ is the power function exponent. This dissertation examines a method by which to estimate the value of ψ when it is unknown and requires estimation. Further, it will be shown that overdispersion in either or both the conditional response and random effects deviances provides information that leads to the estimate of ψ. The uniqueness of this research derives from insight into the power normal distribution parameterization flexibility as regards modeling deviance truncation and skewness, which can characterize the random effects variance exponent ψ. The insight stems from realizing the hierarchical generalized linear model deviance holds information that can characterize ψ.

Three data sets are used to show the efficacy of employing a random effects power function mean-variance relationship quasi distribution to reduce the amount of overdispersion. Two data sets are used in the literature for model comparisons: They are the fabric data of Bissell (1972), and the rats data of Myers & Montgomery (2002). It is shown that the random effects power function mean-variance relationship quasi distribution parameters are identical when ψ represents either a normal distribution or a gamma distribution. The third data set is sunspot counts data supplied by the American Association of Variable Star Observers Solar Section. The power function mean-variance relationship for the random effects is shown to be no worse than the overdispersion in which the random effects are modeled by either a normal or a gamma distribution.

1.3 Objectives

The classical linear modeling of overdispersed data often involves a response variable transformation that attempts to satisfy the assumption of a normally-distributed response. As the response variable of interest is counts, they are subject to a large quantity of apparent outliers. These apparent outliers generally result as the more appropriate Poisson distribution for the response variables. However, even when the counts data are modeled by a hierarchical generalized linear model with the counts conditional upon random effects that follow a normal or a gamma distribution, significant overdispersion remains. This dissertation explores the use of a power function mean-variance relationship for the random effects, and its effect on the random effects mean model deviation.

The objective of this research is to account for overdispersion in HGLMs with counts conditional response variables that follow a Poisson distribution, and the random effects mean-variance relationship has a power function whose exponent, ψ, must be estimated.

The research questions are:

- Q1 Can the power function exponent, ψ, of the random effects mean-variance relationship quasi distribution can be estimated?

- Q2 Will estimates of the power function exponent, ψ, reduce the amount of overdispersion in the conditional response over those when the random effects follow a normal or the gamma distribution?

- Q3 Will estimates of the power function exponent, ψ, reduce the random effects estimated standard errors over the standard errors from random effects of either a normal or a gamma distribution?

Chapter 2 provides the background and literature review required to develop the use of a power function mean-variance relationship quasi distribution for the random effects. Chapter 3 develops the methods and techniques used to define the overdispersion in HGLMs with a power function mean-variance relationship for the random effects. Chapter 4 presents the efficacy of the power function on overdispersion and random effects standard errors for the fabric, rats, and sunspot data.

Chapter 5 draws conclusions on the power function utilization along with future developments.

Chapter 2

Literature Review

This chapter reviews the literature on the development of the class of statistical models known as generalized linear models with the purpose of providing the background necessary to support the enhancement to these models that this dissertation advances. The Introduction Section introduces the topic of generalized linear models as an extension of general linear models. The Generalized Linear Models Section gives the generalized linear model details necessary for the development of model extensions. The Quasi-Generalized Linear Models and Quasi-Likelihood Section describes quasi-generalized linear models. The Extended Quasi-Generalized Linear Model Section describes extended quasi-generalized linear models. The Generalized Extended Quasi-Generalized Linear Model Section describes generalized extended quasi-generalized linear models. The Joint Generalized Linear Model Estimation Section describes joint generalized linear model estimation. The Hierarchical Generalized Linear Models Section describes hierarchical generalized linear models. The IWLS for DEQL Section describes the iterated weighted least squares procedure for double extended quasi-likelihood estimation. Finally, the Deviance-Based Criterion For GLM Selection Section describes how GLMs are selected from their deviances.

2.1 Introduction

Linear models theory has been applied successfully to many data sets. Perhaps the best known linear models are those using analysis of variance and regression analysis, which use least-squares methods to estimate model parameters. Least squares estimation requires the response variables follow a normal distribution, and the relationship between the response variables and the covariates is the identity.

In linear models, the response variables are continuous data. Thus, for a vector $\boldsymbol{Y} = (Y_1,\ Y_2,\ \ldots,\ Y_n)^T$ of random variables, $\boldsymbol{Y} \overset{iid}{\sim} \mathcal{N}(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$, where $\overset{iid}{\sim}$ means identically independently distributed, $\mathcal{N}(\cdot,\cdot)$ indicates a normal distribution, $\boldsymbol{X}$ is the $n \times p$ design matrix of covariates, $\boldsymbol{\beta} = (\beta_1,\ \beta_2,\ \ldots,\ \beta_p)^T$ is the vector of parameters to be estimated via least squares, and $\boldsymbol{I}$ is the $p \times p$ identity matrix that equivalently distributes the variance σ^2 across the p covariates. The linear expectation model for fixed effects may be written in matrix form as

$$\mathcal{E}(\boldsymbol{Y}) = \boldsymbol{X\beta}. \tag{2.1}$$

The linear expectation model with both fixed and random effects may be written

$$\mathcal{E}(\boldsymbol{Y} \mid \boldsymbol{u}_1,\ \boldsymbol{u}_2,\ \ldots,\ \boldsymbol{u}_k) = \boldsymbol{X\beta} + \boldsymbol{Z_1u_1} + \cdots + \boldsymbol{Z_ku_k}, \tag{2.2}$$

where $\boldsymbol{Y}$, $\boldsymbol{X}$, and $\boldsymbol{\beta}$ are as in Equation 2.1; $\boldsymbol{Z_1u_1} + \cdots + \boldsymbol{Z_ku_k}$ is the random effects part with $\boldsymbol{u_i} \overset{iid}{\sim} \mathcal{N}(\boldsymbol{0}, \sigma_i^2\boldsymbol{I})$, and $\boldsymbol{Z_i}$ are the random effects design matrices of known constants, for $i = 1, 2, \ldots, k$; and the population parameters of this mixed effects model are $\boldsymbol{\beta}$ for the fixed effects, and $\sigma_1^2,\ \sigma_2^2, \ldots, \sigma_k^2$ for the k random effects.

The analysis of the random part includes estimating, hypothesis testing, and confidence interval construction about the σ_i^2; and the analysis of the fixed part includes estimating, hypothesis testing, and confidence interval construction about estimable functions of $\boldsymbol{\beta}$.

Standard estimation methods include ordinary least squares, generalized least squares, and the mixed model equations described by Milliken & Johnson (1992). Although linear models can be useful for nonnormal data as well as for normal data, these standard estimation methods do not necessarily produce usable results for nonnormal data modeling. Some estimation problems include inefficient or inaccurate estimates, estimates outside the range of permissible values, and misleading significance values in hypothesis tests. See Box & Cox (1964); Box, Hunter & Hunter (1978); and Belsley, Kuh & Welsch (1980).

2.2 Generalized Linear Models

The generalized linear model Nelder & Wedderburn (1972) extends linear model theory to allow for responses that do not necessarily follow a normal distribution or a constant-variance distribution, and it allows for response-to-covariates relationships to be other than the identity. The generalized linear model (GLM) was developed for distributions included in the exponential family of distributions, and use likelihood estimation for parameter evaluation. Likelihood function extensions, which relax the generalized linear model from distribution specific models to models that specify only the mean-variance relationships, provide less restrictive parameter estimation than the maximum likelihood. These extensions also allow modeling of both fixed and random effects which may be applied to data sets that are structurally over-dispersed, i.e., possess more variation than assumed by the host probability distribution or mean-variance relationship. First, the structures of GLMs is needed.

2.2.1 Generalized Linear Model Structure

The generalized linear model for exponential family (EF) response variables consists of three components. The first component is a random component. The second component is a systematic component. The third component is a link function.

The first component is a random component that describes the variation in the data. The random component is a random variable Y_i $i = 1, 2. \ldots, n$, with a probability distribution function from the EF, such as the normal, Poisson, gamma, and binomial, etc., as described above.

The second component is a systematic component

$$\eta_i = \boldsymbol{x}_i^T \boldsymbol{\beta}, \tag{2.3}$$

that specifies the variation in the response variable accounted for by the p-dimensional vector of design matrix covariates $\boldsymbol{x}_i^T = (\boldsymbol{x}_{i1},\ \boldsymbol{x}_{i2},\ \ldots,\ \boldsymbol{x}_{ip})^T$, $i = 1, 2, \ldots, n$. These covariates may be quantitative, nominal, or they may be a mixture or both. The vector of model parameters, $\boldsymbol{\beta} = (\beta_1,\ \beta_2,\ \ldots,\ \beta_p)^T$, are considered unknown values that require estimation.

The third component is a link function

$$\eta_i = g(\mu_i), \tag{2.4}$$

that specifies the relationship between the random component and the systematic component. The link function, $g(\cdot)$, is strictly monotonic and is twice differentiable. The mean, μ_i, is the expected value of the random response Y_i such that $\mu_i = \mathcal{E}(Y_i)$.

The GLM structure assumes a random response variable with a distribution from the EF such that, for an individual random response variable, Y_i, $i =$

$1, 2, \ldots, n$, a mean model is

$$
\begin{aligned}
&Y_i \sim EF(\mu_i, a(\phi)V(\mu_i)),\\
&\eta_i = \boldsymbol{x}_i^T\boldsymbol{\beta},\\
&\eta_i = g(\mu_i).
\end{aligned}
\tag{2.5}
$$

In equation 2.5, EF denotes a response variable distribution from the exponential family, μ_i is the response variable mean, ϕ is the EF dispersion parameter used in the constant function $a(\cdot)$, $V(\mu_i)$ is the response variable variance function, η_i is the systematic component, the $\boldsymbol{x}_i^T = \{\boldsymbol{x}_{i1},\ \boldsymbol{x}_{i2},\ \ldots,\ \boldsymbol{x}_{ip}\}^T$ is the i^{th} row of the covariates design matrix, $\boldsymbol{\beta} = \{\beta_1,\ \beta_2,\ \ldots,\ \beta_p\}$ is the vector of covariates parameters, and $g(\cdot)$ is the link function.

2.2.2 Parameter Estimation for Generalized Linear Models

Various methods for GLM parameter estimation have been presented by Nelder & Wedderburn (1972); McCullagh & Nelder (1989); Lee, Nelder & Pawitan (2006). The GLM with predictors used to estimate the parameters of the mean model with systematic component $\eta_i = \boldsymbol{x}_i^T\boldsymbol{\beta}$, and link function $g(\mu_i)$, has the log-likelihood

$$
\ell(\boldsymbol{\theta}, \phi; \boldsymbol{y}) = \sum_{i=1}^{n} \left[\frac{y_i g(\boldsymbol{x}_i^T\boldsymbol{\beta}) - \boldsymbol{x}_i^T\boldsymbol{\beta}}{a(\phi)} + c(y_i, \phi) \right]. \tag{2.6}
$$

The GLM for a response variable that follows a Poisson distribution is

$$
\ell(\boldsymbol{\theta}, \phi; \boldsymbol{y}) = \sum_{i=1}^{n} \left(y_i \boldsymbol{x}_i^T\boldsymbol{\beta} - \exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) - \ln y_i! \right). \tag{2.7}
$$

Descriptions of the method of Maximum Likelihood and Iterated Weighted Least Squares are found in McCullagh & Nelder (1989).

2.2.3 Deviance in Generalized Linear Models

Lee et al. (2006) state that the main use of deviance in GLMs is model comparison, and the analysis of deviance is a generalization of the classical analysis of variance. Also, deviance is used as a measure of lack-of-fit. For a measure of goodness-of-fit, analogous to the residual sum of squares for normal models, two such measures are in common use: the first is the generalized Pearson X^2 statistic, and the second is the log likelihood-ratio statistic, called the Deviance in GLMs. These are, respectively,

$$X^2 = \sum_i \frac{(y_{-}\hat{\mu}_i)^2}{V(\hat{\mu}_i)} \tag{2.8}$$

and

$$D = 2\phi[\ell(y; y) - \ell(\hat{\mu}, y)], \tag{2.9}$$

where ℓ is the log-likelihood of the distribution. For normal models, the scaled deviances X^2/ϕ and D/ϕ are identical and become the scaled residual sum of squares, having an exact χ^2 distribution with $n - p$ degrees of freedom. Usually they are different and asymptotic approximation is used for non-normal distributions.

The deviance is considered to have an advantage as a measure of discrepancy in that it is additive for nested sets of models, leading to likelihood-ratio tests. The χ^2 approximation is usually quite accurate for the differences of deviances even though it could be inaccurate for the deviances themselves. Another advantage of the deviance over the X^2 is that it leads to the best normalizing residuals (Pierce & Schafer, 1986), which will be important in Chapter 4.

2.2.4 Estimating the Generalized Linear Model Dispersion Parameter

Suppose the dispersion parameter, $\phi \neq 1$, as in the overdispersed Poisson distribution. If $c(y_i, \phi)$ from Equation 2.6 is known, then the full likelihood is used to estimate $\boldsymbol{\beta}$ and ϕ jointly. However, usually $c(y_i, \phi)$ is not available, so estimation of ϕ needs special consideration (Lee et al., 2006). For the GLM, ϕ may be estimated using either X^2 or D, divided by the appropriate degrees of freedom. Given the correct model, X^2 is asymptotically unbiased whereas D is not. However, D often has smaller sampling variance so that, in terms of Mean Square Error, neither is uniformly better (Lee & Nelder, 1996).

2.2.5 Generalized Linear Model Deviance Residuals

In GLMs the deviance is given by the sum of deviance components

$$D = \sum d_i, \tag{2.10}$$

where the deviance component d_i is

$$d_i = 2\int_{\hat{\mu}_i}^{y_i} \frac{y_i - s}{V(s)} ds. \tag{2.11}$$

For responses y_i that follow a Poisson distribution, the deviance component is

$$2\left[y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right]. \tag{2.12}$$

Residuals $r = y - \hat{\mu}$ are central in model checking for normal models. Two different types of residuals have been extended for use with GLMs: standardized (Studentized) and deletion residuals. Lee et al. (2006) suggest using standardized residuals from GLMs for checking assumptions about components. Note that $Var(d_i) = \phi(1-q))$, so that a residual with a high leverage tends to have large variance. The standardized residuals are

$$r = \frac{y - \hat{\mu}}{\sqrt{\phi(1-q)}}. \tag{2.13}$$

The standardized Pearson residual is

$$r_p^s = \frac{r_p}{\sqrt{\phi(1-q)}} = \frac{y - \hat{\mu}}{\sqrt{\phi V(\hat{\mu})(1-q)}}. \tag{2.14}$$

The standardized Pearson residual for a Poisson-distributed response is

$$r_p^s = \frac{y - \hat{\mu}}{\sqrt{\phi\hat{\mu}(1-q)}}. \tag{2.15}$$

The standardized deviance residual is

$$r_d^s = \frac{r_D}{\sqrt{\phi(1-q)}} = \frac{\text{sign}(y - \hat{\mu})\sqrt{d}}{\sqrt{\phi(1-q)}}. \tag{2.16}$$

Pierce & Schafer (1986) state that the deviance residuals give a good approximation to normality for all GLM distributions, so normal probability plots can be used for model checking. Besides the use of normal probability plot model checking, Nelder (1990) describes the use of the standardized residuals in plots against the fitted values on the constant-information scale, and the plot of the absolute residuals.

The desirable models have Studentized residuals that follow a normal distribution, and this quality is exploited in determining the power on the mean-variance relationship. However, the Studentized residuals are not often normally distributed. The power normal distribution is a distribution with useful properties for describing GLM residuals, and a discussion follows.

2.2.6 Power Normal Distribution

Studentized deviance residuals are assumed to be normal, thought this is often not the case. A common approximation is the power normal (PN) distribution. The PN is useful to characterize GLM residuals when they are approximately normally distributed, as well as when they are skewed, particularly, right skewed.

Power normal distribution parameters are usually estimated under the assumption that the transformed distribution is normal, though actually it is a truncated normal distribution. This typically is not a problem when the intent is to achieve an approximately normal distribution from the transformed data. The estimators of the power normal distribution percentiles, for example, based on the likelihood estimation method, are not consistent estimators. Therefore, it is necessary to find unbiased and consistent parameter estimates and functions of the the parameter estimates when the parameters are determined under the assumption that the transformed distribution is a normal distribution.

Suppose X is a random variable with support on the positive real numbers and Y defined as

$$Y = \begin{cases} \frac{X^\lambda - 1}{\lambda} & ,\lambda \neq 0 \\ \ln(X) & ,\lambda = 0 \end{cases} \tag{2.17}$$

where λ is the transformation parameter (Box & Cox, 1964). The inverse of the normal random variable Y is

$$X = \begin{cases} (\lambda Y + 1)^{\frac{1}{\lambda}} & ,\lambda \neq 0 \\ \exp(Y) & ,\lambda = 0 \end{cases} \tag{2.18}$$

Y is more accurately represented as a truncated normal (TN) distribution than a normal distribution, thus,

$$Y = \begin{cases} \mathcal{TN}(\mu_Y, \sigma_Y^2, -\frac{1}{\lambda}) & ,\lambda \neq 0 \\ \mathcal{N}(\mu_Y, \sigma_Y^2) & ,\lambda = 0, \end{cases} \tag{2.19}$$

where $\frac{1}{\lambda}$ is the left or right truncation value. The probability distribution function (pdf) of Y is

$$g\left(Y \mid \mu, \sigma^2, -\frac{1}{\lambda}\right) = \frac{1}{K(T)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(Y-\mu)^2\right]. \tag{2.20}$$

The constant $K(T)$ is

$$K(T) = \begin{cases} \Phi[\text{sgn}(\lambda)T] & ,\lambda \neq 0 \\ 1 & ,\lambda = 0, \end{cases} \tag{2.21}$$

where Φ is the cdf of the standard normal distribution and $T = \frac{1}{\lambda\sigma} + \frac{\mu}{\sigma}$, which makes $K(T)$ the normalizing constant corresponding to the left or right point of truncation. Therefore, with only positive real number support for a random variable X,

$$Y(\lambda) = \begin{cases} \frac{X^{\lambda}-1}{\lambda} \sim \mathcal{TN}(\mu, \sigma^2, -\frac{1}{\lambda}) & ,\lambda \neq 0 \\ \ln(X) \sim \mathcal{N}(\mu, \sigma^2) & ,\lambda = 0. \end{cases} \tag{2.22}$$

Redefine the random variable X such that $X \sim \mathcal{PN}(\lambda, \mu, \sigma)$, i.e., X follows a power normal distribution. The pdf is

$$f\left(X \mid \lambda, \mu, \sigma^2\right) = \frac{1}{K(T)} \frac{1}{\sqrt{2\pi\sigma^2}} X^{\lambda-1} \exp\left[-\frac{1}{2\sigma^2}(p_\lambda(X) - \mu)^2\right], X > 0, \tag{2.23}$$

where $K(T)$ is as above, and

$$p_\lambda(X) = Y. \tag{2.24}$$

The log-likelihood function of $X \dot{\sim} PN(\lambda, \mu, \sigma^2)$ is given by

$$\ell(\lambda, \mu, \sigma^2 \mid \boldsymbol{y}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i(\lambda) - \mu)^2 + (\lambda-1)\sum_{i=1}^{n} y_i - n\ln K(T). \tag{2.25}$$

The parameter estimation procedure proposed by Box & Cox (1964) assumes that $K(T) = 1$, and constructs the following profile likelihood for λ:

$$l^*(\lambda) = -\frac{n}{2}\ln\hat{\sigma}_\lambda^{*2} + (\lambda - 1)\sum_{i=1}^{n} y_i - \frac{n}{2}(\ln 2\pi + 1), \tag{2.26}$$

where

$$\hat{\sigma}_\lambda^{*2} = \frac{1}{n}\sum_{i=1}^{n}(y(\lambda) - \hat{\mu}_\lambda^*)^2, \tag{2.27}$$

and

$$\hat{\mu}_\lambda^* = \frac{1}{n}\sum_{i=1}^{n} y_i. \tag{2.28}$$

The parameter λ is estimated via MLE. The assumption that $K(T) = 1$ has that Y_i follows a normal distribution (Maruo, Shirahata & Goto, 2011), but Y_i is strictly speaking a truncated normal. Therefore the estimators of λ, μ, and σ^2, based on MLE, are not consistent estimators, and the estimators of the functions of these parameters are not consistent (Maruo et al., 2011). To obtain consistent estimators of these parameters, Maruo et al. (2011) use a Newton-Raphson algorithm that accounts for $K(T)$ in the form of a profile likelihood:

$$l_n(\lambda) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \hat{\sigma}^2_\lambda - \frac{1}{2}\hat{\sigma}^2_\lambda \sum_{i=1}^{n}(y_i(\lambda) - \hat{\mu}_\lambda)^2 + (\lambda - 1)\sum_{i=1}^{n}\ln y_i - n\ln K(T(\lambda)). \tag{2.29}$$

However, Freeman & Modarres (2006a) and Freeman & Modarres (2006b) provide consistent moment estimators in which $K(T) \neq 1$ is possible. A series estimate for the moments, after Freeman & Modarres (2006a) and Freeman & Modarres (2006b), is now discussed.

2.2.7 Power Normal Parameter Estimation

As the MLEs of the parameters of the PN distribution are not asymptotically consistent (Maruo et al., 2011), Freeman & Modarres (2006a) and Freeman & Modarres (2006b) provide useful forms for the moments of a PN distribution. The r^{th} moment of X, the pre-transform random variable, is

$$\mathcal{E}X^r = \begin{cases} \int_{-\frac{1}{\lambda}}^{\infty}(\lambda y + 1)^{\frac{r}{\lambda}}\phi\left(\frac{y-\mu}{\sigma}\right)\frac{dy}{d\sigma} & ,\lambda > 0 \\ \exp(r\mu + r^2\sigma^2/2) & ,\lambda = 0. \end{cases} \tag{2.30}$$

In the case when $\lambda > 0$, Y follows a truncated normal distribution as

$$Y \sim \mathcal{TN}\left(\mu, \sigma^2, -\frac{1}{\lambda}\right). \tag{2.31}$$

When Y has a truncated normal distribution, $X = (\lambda Y + 1)^{\frac{1}{\lambda}}$ can have a power normal distribution. This is useful to obtain a mean and variance for X, which is now developed from Freeman & Modarres (2006a) and Freeman & Modarres (2006b).

Let $S(y) = (\lambda y + 1)^{\frac{r}{\lambda}}$. Expand $S(y)$ in a power series around μ to obtain

$$S(y) = \sum_{k=0}^{\infty}\frac{1}{k!}S^{(k)}(\mu)\ (y - \mu) \tag{2.32}$$

where

$$S^{(k)}(y) = (\lambda y + 1)^{\frac{r}{\lambda} - k} \prod_{l=1}^{k-1} (r - l\lambda). \tag{2.33}$$

Lemma 2.2.1 *Let $X \sim \mathcal{PN}(\lambda, \mu, \sigma^2)$. If*

$$Y = \frac{X^\lambda - 1}{\lambda} \tag{2.34}$$

and

$$Z = \frac{Y - \mu}{\sigma}, \tag{2.35}$$

then

$$\mathcal{E}X^r = \begin{cases} \sum_{k=0}^{\infty} \frac{1}{k!} S^{(k)}(y) \sigma^k \mathcal{E} Z^k & , \lambda > 0 \\ \exp(r\mu + r^2\sigma^2/2) & , \lambda = 0, \end{cases} \tag{2.36}$$

where $Z \sim \mathcal{TN}(0, 1, T)$ and

$$\mathcal{E}Z^k = \frac{\phi(T)}{1 - \Phi(T)} H_{k-1}(T) + R, \tag{2.37}$$

for T as in Equation 2.21, R a polynomial of degree $k - 2$ in Z, and H_{k-1} is the $(k-1)^{th}$ Chebyshev-Hermite polynomial.

When Y follows an approximately normal distribution,

$$\mathcal{E}(Y - \mu)^r = \begin{cases} \frac{\sigma^k k!}{s^{\frac{k}{2}} (\frac{k}{2})!} & , k \text{ even} \\ 0 & , k \text{ odd}, \end{cases} \tag{2.38}$$

this leads to an infinite series moment generating function, as per the Lemma 2.

Lemma 2.2.2 *Let $X \sim \mathcal{PN}(\lambda, \mu, \sigma^2)$, $\lambda \neq 0$ and $Y \sim \mathcal{N}(\mu, \sigma^2)$, then*

$$\mathcal{E}X^r = \sum_{\text{even } k \geq 0} \frac{\sigma^k k!}{s^{\frac{k}{2}} \left(\frac{k}{2}\right)!} S^{(k)}(y). \tag{2.39}$$

Hence,

$$\mathcal{E}X^r = \begin{cases} \sum_{\text{even } k\geq 0} \frac{\sigma^k k!}{s^{\frac{k}{2}}\left(\frac{k}{2}\right)!}(\lambda Y+1)^{\frac{r}{\lambda}-k}\prod_{l=1}^{k-1}(r-l\lambda) & ,\lambda \neq 0 \\ \exp(r\mu + r^2\sigma^2/2) & ,\lambda = 0. \end{cases} \tag{2.40}$$

The series approximation of the first moment, μ, is

$$\mu = \mathcal{E}X^1 = \begin{cases} (\lambda Y+1)^{\frac{1}{\lambda}}(1-\lambda^2) & ,\lambda \neq 0 \\ \exp(\mu + \sigma^2/2) & ,\lambda = 0, \end{cases} \tag{2.41}$$

the calculation of which is found found in Appendix A.

The properties of the PN described above will be used in the development of GLMs that depend on random effects.

2.3 Quasi-Generalized Linear Models and Quasi-Likelihood

The discussion of GLMs thus far assumed a complete probability density function from the EF was required to obtain estimates of the parameters for a mean model. However, many data sets exist in which a complete probability density function specification is not available. The subset of these data sets for which knowledge only of the first two moments is available, i.e., the mean and the variance, may have tractable mean model estimators, and may be obtained via quasi-likelihood methods.

Wedderburn (1974) introduced a form of the likelihood estimation for data series in which only the mean and variance may be determined. He called this form of the likelihood the quasi-likelihood. The difference between the true likelihood and the quasi-likelihood (QL) is that the true likelihood requires complete probability density function specification, whereas the QL requires only knowledge of the first two moments. Models that may have model parameters estimated by QL still have a random component, a systematic component, and a link function component. These three components and model parameter estimation via the quasi-likelihood result in a relaxed form of the GLM, called a quasi-GLM (QGLM).

2.3.1 Definition of Quasi-Likelihood

Wedderburn (1974) defined a QL function such that, for a single random response $\boldsymbol{Y} = (Y_1,\ Y_2,\ \ldots,\ Y_n)^T$ with mean $\mathcal{E}Y_i = \mu_i$ and variance $Var(Y_i) = a(\phi)V(\mu_i)$, for

μ_i a function of unknown parameters $\boldsymbol{\beta} = (\beta_1,\ \beta_2,\ \ldots,\ \beta_p)^T$, and $V(\mu_i)$ a known function, then

$$\frac{\partial Q(\mu_i; y_i)}{\partial \mu_i} = \frac{y_i - \mu_i}{a(\phi)V(\mu_i)}, \tag{2.42}$$

or equivalently, for some $c(y_i)$,

$$Q(\mu_i; y_i) = \int_i^{\mu} \frac{y_i - \mu_i'}{a(\phi)V(\mu_i)} d\mu_i' + c(y_i). \tag{2.43}$$

For independent observations, the total QL is

$$q(\boldsymbol{\mu}; \boldsymbol{y}) = \sum_i Q(\mu_i; y_i). \tag{2.44}$$

The QL, as characterized by a mean-variance relationship, behaves largely as a true likelihood when ϕ is known (Wedderburn, 1974). Hence,

$$\mathcal{E}\left(\frac{\partial q}{\partial \mu}\right) = 0, \tag{2.45}$$

and

$$\left[\mathcal{E}\left(\frac{\partial q}{\partial \mu}\right)\right]^2 + \mathcal{E}\left(\frac{\partial^2 q}{\partial \mu^2}\right) = 0. \tag{2.46}$$

Further, if the true log-likelihood is $\ell(\mu)$, by the Cramér-Rao lower bound theorem,

$$-\mathcal{E}\left(\frac{\partial^2 q}{\partial \mu^2}\right) = \frac{1}{V(\mu)} \leq -\mathcal{E}\left(\frac{\partial^2 l}{\partial \mu^2}\right), \tag{2.47}$$

with equality if the true likelihood has the EF form. If $a(\phi)$ is not known, the quasi-distribution is generally not in the EF (Wedderburn, 1974).

2.3.2 Quasi-Likelihood Models

With the QL approach, for individual random responses $Y_i,\ i - 1, 2, \ldots, n$, predictors $\boldsymbol{x}_i^T$, and using a known link function $g(\cdot)$, the expected values of the random responses are,

$$\mathcal{E}Y_i = \mu_i = g(\boldsymbol{x}_i^T \boldsymbol{\beta}), \tag{2.48}$$

where $g(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$ is the systematic component operating through a link function $g(\cdot)$. The variances of the random responses are,

$$
\begin{aligned}
Var(Y_i) &= a(\phi)V(\mu_i) && (2.49)\\
&= V(\beta_i, \phi). && (2.50)
\end{aligned}
$$

The estimating equation for $\boldsymbol{\beta}$ is, for $a(\phi) = 1$,

$$
\sum_i x_i \frac{1}{V_i}(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}) = 0, \tag{2.51}
$$

where $V_i = V(\beta_i, \phi)$. The weighted least squares estimate is

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left(\sum_i \frac{\boldsymbol{x}_i \boldsymbol{x}_i^T}{V_i} \right)^{-1} \sum_i \frac{\boldsymbol{x}_i^T \boldsymbol{y}}{V_i} && (2.52)\\
&= (\boldsymbol{X}_i^T \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{y}, && (2.53)
\end{aligned}
$$

where $\boldsymbol{X}$ is the $n \times p$ model matrix $(\boldsymbol{X}_{i1},\ \boldsymbol{X}_{i2},\ \ldots,\ \boldsymbol{X}_{in})$, $\boldsymbol{V}$ is the variance matrix consisting only of $diag(V_i)$, and the response vector $\boldsymbol{y} = (y_1,\ y_2,\ \ldots,\ y_n)^T$.

The QL extends the standard GLM by allowing a dispersion function, $a(\phi)$, with parameter ϕ to EF models. This extension gives a more flexible and direct modeling of the variance function (Lee et al., 2006). The QL estimators may be obtained using IWLS.

2.3.3 Deviance in Quasi-Generalized Linear Models

Wedderburn's QL function for an individual observation was given by Equation 2.42 as

$$
\frac{\partial Q(\mu_i; y_i)}{\partial \mu_i} = \frac{y_i - \mu_i}{a(\phi)V(\mu_i)}. \tag{2.54}
$$

The deviance function, which measures the discrepancy between the observation and its expected value, is obtained from the analogue of the log-likelihood-ratio statistic

$$
\begin{aligned}
D(y_i; \mu_i) &= -2[Q(y_i; \mu_i) - Q(y_i; y_i)] && (2.55)\\
&= -2 \int_{y_i}^{\mu_i} \frac{y_i - \mu_i}{V(\mu_i)}. && (2.56)
\end{aligned}
$$

Wedderburn (1974) and McCullagh (1983) show that quasi-likelihoods, and their associated maximum quasi-likelihood estimates have many properties analogous to those of likelihoods and their associated maximum likelihood estimates. In particular, the maximum quasi-likelihood estimate $\hat{\boldsymbol{\beta}}$ is asymptotically normal with mean $\boldsymbol{\beta}$, and asymptotic covariances may be derived in the usual fashion from the second derivative matrix of Q. Further, if H_A and H_B are two nested hypotheses for comparing models A and B of dimension $A < B$, then, under H_A, the change in deviance

$$D(\hat{\mu_B};\hat{\mu_A}) = D(y;\hat{\mu_A}) - D(y;\hat{\mu_B}) \tag{2.57}$$

has an asymptotic χ^2_{B-Z} distribution.

2.3.4 Estimating the Quasi Generalized Linear Model Dispersion Parameter

Wedderburn relaxes the assumption of a known variance function of Y_i by allowing an unknown constant of proportionality ϕ, so that $Var(Y_i) = \phi V(\mu_i)$, where $a(\phi_i) = \phi$. The introduction of the dispersion parameter does not alter the estimation of the regression coefficients $\boldsymbol{\beta}$. However, ϕ does appear as a scale factor in the asymptotic distributions described above, and for these purposes an estimate is required. Wedderburn suggested the bias-corrected mean X^2 statistic

$$\begin{aligned}\hat{\phi} &= \frac{X^2}{n-q} &(2.58)\\ &= \frac{1}{n-q}\sum\frac{(y_i-\hat{\mu}_i)^2}{V(\hat{\mu}_i)}. &(2.59)\end{aligned}$$

Many of the ideas about, and procedures for fitting generalized linear models can be extended when likelihoods are replaced by quasi-likelihoods (Nelder & Pregibon, 1987). One remaining problem concerns the comparison of difference variance functions on the same data set. As the variance function determines the units of measurement for $D(y;\hat{\mu})$ and X^2, differencing these discrepancy measures across variance functions is not possible. To assess variance function it is necessary to extend the definition of quasi-likelihood, which is provided by the extended QL given in the next section.

2.4 Extended Quasi-Generalized Linear Model

In applications for which the generalized linear model is well suited, there are several circumstances when overdispersion may be present. Williams (1982) and Anderson (1988) report that the appearance of overdispersion may occur as the result of unmeasured covariates or factors in the model. Problems may also occur if the functional form of the systematic component is incorrectly specified. The presence of outliers may also give the appearance of overdispersion. The removal of such observations may reduce the amount of overdispersion or remove the overdispersion altogether. An improper selection of a link function may also appear to exhibit overdispersion. Sparseness of the data is another potential cause that can lead to overdispersion. Of particular interest here in the incorrect specification of the distribution or the first and second moments of the random effects. An incorrect specification is considered yet another cause of overdispersion.

Several approaches to modeling overdispersion were outlined by Wilson (1989), Anderson (1988) and Jorgensen (1987) to name a few. One approach examines overdispersion as it relates to the random component of a GLM by constructing models where parameters may vary according to a known distribution, for example, Griffiths (1973), Williams (1975), Crowder (1978), Koehler & Wilson (1986), and Wilson & Koehler (1991); or an unknown distribution, for example, Williams (1982), Breslow (1984), and Wilson (1989). Others have modeled overdispersion by introducing a linear predictor for the dispersion via likelihoods, for example Efron (1986), Aitkin (1987), and Smyth (1989); extended quasi-likelihoods, e.g. Nelder & Pregibon (1987) and McCullagh & Nelder (1989); and pseudo-likelihoods, e.g., Carroll & Ruppert (1982). Yanez (1993) generalized the extended quasi-likelihood (EQL) function, where the dispersion parameters are modeled with respect to covariates in a manner similar to that of the mean parameters without the specification of a likelihood equation.

The QL has no provision to estimate the dispersion parameter ϕ, which disallows treatment of overdispersion; i.e., if $\phi > 1$, then $Var(Y_i) = \phi V(\mu_i)$ is indeterminate. However, for independent observations and using Equation 2.42,

$$\frac{\partial}{\partial \mu_i} q(\mu_i; y_i) = \frac{y_i - \mu_i}{\phi V(\mu_i)}, \tag{2.60}$$

Wedderburn (1974) uses the method of moments to find ϕ as

$$Var\left(\frac{y_i - \mu_i}{\sqrt{V(\mu_i)}}\right) = \phi, \tag{2.61}$$

and a consistent estimate is

$$\hat{\phi} = \frac{1}{n-p} \sum_i \frac{(y_i - \mu_i)^2}{V(\mu_i)}, \tag{2.62}$$

where μ_i is evaluated from the estimated parameters $\boldsymbol{\beta}$, and p is the number of these parameters in the model.

Nelder & Pregibon (1987) have proposed an Extended Quasi-Likelihood (EQL) for estimating ϕ, thus giving an extended GLM (EQGLM). They give an approximate log-likelihood of a single response y_i as

$$Q_i^+(\mu_i, \phi; y_i) = -\frac{1}{2} \ln\left[2\pi\phi v(y_i)\right] - \frac{1}{2\phi} D(y_i, \mu_i), \tag{2.63}$$

where $D(y_i, \mu_i)$ is as in Equation 2.55. The total EQL is denoted

$$q^+ = \sum_i Q_i^+. \tag{2.64}$$

Nelder & Pregibon (1987) call this the extended quasi-likelihood. This is exact if $y_i \sim \mathcal{N}(\mu_i, \sigma^2)$. The approximation is reasonable if the likelihood of y_i is regular.

The EQL has a deviance statistic that is $\phi \cdot \chi_1^2$-variate, which is a gamma distribution with mean ϕ and variance $2\phi^2$. This is equivalent to assuming that the deviance residual

$$r_d^s \equiv \mathrm{sign}(y_i - \mu_i)\sqrt{d_i} \sim \mathcal{N}(\cdot, \cdot). \tag{2.65}$$

For one parameter EFs such as

$$\ln f(\mu_i; y_i) = y_i\theta - b(\theta) + c(y_i), \tag{2.66}$$

the deviance residual has been shown to be the normalizing transformation (Pierce & Schafer, 1986). In simple problems with a single dispersion parameter, the EQL allows a GLM to estimate the dispersion parameter using the deviance as the data. Using the deviance as data, which was discussed above. Unlike the PL, EQL forms the basis of the joint modeling of structured mean and variance parameters, both of which are in the GLM framework.

2.5 Generalized Extended Quasi-Generalized Linear Model

Generalized Extended Quasi-Likelihood (GEQL) is used to estimate the GLM parameters of the mean and variance when the data are over dispersed. Overdispersion can result from an insufficient number of covariates, incorrect link specification, or mis-specified response variable distribution or first and second moments. Some remedies include, for small overdispersion, a known or estimated inflation parameter (ϕ). For large overdispersion, modeling the variance function $V(\mu_i)$ is given by Nelder & Pregibon (1987) and Carroll & Ruppert (1982). For both methods, assumptions are made of the first and second moments for the mean model and a gamma distribution is implicitly assumed for the dispersion parameters.

Mean Model Components	Dispersion Model Components
$Y_i \sim (\mu_i, (\psi - 1)\phi_i^{\psi-1} V_\tau(\mu_i)^\tau)$	$d_i \sim (\phi_i, 2\phi_i^\psi)$
$\eta_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$	$\eta_{di} = \boldsymbol{g}_{di}^T \boldsymbol{\gamma}$
$\eta_i = g(\mu_i)$	$\eta_{di} = g_d(\phi_i)$

Figure 2.1 The GLM structures for the GEGLMs include a mean model and a dispersion model.

The GEQL, Q^*, is similar to Q^+ in that it does not require full distributional assumptions for the mean model, i.e., first and second moments for both the mean model and the dispersion model. EQL is a limiting form of GEQL given by Yanez (1993)

$$\lim_{\psi \to 2} Q^* = Q^+, \tag{2.67}$$

making Q^* a linear function of Q^+, similar to Nelder & Pregibon (1987) in which Q^+ is a linear function of Q.

In the definition of Q^+, the form of the variance function for dispersion, namely, $V_D(\phi_i) = \phi_i^2$ is implicitly assumed as a gamma random variable. This is only approximately conrrect for non normal y_i. Q^* adjusts for this via the parameter ψ. The variance function in Q^* assumes only a power form indexed by a parameter ψ so that

$$V_D(\phi_i) = \phi_i^\psi, \quad \psi > 1, \tag{2.68}$$

with $\psi = 2$ a special case. Thus, less restrictive extra variation in the mean and dispersion can be modeled through ψ as it affects the variances of both y_i and d_i respectively.

The generalization of the EQL to GEQL allows for the fitting of models where only the form of the first two moments of the mean response and the first two moments of the dispersion are specified. A broader class of models is available in the joint modeling framework where only the form of the first two moments of the mean the dispersion are needed to fit these two models. By the definition of the GEQL function, information in the data can assist in selecting the form of the variance function in the dispersion model in a manner similar to that proposed for the EQL function of Nelder & Pregibon (1987).

2.6 Joint Generalized Linear Model Estimation

Suppose there are two interlinked models for the mean and dispersion based on the observed responses $y_1,\ y_2,\ \ldots,\ y_n$, deviance $d_1,\ d_2,\ \ldots,\ d_n$, with model components as in Figure 2.2. These interlinked models were called Joint Generalized Linear Models (JGLMs) by Nelder & Pregibon (1987). Note $\mathcal{E}d_i = \phi_i$ and $Var(d_i) = 2\phi_i^2$ are the dispersion model mean and variance and hence have the same mean-variance relationship as a gamma distribution. The dispersion model parameters are no longer constant, but can vary with the mean model parameters. Figure 2.2 displays the JGLM mean and dispersion submodels structures.

Mean Model Components	Dispersion Model Components
$Y_i \sim (\mu_i, \phi_i V_\tau(\mu_i))$	$d_i \sim (\phi_i, 2\phi_i^2)$
$\eta_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$	$\eta_{di} = \boldsymbol{g}_{di}^T \boldsymbol{\gamma}$
$\eta_i = g(\mu_i)$	$\eta_{di} = g_d(\phi_i)$

Figure 2.2 The GLM structures for the JGLMs include a mean model and a dispersion model.

The consequence is that the dispersion model values are now needed in the IWLS algorithm for estimating the regression parameters, and that these values have a direct effect on the estimates of the regression parameters. The EQL q^+ yields a fitting algorithm, which can be computed iteratively using two interconnected IWLS procedures:

1. Given $\hat{\gamma}$ and the dispersion estimates $\hat{\phi}_i$'s, use IWLS to update $\hat{\beta}$ for the mean model;
2. Given $\hat{\beta}$ and the estimated means $\hat{\mu}_i$'s, use IWLS to update $\hat{\gamma}$ with the deviances as data;
3. Iterate step 1 and 2 until convergence.

For the mean model in the first step, the updated equation is

$$(\boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{X})\boldsymbol{\beta} = \boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{z}, \tag{2.69}$$

where

$$z_i = \eta_i + \frac{\partial \eta_i}{\partial \mu_i}(y_i - \mu_i) \tag{2.70}$$

is the adjusted dependent variable and $\boldsymbol{\Sigma}$ is a diagonal matrix with elements

$$\Sigma_{ii} = \phi_i \left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 V(\mu_i). \tag{2.71}$$

Use $\phi_i \equiv \phi$ as a starting value, so no actual value of ϕ is needed. Thus, this GLM is specified by a response variable $\boldsymbol{y}$, a variance function $V(\cdot)$, a link function $g(\cdot)$, a linear predictor η_i, and a prior weight $1/\phi$.

For the dispersion model, first compute the observed deviances $d_i = d(y_i, \hat{\mu}_i)$ using the estimated means. Let $d_i^* = d_i/(1-q_i)$ with $q_i = 0$. The REML adjustment is the GLM leverage for q_i described below in the REML procedure for QL models. The updating formula for $\hat{\gamma}$ is

$$\boldsymbol{G}^T \boldsymbol{\Sigma_d}^{-1} \boldsymbol{G\gamma} = \boldsymbol{G}^T \boldsymbol{\Sigma_d}^{-1} \boldsymbol{z_d}, \tag{2.72}$$

where the dependent variables are defined as

$$z_{di} = \eta_{di} + \frac{\partial \eta_{di}}{\partial \phi_i}(d_i^* - \phi_i) \tag{2.73}$$

and $\boldsymbol{\Sigma_d}$ is a diagonal matrix with elements

$$\Sigma_{d_i} = 2\left(\frac{\partial \eta_{di}}{\partial \phi_i}\right)^2 \phi_i^2. \tag{2.74}$$

This GLM is characterized by a response d_i^* for the deviance residual, a gamma error, a link function $g_d(\cdot)$, a linear predictor η_{di}, and a prior weight $\frac{1}{2}(1-q)$. At convergence, the standard error of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ can be computed. If the GLM deviance is used, this algorithm yields estimators using the EQL, while the Pearson deviance that results are those from the PL.

The deviance components d_i^* become the responses for the dispersion GLM. Then the reciprocal of the fitted values from the dispersion GLM provide the prior weights of the next iterations for the mean GLM. The resulting back-and-forth algorithm is very fast to converge. This means that all the inferential tools used for GLMs can be used for the GLMs for the dispersion parameters.

2.7 Hierarchical Generalized Linear Models

A class of GLM that combines both fixed effects and random effects is the Hierarchical Generalized Linear Model (HGLM). The random components may come from an arbitrary distribution, in one case, the distribution is conjugate to that of the response vector $\boldsymbol{y}$. An additional random term in the fixed effects mean submodel accounts for clustering in data. For estimation of fixed effects and random effects, the corresponding joint likelihood is generalized by Lee & Nelder (1996) to a hierarchical or h-likelihood.

Let y_{ij} be the observed response variable for $i = 1, \ldots, n$, $j = 1, \ldots, m$, and u_i be the unobserved random component. Lee & Nelder (1996) define the following HGLM, represented in Figure 2.3, utilizing the conditional log-likelihood for y_{ij} given u_i with the EF form

$$\ell(\theta, \phi; y_{ij} \mid u_i) = \frac{1}{a_{ij}(\phi)}[y_{ij}\theta_{ij} - b(\theta_i)] + c(y_{ij}, \phi_{ij}), \tag{2.75}$$

where θ_{ij} denotes the canonical parameter and ϕ is the dispersion parameter. Let μ_{ij} be the conditional mean of y_{ij} given u_i, where $\eta_{ij} = g(\mu_{ij})$, i.e., $g(\cdot)$ is the link function for the GLM describing the conditional distribution of y_{ij} given u_i. The linear predictor η_{ij} has the form given in Figure 2.3 for some strictly monotonic function $v_i = v(u_i)$. Modeling η_{ij} involves fixed effects mean modeling and dispersion modeling for v_i, which describes the overdispersion.

Effect	Mean Model Components	Dispersion Model Components
Fixed	$Y_{ij} \mid u_i \sim (\mu_{ij}, a_{ij}(\phi)V(\mu_{ij}))$ $\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{v}$ $\eta_{ij} = g(\mu_{ij})$	Not applicable
Random	$u_i \sim (\mu_{Ri}, \zeta V_R(\mu_{Ri}))$ $v_i = g_R(u_i)$	Not applicable

Figure 2.3 The GLM structures for the HGLMs include a mean model, a dispersion model, and a random effects model.

As an example, suppose that the distribution of y_{ij} given u_i is Poisson with mean $\mu_{ij} = \mathcal{E}(y_{ij} \mid u_i) = \mu_{ij}u_i$. With a a log-link, $v_i = \ln u_i$. If the distribution of u_i is gamma, then the model is called the Poisson-gamma HGLM, with v_i has the log-gamma distribution.

2.7.1 Hierarchical Likelihood

The systematic component of the model specification that assumes a particular mean and variance function for an individual response y_{ij} conditional on a random effect u_i is

$$\eta_{ij} = g(\mu_{ij}) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{v}. \tag{2.76}$$

The outcome y_{ij} is independently distributed with mean μ_{ij} and variance $\phi_{ij} V(\mu_{ij})$, where $a_{ij}(\phi) = \phi_{ij}$.

The marginal variance of y_{ij} is

$$Var(y_{ij} \mid v_i) = \mathcal{E}[Var(y_{ij} \mid v_i)] + Var[\mathcal{E}(y_{ij} \mid v_i)], \tag{2.77}$$
$$= \phi_{ij}\mathcal{E}[g(\mu_{ij})] + Var(\mu_{ij}). \tag{2.78}$$

Overdispersion may appear in both the right-hand side terms, the first term accounts for the contribution of the dispersion family, and the second term is the contribution of the random effects.

The development of the likelihood for HGLMs is analogous to the development of likelihoods for GLMs, QGLMs, and EQGLMs. Therefore, the start of development for HGLMs is with random effects that follow a distribution from the EF, as given by Lee & Nelder (1996). Consider a HGLM as presented in Figure 2.3, with fully specified distributions for a response vector $\boldsymbol{y} \mid \boldsymbol{u}$. Let the fixed effects dispersion parameter vector $a(\boldsymbol{\phi}) = \boldsymbol{\phi}$, and the random effects mean model parameter vector be $\boldsymbol{\zeta}$. Note that $\boldsymbol{v} = v(\boldsymbol{u})$ is the response canonical transformation applied to $\boldsymbol{u}$. The joint likelihood for normal models (Henderson, 1975) are generalized to form the h-likelihood, denoted by h, is

$$h = \ell(\boldsymbol{\theta}; \boldsymbol{y} \mid \boldsymbol{u}) + \ell(\boldsymbol{\theta}_R; \boldsymbol{v}), \tag{2.79}$$

where $\boldsymbol{\theta}$ is the canonical parameter vector for the conditional response distribution and $\boldsymbol{\theta}_R$ is the parameter vector for the distribution of $\boldsymbol{v}$ (Lee & Nelder, 1996). The first term of Equation 2.79 is the log-likelihood of the conditional response distribution, and the second term is the log-likelihood of the random effects distribution of $\boldsymbol{v}$. The vector $\boldsymbol{v}$ is estimated as a random effects mean model parameter. As $\boldsymbol{\theta}_R$ is a parameter in the distribution of $\boldsymbol{v}$, $\boldsymbol{\theta}_R$ is considered to be a dispersion parameter vector.

The maximum h-likelihood estimates are derived from maximizing the h-

likelihood function by solving

$$\frac{\partial h}{\partial \boldsymbol{\beta}} = 0 \tag{2.80}$$

$$\frac{\partial h}{\partial \boldsymbol{v}} = 0. \tag{2.81}$$

These solutions have the advantage that integration of the random effects is not needed for estimates of the random effects mean model parameters. From the definition of the h-likelihood, Equation 2.79, it is easy to see that the maximum h-likelihood estimates (MHLEs) for $\boldsymbol{\beta}$ given $\boldsymbol{u}$ are obtained by the GLM equations with $v(\boldsymbol{u})$ as an offset. As such, the IWLS may be used to obtain the maximum HGLMs (MHLEs). However, an augmented form of the GLM IWLS is used, thereby permitting the simultaneous determination of both the fixed effects and random effects mean model parameters (Lee & Nelder, 2006).

The h-likelihood estimation for an HGLM can be viewed as that for an augmented GLM with the response variables $(\boldsymbol{y}^T, \boldsymbol{\xi}^T)^T$, where $\mathcal{E}(\boldsymbol{Y} \mid \boldsymbol{u}) = \boldsymbol{\mu}$, $Var(\boldsymbol{Y} \mid \boldsymbol{u}) = \boldsymbol{\phi} V(\boldsymbol{\mu})$, $\mathcal{E}(\boldsymbol{\xi}) = \boldsymbol{u}$, and $Var(\boldsymbol{\xi}) = \boldsymbol{\zeta} V_R(\boldsymbol{u})$. These statistics are from the EF individual observation log-likelihoods $\sum[y_{ij}\theta(\mu_{ij}) - b(\theta(\mu_{ij}))]/\phi_{ij}$ for fixed effects mean model, and $\sum[\xi_i\theta(u_i) - b(\theta(u_i))]/\zeta_i$, with ξ_i the pseudo-response of the random effects mean model. The augmented response variables are used with the augmented linear predictor

$$(\boldsymbol{\eta}^T, \boldsymbol{v}^T)^T = \boldsymbol{T}\boldsymbol{\omega}, \tag{2.82}$$

where $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{v}$, $\boldsymbol{v} = g_R(\boldsymbol{u})$, and $\boldsymbol{\omega} = (\boldsymbol{\beta}^T, \boldsymbol{v}^T)^T$ are fixed, unknown parameters and quasi-parameters, and the augmented model matrix is

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Z} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}, \tag{2.83}$$

where $\boldsymbol{I}$ is the $m \times m$ identity matrix.

Given $(\boldsymbol{\phi}, \boldsymbol{\zeta})$, the estimate of the two components $\boldsymbol{\omega} = (\boldsymbol{\beta}^T, \boldsymbol{v}^T)^T$ can be computed by IWLS from the augmented GLM as

$$\boldsymbol{T}_R^T \boldsymbol{\Sigma}_R^{-1} \boldsymbol{T}_R \boldsymbol{\omega} = \boldsymbol{T}_R^T \boldsymbol{\Sigma}_R^{-1} \boldsymbol{z}_{Ra}, \tag{2.84}$$

where $\boldsymbol{z}_{Ra} = (\boldsymbol{z}^T, \boldsymbol{z}_R^T)^T$ and

$$\boldsymbol{\Sigma}_R = \boldsymbol{\Gamma}_R \boldsymbol{W}_{Ra}^{-1}, \tag{2.85}$$

with $\boldsymbol{\Gamma}_R = diag(\phi_1,\ \phi_2,\ \ldots,\ \phi_n, \zeta_1,\ \zeta_2,\ \ldots,\ \zeta_m)$. The adjusted dependent variables $z_{ai} = (z_i, z_{Ri})$ are defined by

$$z_i = \eta_i + (y_i - \mu_i)\frac{\partial \eta_i}{\partial \mu_i}, \tag{2.86}$$

for $i = 1, \ldots, n$, and for $j = 1, \ldots, m$,

$$z_{Rj} = v_j + (\xi_j - u_j)\frac{\partial v_j}{\partial u_j}. \tag{2.87}$$

The iterative weight matrix

$$\boldsymbol{W}_a = diag(\boldsymbol{W}^{-1}, \boldsymbol{W}_R^{-1}) \tag{2.88}$$

has the elements

$$W_i^{-1} = \left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 V(\mu_i) \tag{2.89}$$

and

$$W_{Rj}^{-1} = \left(\frac{\partial v_j}{\partial u_j}\right)^2 V_R(u_j). \tag{2.90}$$

The standard errors of the random effects parameters are obtained as

$$s.e.(\boldsymbol{v}) = \sqrt{diag\{\boldsymbol{\Sigma}_R\}}. \tag{2.91}$$

2.7.2 Double Extended Quasi-Likelihood

An extension to the h-likelihood was proposed by Lee & Nelder (2001) that generalizes the HGLM from distribution specifications for both the fixed effects and random effects, to requiring only the first and second moments. This relaxed distribution specification is called the double extended quasi-likelihood (DEQL), and employs EQLs for both the fixed effects mean model and the random effects mean model parameter estimation. These HGLMs are known as quasi-HGLMs (QHGLMs) and the relationships are as in Figure 2.4.

The DEQL derives from the Exponential Family (EF) such that

$$h^+ = q^+(\theta(\boldsymbol{\mu})), \boldsymbol{\phi}; \boldsymbol{y} \mid \boldsymbol{u}) + q_R^+(\boldsymbol{u}; \boldsymbol{\xi}) \tag{2.92}$$

Effect	Mean Model Components	Dispersion Model Components
Fixed	$Y_{ij} \mid u_i \sim (\mu_{ij}, a_{ij}(\phi)V(\mu_{ij}))$ $\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{v}$ $\eta_{ij} = g(\mu_{ij})$	$d_{ij} \sim (\phi_{ij}, 2\phi_{ij}^2)$ $\eta_{dij} = \boldsymbol{g}_i^T\boldsymbol{\gamma}$ $\eta_{dij} = g_d(\phi_{ij})$
Random	$u_i \sim (\mu_{Ri}, \zeta V_R(\mu_{Ri}))$ $v_i = g_R(u_i)$	$d_{Ri} \sim (\zeta, 2\zeta^2)$ $\eta_{dRi} = \boldsymbol{g}_{dRi}^T\boldsymbol{\delta}$ $\eta_{dRi} = g_{dRi}(\zeta_i)$

Figure 2.4 The GLM structures for the DHGLMs include a mean model, a dispersion model, a random effects model, and a random effects dispersion model.

where

$$q(\theta(\boldsymbol{\mu})), \boldsymbol{\phi}; \boldsymbol{y} \mid \boldsymbol{u}) = -\frac{1}{2}\sum_i \left\{ \frac{d_i}{\phi} + \ln(2\pi\phi V(y_i \mid u_i)) \right\} \tag{2.93}$$

$$= -\frac{1}{2}\sum_i \{d_i + \ln(2\pi V(y_i \mid u_i))\}, \quad \phi_i = 1, \tag{2.94}$$

$$q_R^+(\boldsymbol{u}) = -\frac{1}{2}\sum_i \{d_{Ri} + \ln(2\pi V_R(\xi_i))\}, \tag{2.95}$$

$$d_i = 2\int_{\mu_i}^{y_i} \frac{y_i - s}{s} ds, \tag{2.96}$$

and

$$d_{Ri} = 2\int_{u_i}^{\xi_i} \frac{\xi_i - s}{s^\psi} ds \tag{2.97}$$

are the deviance components of $\boldsymbol{y} \mid \boldsymbol{u}$ and $\boldsymbol{u}$ respectively. The function $q_R(\boldsymbol{u}; \boldsymbol{\xi})$ has the form of an EQL for the quasi-data $\boldsymbol{\xi}$. The $\boldsymbol{\xi} = (\xi_1, \ \xi_2, \ \ldots, \ \xi_m)^T$ is an unobservable random variable whose expected value is the mean parameter $\boldsymbol{u}$, and whose dispersion structures is the same form as that of $\boldsymbol{u}$,

$$\boldsymbol{\xi} \sim (\boldsymbol{u}, \boldsymbol{\zeta} V_R(\boldsymbol{u})). \tag{2.98}$$

Lee et al. (2006) states that the random effects $\boldsymbol{u}$ are treated as fixed parameters of the distribution of $\boldsymbol{\xi}$ after the response vector $\boldsymbol{y}$ has been observed.

Lee & Nelder (2001) showed that h^+ is equivalent to the first-order Laplace approximation of the h-likelihood used when the conditional response distribution and the random effects distribution are known. The DEQL h^+ can be used to estimate the model mean parameters $\boldsymbol{\omega} = (\boldsymbol{\beta}^T, \boldsymbol{v}^T)^T$. These estimates are obtained by setting the first derivatives of h^+ to zero, thus,

$$\frac{\partial h^+}{\partial \beta_j} = \sum_{i=1}^{n} \boldsymbol{x}_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) \left(\frac{y_i - \mu_i}{\phi_i V(\mu_i)}\right) = 0, \tag{2.99}$$

and

$$\frac{\partial h^+}{\partial v_k} = \sum_{i=1}^{n} \boldsymbol{z}_{ik} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) \left(\frac{y_i - \mu_i}{\phi_i V(\mu_i)}\right) + \sum_{i=j}^{m} \boldsymbol{I}_{jk} \left(\frac{\partial u_j}{\partial v_j}\right) \left(\frac{\xi_j - u_j}{\zeta_j V_R(u_j)}\right) = 0, \tag{2.100}$$

where $\boldsymbol{I}_{jk}$ is the $(j,k)^{th}$ element of the identity matrix. To estimate the dispersion parameters, and adjusted profile h^+-likelihood procedure is used. For the DEQL, the adjusted h^+-likelihood is

$$h_a^+ = h^+ - \frac{1}{2} \ln \left[\left| -\frac{1}{2\pi} \left(\frac{\partial^2 h^+}{\partial \boldsymbol{\omega}} \right) \right| \right]. \tag{2.101}$$

Dispersion parameters are estimated using the first derivatives of h_a^+ with current estimates of the mean parameters substituted in. Lee & Nelder (2001) showed that the adjusted profile h^+-likelihood is equivalent to the first-order Laplace approximation of the restricted likelihood used when the marginal response likelihood is known. Thus, for $\phi_i > 0$, $\zeta_j > 0$, $\boldsymbol{g}_{di}\boldsymbol{\gamma} = \gamma$, and $\boldsymbol{g}_{Rdi}\boldsymbol{\delta} = \delta$,

$$\frac{\partial h_a^+}{\partial \gamma} = \gamma \sum_{i=1}^{n} (1 - q_i) \left(\frac{d_i - \phi_i}{\phi_i}\right) = 0, \tag{2.102}$$

and

$$\frac{\partial h_a^+}{\partial \delta} = \delta \sum_{j=1}^{m} (1 - q_{Rj}) \left(\frac{d_{Rj} - \zeta_j}{\zeta_j}\right) = 0, \tag{2.103}$$

where $\boldsymbol{q}_h = (\boldsymbol{q}^T, \boldsymbol{q}_R^T)^T$ are the diagonal leverage estimates from the mean model

$$\boldsymbol{q}_h = diag(\boldsymbol{T}(\boldsymbol{T}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{T})^{-}\boldsymbol{T}^T\boldsymbol{\Sigma}^{-1}). \tag{2.104}$$

The term $\boldsymbol{T}$ is defined in Equation 2.83, $\boldsymbol{\Sigma}$ is defined in Equation 2.85, d_i is defined in Equation 2.96, and d_{Ri} is defined in Equation 2.97. Note that the dispersion model score equations are chosen to have the same form as the mean model score equations. The iteration algorithm then alternates between the h^+ mean model parameters and the h_a^+ dispersion model parameters.

2.7.3 Iterated Weighted Least Squares for Double Extended Quasi-Likelihood

The double extended model can be fitted by solving IWLS estimating equations of three GLMs as follows:

1. Given (ϕ_i, ζ_j), the two components $\boldsymbol{\omega} = (\boldsymbol{\beta}^T, \boldsymbol{v}^T)^T$ can be estimated by Equation 2.84 which are the IWLS equations for the augmented GLM in Subsection 2.7.1.
2. Given $(\boldsymbol{\omega}, \boldsymbol{\zeta})$, γ estimates ϕ_i by the IWLS equations
$$\boldsymbol{G}^T \boldsymbol{\Sigma}_d^{-1} \boldsymbol{G}\boldsymbol{\gamma} = \boldsymbol{G}^T \boldsymbol{\Sigma}_d^{-1} \boldsymbol{z}_d, \tag{2.105}$$
where $\boldsymbol{\Sigma}_d = \boldsymbol{\Gamma}_d \boldsymbol{W}_d^{-1}$ with $\boldsymbol{\Gamma}_d = diag(2/(1-q_i))$,
$$q_i = \boldsymbol{x}_i (\boldsymbol{X}^T \boldsymbol{W}_R \boldsymbol{X})^{-1} \boldsymbol{x}_i^T, \tag{2.106}$$
the weight functions $\boldsymbol{W}_d = diag(\boldsymbol{W}_{di})$ are defined as
$$\boldsymbol{W}_{di} = \left(\frac{\partial \phi_i}{\partial \eta_{di}} \right)^2 \frac{1}{2\phi_i^2}, \tag{2.107}$$
and the dependent variables are defined as
$$z_{di} = \eta_{di} + (d_i^* - \phi_i) \frac{\partial \eta_{di}}{\partial \phi_i}, \tag{2.108}$$
with GLM deviance components
$$d_i^* = \frac{d_i}{1 - q_i}, \tag{2.109}$$
and d_i is as in Equation 2.96. This GLM is characterized by a response d^*, gamma error, link function $g_d(\cdot)$, linear predictor $\boldsymbol{G}$ and prior weight $(1-q)/2$.
3. Given $(\boldsymbol{\omega}, \boldsymbol{\phi})$, estimate δ for $\boldsymbol{\zeta}$ by the IWLS equations
$$\boldsymbol{G}_{dR}^T \boldsymbol{\Sigma}_{dR}^{-1} \boldsymbol{G}_{dR} \boldsymbol{\delta} = \boldsymbol{G}_{dR}^T \boldsymbol{\Sigma}_{dR}^{-1} \boldsymbol{z}_{dR}, \tag{2.110}$$
where $\boldsymbol{\Sigma}_{dR} = \boldsymbol{\Gamma}_{dR} \boldsymbol{W}_{dR}^{-1}$ with $\boldsymbol{\delta} = diag(2/(1-q_{dRi})$; $\boldsymbol{W}_{dR} = diag(\boldsymbol{W}_{dRi})$ are defined by
$$\boldsymbol{W}_{dRi} = \left(\frac{\partial \zeta_i}{\partial \eta_{dRi}} \right)^2 \frac{1}{2\zeta_i^2}, \tag{2.111}$$

and the dependent variables are defined as

$$z_{dRi} = \eta_{dRi} + (d^*_{Ri} - \zeta_i)\frac{\partial \eta_{dRi}}{\partial \zeta_i}, \tag{2.112}$$

with GLM deviance components

$$d^*_{Ri} = \frac{d_{Ri}}{1 - q_{dRi}}, \tag{2.113}$$

and d_{Ri} is as in Equation 2.97. The weight q_{dR} extends leverage to HGLMs. This GLM is characterized by a response d^*_{dR}, gamma error, link function $g_{dR}(\cdot)$, linear predictor $\boldsymbol{G}_{dR}$ and prior weight $(1 - q_{dR})/2$.

Let the observed response vector with elements y_{ij} have mean μ_{ij} and $g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{v}$. Given initial values of $\boldsymbol{\beta}^0$ and $\boldsymbol{v}^0$, the EF log-likelihood can be approximated by

$$-\frac{1}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{v})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{v}), \tag{2.114}$$

where $\boldsymbol{Y}$ is a working vector with elements

$$Y_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}^0 + \boldsymbol{z}_i^T\boldsymbol{v}_i^0 + \frac{\partial h}{\partial \mu_{ij}}(y_{ij} - \mu_{ij}^0), \tag{2.115}$$

and $\boldsymbol{\Sigma}$ is a diagonal matrix of the variance of the working vector with elements

$$\Sigma_{ii} = \left(\frac{\partial h}{\partial \mu_{ij}}\right)^2 \phi v(\mu_{ij}^0), \tag{2.116}$$

where $\phi v(\mu_{ij}^0)$ is the conditional variance of y_{ij} given v_i. The derivative $\partial h/\partial \mu_{ij}$ is also evaluated at the current vectors values of $\boldsymbol{\beta}$ and $\boldsymbol{v}$.

If the random effects parameter vector $\boldsymbol{v}$ is assumed normal with mean zero and variance $\boldsymbol{D} = D(\boldsymbol{\theta})$, the log-likelihood has the familiar normal-based formula:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{v}) = -\frac{1}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{v})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{v}) - \frac{1}{2}\ln|\boldsymbol{D}| - \frac{1}{2}\boldsymbol{v}^T\boldsymbol{D}^{-1}\boldsymbol{v}. \tag{2.117}$$

This yields the usual mixed model equations to update $\boldsymbol{\beta}$ and $\boldsymbol{v}$:

$$\begin{pmatrix} \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} + \boldsymbol{D}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{v} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \\ \boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \end{pmatrix} \tag{2.118}$$

The IWLS is used to find the solution. The iteration continues by recomputing $\boldsymbol{Y}$ and $\boldsymbol{\Sigma}$. Hence the computation of estimates in HGLM involves repeated applications of normal-based formulae.

If $\boldsymbol{v}$ is not normally distributed, an extra step is needed first to approximate its log-likelihood by a quadratic form. Let $u_i \overset{iid}{\sim} (\mu_{Ri}, \zeta V_R(\mu_{Ri}))$, $v_i = g_R(u_i)$ for some $g_R(\cdot)$, and $\ell(v_i) = \ln g_R(u_i)$. Using the initial value v_i^0

$$\ln g_R(u_i) \approx \ln g_R(u_i^c) + \frac{1}{2}\ell''(v_i^0)(v_i - v_i^c)^2, \tag{2.119}$$

where

$$v_i^c = v_i^0 - \frac{\ell'(v_i^0)}{\ell''(v_i^0)}. \tag{2.120}$$

Now let $\boldsymbol{D}^{-1} = diag[-\ell''(v_i^0)]$, the Fisher information matrix of $\boldsymbol{v}$ based on $g_R(\cdot)$, and let $\ell'(\boldsymbol{v}^0)$ be the vector of $\ell'(v_i^0)$, so in vector notation

$$\boldsymbol{v}^c = \boldsymbol{v}^0 + \boldsymbol{D}l'(\boldsymbol{v}^0). \tag{2.121}$$

Then,

$$\ell(\boldsymbol{v}) = \ell(\boldsymbol{v}^c) - \frac{1}{2}(\boldsymbol{v} - \boldsymbol{v}^c)^T \boldsymbol{D}^{-1}(\boldsymbol{v} - \boldsymbol{v}^c). \tag{2.122}$$

In the normal case $\boldsymbol{D}$ is the covariance matrix, and $\boldsymbol{v}^c = 0$.

After combining this with the quadratic approximation of $\ell(\boldsymbol{y}|\boldsymbol{u})$, then taking the derivatives with respect to $\boldsymbol{\beta}$ and $\boldsymbol{v}$ and finding the updating equation

$$\begin{pmatrix} \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} + \boldsymbol{D}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{v} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \\ \boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} + \boldsymbol{D}^{-1}\boldsymbol{v}^c \end{pmatrix}. \tag{2.123}$$

This is similar to 2.118, except for the term $\boldsymbol{D}^{-1}\boldsymbol{v}^c$.

As in the normal case, the approximate MLEs of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and $\boldsymbol{v}$ are the joint maximizers of

$$q(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{v}) = \ell(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{v}) - \frac{1}{2}\ln|\boldsymbol{Z}^T\boldsymbol{\Sigma}^{-1}, \boldsymbol{Z} + \boldsymbol{D}^{-1}|. \tag{2.124}$$

To derive an iterative estimation procedure the first term is approximated by a quadratic form of Equation 2.117. However, in contrast with the normal mixed models, because of the dependence of $\boldsymbol{\Sigma}$ on $\boldsymbol{\beta}$ and $\boldsymbol{v}$. An iterative algorithm is:

1. Compute $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{v}}$ given $\boldsymbol{\theta}$ by solving Equation 2.118.
2. Fixing $\boldsymbol{\beta}$ and $\boldsymbol{v}$ at the values $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{v}}$, update $\boldsymbol{\theta}$ by maximizing q.
3. Iterate between steps 1 and 2 until convergence.

The derivative of $\ell(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{v})$ with respect to $\boldsymbol{\beta}$ is

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zv}). \tag{2.125}$$

Combining this with the derivative with respect to $\boldsymbol{v}$ and setting them to zero, gives Equation 2.118. Equation 2.118 suggest the that the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{v}$ may be calculated by computing the marginal variance $\boldsymbol{V}$ or its inverse. Rather, this is given by the iterative back fitting algorithm (Gauss-Seidel method). In the algorithm, $\boldsymbol{\beta}$ and $\boldsymbol{v}$ are computed in as follows:

1. Start with an estimate of $\boldsymbol{\beta}$, for example, the ordinary least-squares estimate
$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}, \tag{2.126}$$
then iterate between steps 2 and 3 below until convergence.
2. Compute a corrected outcome
$$\boldsymbol{y}^c = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{2.127}$$
and estimate $\boldsymbol{v}$ from a random effects model
$$\boldsymbol{y}^c = \boldsymbol{Zv}, \tag{2.128}$$
based on
$$(\boldsymbol{Z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{Z} + \boldsymbol{D}^{-1})\boldsymbol{v} = \boldsymbol{Z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{y}^c. \tag{2.129}$$
3. Recompute a corrected outcome
$$\boldsymbol{y}^c = \boldsymbol{y} - \boldsymbol{ZV} \tag{2.130}$$
and estimate $\boldsymbol{\beta}$ from a fixed effects model
$$\boldsymbol{y}^c = \boldsymbol{X\beta}, \tag{2.131}$$
which updates $\hat{\boldsymbol{\beta}}$ from the solution of
$$(\boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{X})\boldsymbol{\beta} = \boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{y}^c. \tag{2.132}$$

2.8 A Deviance-Based Criterion for Generalized Liner Model Selection

The use of deviance to determine model parameters is most recently explored by Sakate & Kashid (2012). It is important to note that, while the following is useful for model selection and link function identification, there is no component that allows for the characterization of a random effects power function mean-variance relationship exponent.

Sakate & Kashid (2012) have devised a deviance-based criterion (DBC) for GLM selection. The DBC is obtained by penalizing the difference between the deviance of the fitted model and the full model. Under certain conditions, DBC is shown to be a consistent model selection criterion, where the selected model asymptotically equals the optimal model relating response and predictors. The DBC is used to identify models with the appropriate mixes of predictors, and to help identify appropriate link functions.

The deviance is a function of the data only and is used to define a statistic for model selection. Let $D(\boldsymbol{y}, \hat{\boldsymbol{\beta}})$ denote the deviance of the full model. If the difference in deviance of a model, say M_a and the full model $D(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_a - D(\boldsymbol{y}, \hat{\boldsymbol{\beta}})$ is small, then the model M_a can be regarded as good as the full model for prediction. This does not serve for model selection as for the model M_a such that $a_* \supset a$, the difference is smaller than that for the model M_a, and is zero when a corresponds to the full model. Thus, it is difficult to identify an optimal model. A good model selection criterion should take into account goodness-of-fit as well as the complexity of the model. The number of parameters, p_a, is a natural measure of model complexity. Therefore Sakate & Kashid (2012) define a model selection criterion for GLMs based on the penalized difference between deviance of model M_a and the full model. The DBC can be expressed as

$$DBC(M_a) = \frac{1}{\phi} D(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_a - D(\boldsymbol{y}, \hat{\boldsymbol{\beta}}) - (k - p_a) + C(n, P_a), \tag{2.133}$$

where ϕ is the dispersion parameter and is either known or estimated. Under normality of the response and $C(n, p_a) = p_a$, the criterion in Equation 2.133 is equivalent to Mallow's C_p.

The DBC defined in Equation 2.133 may be used for model selection when the true link function is known. If the link function is unknown and is to be selected from a finite set of continuous monotone link functions, the DBC in Equation 2.133 cannot be used as it contains the full model deviance. Let $g \in \mathcal{G}$ be on of the finitely many link functions in $\mathcal{G}$, and let M_b be the GLM when the link function is g. Denote the mean of the response of the model M_b by $\boldsymbol{\mu}_b$, and the regression parameter vector by $\boldsymbol{\beta}_b$. Then the DBC for model selection when the link function is unknown is defined as

$$DBC(M_b) = \frac{1}{\phi} D(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_b{}^g) - D(\boldsymbol{y}, \hat{\boldsymbol{\beta}}^{g*}) - (k - p_a) + C(n, P_a), \tag{2.134}$$

where $D(\boldsymbol{y}, \hat{\boldsymbol{\beta}}^g)$ is the deviance of the full model corresponding to the link function $g* \in \mathcal{G}$ such that

$$\min_{g \in \mathcal{G}} D(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_{ba}) = D(\boldsymbol{y}, \hat{\boldsymbol{\beta}}^{g*}). \tag{2.135}$$

Using the term $D(\boldsymbol{y}, \hat{\boldsymbol{\beta}}^{g*})$ derives from Pregibon's (Pregibon, 1981) test for checking whether a modification to the hypothesized link function is necessary. Both Sakate & Kashid (2012) and Pregibon (1981) utilize deviance for GLM parameterization identification.

Chapter 3

A Random Effects Quasi Distribution

A random effects quasi distribution with mean μ_{Ri} and variance μ_{Ri}^{ψ} is described that intends to reduce the conditional response overdispersion of a hierarchical generalized linear model, and the standard error of the random effects estimates. The description includes a proposition that addresses the research questions given in Chapter 1.

3.1 Introduction

Chapters 1 and 2 introduced and expanded on the need to specify either the distribution of the random effects (REs), or provide the first and second moments of the random effects in hierarchical generalized linear models (HGLMs). Incorrect distribution or moments specification can lead to overdispersion in both the mean models and the dispersion models of the conditional response (CR) and the random effects (McCullagh & Nelder, 1989) and (Sakate & Kashid, 2012), which are independent of each other. The overdispersion is due to inadequate modeling of data clustering by the random effects. The mean-variance relationships with the limited development in the generalized linear models (GLMs) literature are the power function relationships, and biased model parameter estimates are highly likely in the presence of overdispersion. The literature (Yanez, 1993) and (Yanez & Wilson, 1995) treats power function mean-variance relationships for the dispersion model of the conditional responses GLMs. The random effects mean model mean-variance relationship of interest is:

$$\zeta V_R(\mu_{Ri}) = \zeta \mu_{Ri}^{\psi}, \tag{3.1}$$

where $V_R(\mu_{Ri})$ is the random effects variance function that is dependent on the mean μ_{Ri} of the random effects, ζ is the dispersion parameter for the random effects dispersion model, and ψ is the power function exponent whose value and estimation method is of interest.

The purpose of this chapter is to describe a method to recover information from the conditional and random effects model deviance residuals on the value of the exponent ψ in Equation 3.1, especially when this deviance is overdispersed. The assumption is that correct specification of the exponent ψ will not result in overdispersed random effects model deviance, thus minimizing the possibility of biased random effects mean and dispersion model parameter estimates, as well as achieving deviance residuals that follow a normal distribution. Recall from that deviance residuals following a normal distribution is an indication of an adequate fit of a model to data.

The first step in estimating ψ is to find the expected value of the random effects mean model deviance, which is dependent on ψ if ψ is the exponent in the random effects mean-variance relationship. Then this expected value is equated to the mean of a deviance distribution approximated by a power normal distribution. The equivalence allows ψ to be found. It will be shown that data are required to find an empirical estimate of ψ.

Four parts are needed to extract values for the power function exponent ψ of the random effects mean-variance relationship in Equation 3.1. These components are: (1) the expected value of the random effects deviance; (2) a suitable approximating distribution for the random effects deviance; (3) a relationship between the outcomes of Parts (1) and (2); and (4), initial values for iterative solutions to the HGLM parameters.

The expected value expression for the random effects deviance of Part (1) will be obtained in the Deviance Properties section. the Deviance Properties section also explores the properties of approximating distributions for the random effects deviance of Part (2). Part (3) establishes an equivalence relationship for the outcomes in Parts (1) and (2), and is given in the Derivation of the Random Effects Mean-Variance Power Function Exponent ψ section. Section HGLM IWLS discusses the generation of iteratively weighted least squares (IWLS) method initial values for solutions to the HGLM parameters. The following section presents a formal proposal of solutions to the random effects power function mean-variance relationship exponent ψ.

3.2 Proposition

Consider the GLM conditional response and random effects mean models, and conditional response and random effects dispersion models from Chapter 1, Figure 1.4, for individual observations reproduced here in Figure 3.1. The $\boldsymbol{\eta}$, $\boldsymbol{\eta}_d$, $\boldsymbol{\eta}_R$, and $\boldsymbol{\eta}_{dR}$ are the model response means resulting from application of the respective link functions, $g(\cdot)$, $g_d(\cdot)$, $g_R(\cdot)$, and $g_{dR}(\cdot)$; $\boldsymbol{x}_i^T$ and $\boldsymbol{z}_i^T$ are the i^{th} row vectors of the design matrices for the conditional response mean model and the random effects mean model, respectively; γ_0 and δ_0 are the conditional response and random effects systematic component parameters; and d_i and d_{Ri} are the deviance values for the conditional response mean model and the random effects mean model, respectively. The concern is when the random effects mean model has a power function mean-variance relationship as in Equation 3.1, namely, $\zeta V_R(\mu_{Ri}) = \zeta \mu_{Ri}^{\psi}$. The value of ψ must be estimated to give valid HGLM parameter estimates.

Effect	Mean Model Components	Dispersion Model Components
Fixed	$Y_{ij} \mid u_i \sim (\mu_{ij}, \phi_{ij} V(\mu_{ij}))$	$d_{ij} \sim (\phi_{ij}, 2\phi_{ij}^2)$
	$\eta_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{v}$	$\eta_{dij} = \gamma_0$
	$\eta_{ij} = g(\mu_{ij})$	$\eta_{dij} = g_d(\phi_{ij})$
Random	$u_i \sim (\mu_{Ri}, \zeta_i \mu_{Ri}^{\psi})$	$d_{R_i} \sim (\zeta_i, 2\zeta_i^2)$
		$\eta_{dRi} = \delta_0$
	$v_i = g_R(u_i)$	$\eta_{dRi} = g_{dR}(\zeta_i)$

Figure 3.1 HGLM with a power function mean-variance random effects model is comprised of a conditional response mean model, a conditional response dispersion model, a power function random effects mean model, and a random effects dispersion model.

Proposition 3.2.1 *The overdispersion from the HGLM conditional response mean model (Equation 3.2) as represented by the conditional response dispersion model Equation 3.3, and the overdispersion in the GLM random effects mean model (Equation 3.4) as represented by the random effects dispersion model Equation 3.5, both contain information that allow for the estimation of the random effects variance power function exponent ψ, from Equation 3.1.*

Recall that the precedent for using deviance as data was established when the extended quasi-likelihood allowed a HGLM estimates of the conditional response mean model dispersion parameter ϕ (Nelder & Pregibon, 1987). Also, Sakate &

Kashid (2012) used deviance for goodness-of-fit and link function identification, and the distribution of d_{dRi} was parameterized as $(\zeta, 2\zeta^2)$. When these elements are combined with the definition of the random effects deviance d_{Ri}, an estimate for ψ may be found.

To support the proposition, the residuals for HGLMs with random effects mean model mean-variance relationships as in Figure 3.1 are examined. The focus is on those characteristics necessary to explore solutions for the random effects mean-variance relationship exponent, ψ. The expected values of both the conditional response deviance d_i and the random effects deviance d_{Ri} are used. These expectations are then related to the random effects deviance distributions. The Studentized (standardized) signed square root form of the deviance residual, Equation 2.16, is commonly used as it is more often approximately normal than the Pearson residuals. The standardized signed square root form of the deviance residual needs to be approximately normal when the model fit is adequate. For overdispersed, right-skewed distributions, the normal distribution must be substituted with a distribution that approximates the skewness, and at the same time degenerates to a normal distribution as the overdispersion is reduced by accurately approximating the mean-variance power function exponent ψ. The power normal distribution is shown to satisfy these requirements, and is a suitable distribution approximation for the deviance. The random effects deviance parameters as estimated by the power normal distribution are related to the expected value of the random effects deviance d_{Ri}. This relationship is shown to provide the estimates of ψ.

3.3 Deviance Properties

In chapters 1 and 2, viable GLM models for data in which as many as four submodels, a conditional response mean model, a conditional response dispersion model, a random effects mean model, and a random effects dispersion model, were presented. In many cases these models are used to fit data that are overdispersed. While extended quasi-generalized linear models (EQGLMs) with integer-valued exponents (Nelder & Pregibon, 1987) and generalized extended quasi-generalized linear models (GEQGLMs) with relatively unrestricted exponent values (Yanez, 1993) were discussed for conditional response models with mean-variance relationships as described by Equation 2.68. What was lacking was a treatment of data with a random effects mean model for which the mean-variance relationship is best represented as in Equation 3.1. As such, this section examines HGLMs, specifically, double extended quasi-likelihood (DEQL) HGLMs, for which Equation 3.1 applies.

As shown in Chapter 2, Equation 2.97, the deviance d_{Ri} for the random effects mean model is

$$
\begin{aligned}
d_{Ri} &= 2\int_{u_i}^{\xi_i} \frac{\xi_i - s}{s^{\psi}} ds \qquad (3.2)\\
&= 2\left[\frac{\xi_i}{1-\psi}\left(\xi^{1-\psi} - u^{1-\psi}\right) - \frac{1}{2-\psi}\left(\xi^{2-\psi} - u^{2-\psi}\right)\right], \quad \psi \neq 1, 2.
\end{aligned}
$$

The expected value of d_{Ri} is

$$
\mathcal{E}(d_{Ri}) = 2\left\{\frac{1}{1-\psi}\left[\mathcal{E}\left(\xi_i^{2-\psi}\right) - \mu_{Ri}^{1-\psi}\mathcal{E}\left(\xi_i^{1-\psi}\right)\right] + \frac{1}{2-\psi}\left[\mathcal{E}\left(\xi_i^{2-\psi}\right) - \mu_{Ri}^{2-\psi}\right]\right\} \qquad (3.3)
$$

$$
= 2\left[\frac{\mathcal{E}\left(\xi_i^{2-\psi}\right) - (2-\psi)\mu_{Ri}^{1-\psi}\mathcal{E}(\xi_i) + (1-\psi)\mu_{Ri}^{2-\psi}}{(1-\psi)(2-\psi)}\right]. \qquad (3.4)
$$

Not surprisingly, $\mathcal{E}(d_{Ri}) \propto f(\psi)$, and yet the expectation function $f(\psi)$ does not seem to lend itself to a closed form solution for ψ. It should be noted, that, except for ψ, all other terms may be observed or estimated from data: for example, the pseudo-responses ξ_i are extracted from the IWLS augmented response vector; while μ_{Ri} are obtained by averaging the d_{Ri}.

To provide estimates of ψ, methods such as maximum likelihood estimation (MLE) or method of moments estimation (MME) are required. Before either of these two methods can be employed, the distribution properties of the random effects deviance must be explored.

Nelder & Pregibon (1987) suggest that, for the conditional response mean model deviance,

$$
d_i = \int_{\mu_i}^{y_i} \frac{y_i - s}{s} ds, \qquad (3.5)
$$

with $\mathcal{E}(d_i) = \phi$ and $Var(d_i) = 2\phi_i^2$, so that

$$
d_i \sim (\phi, 2\phi^2) \dot{\sim} \, gamma(\phi, 2\phi^2), \qquad (3.6)
$$

where $\hat{d}_i = (y_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}})^2$ and $\hat{\phi}$ are found iteratively via IWLS.

For Equation 3.1, due to the power function exponent ψ, the expected value of the random effects deviance residuals is $\mathcal{E}(d_{Ri}) = \zeta$, and $Var(d_{Ri}) = 2\zeta^2$. Thus,

$$
d_{Ri} = \int_{\mu_{Ri}}^{\xi_i} \frac{y_i - s}{s^{\psi}} ds, \qquad (3.7)
$$

so

$$d_{Ri} \dot{\sim} (\zeta, 2\zeta^2). \tag{3.8}$$

As stated in Chapter 2, an attribute indicating an adequate fit of data to a GLM is that the Studentized deviance residuals follow a normal distribution. If this attribute holds, then

$$gamma(\zeta, 2\zeta^2) \dot{\sim} \mathcal{N}(f(\zeta), 2f(\zeta)^2). \tag{3.9}$$

This near equivalence of the gamma and normal distributions is not common. As the gamma distribution is a right-skewed distribution, it is desired to find a right-skewed distribution that degenerates to a normal distribution as the skewness decreases.

The statistical analysis of many research data often reveal skewed distributions. Transformations like the Box-Cox transformation (Box & Cox, 1964), are applied to the data to render them more tractable to the ordinary normal theory analysis. The power normal (PN) distribution (Freeman & Modarres, 2006a), (Freeman & Modarres, 2006b), and (Maruo et al., 2011) is one that specifies observation parameterization prior to transformation. This power normal distribution parameterization prior to transformation is just what is needed to characterize the random effects deviance, and is described below.

As stated in Chapter 2, the power normal distribution parameters are usually estimated under the assumption that the transformed distribution is normal. This typically is not a problem when the intent is to achieve an approximately normal distribution from the transformed data, such as to find a power function exponent ψ from an overdispersed random effects model with a mean-variance relationship as in Equation 3.1. Maruo et al. (2011) state difficulties arise when identifying the power normal distribution on the original scale of the observations, in this case, the random effects deviance d_{Ri}. The estimators of the power normal distribution percentiles, for example, based on the likelihood estimation method, are not consistent estimators. Therefore, it is necessary to find unbiased and consistent parameter estimates and functions of the the parameter estimates when the parameters are determined under the assumption that the transformed distribution is a normal distribution.

As the MLEs of the parameters of the PN distribution are not asymptotically consistent (Maruo et al., 2011), (Freeman & Modarres, 2006a) and (Freeman & Modarres, 2006b) provide useful forms for the moments of a PN distribution for which the series approximation of the first moment, μ, is

$$\mu = \mathcal{E}X^1 = \begin{cases} (\lambda Y + 1)^{\frac{1}{\lambda}}(1 - \lambda^2) & , \lambda \neq 0, \\ \exp(\mu + \sigma^2/2) & , \lambda = 0, \end{cases} \tag{3.10}$$

where X is a random variable and Y is defined as

$$Y = \begin{cases} \frac{X^{\lambda}-1}{\lambda} & , \lambda \neq 0, \\ \ln(X) & , \lambda = 0, \end{cases} \tag{3.11}$$

and λ is the transformation parameter (Box & Cox, 1964).

The series approximation for the first moment of the PN distribution now is applied to the HGLM random effects deviance, and the corresponding properties are examined.

3.4 Derivation of the Power Function Exponent ψ

Pierce & Schafer (1986) state that GLM deviance residuals are used to identify poorly fitting observations, to plot to examine for the effects of potential new covariates or nonlinear effects of those already used to fit the GLM, to combine them into overall goodness-of-fit tests, and as components for case-influence diagnostics. The standardized signed square root form of the deviance residual, Equation 2.16, is commonly used as it is more often approximately normal than the Pearson residuals (Chapter 2). It was stated above that the deviance residuals, d_{Ri}, follow an approximate gamma distribution, and that it needs to degenerate to a normal distribution when the model fit is adequate. Also, it was indicated that for right-skewed gamma distributions, the normal distribution must be substituted with a distribution that approximates the skewness, and must degenerate to a normal distribution as the overdispersion is reduced by accurately approximating the power function mean-variance relationship quasi distribution exponent ψ.

If a gamma distribution is to converge to a normal, by the Central Limit Theorem, the number of observations must be large. The number of observations used to construct a HGLM may not be large enough to allow the normal distribution to approximate the gamma distribution. The PN is immune to small sample size, skewed model residuals, and degenerates to a normal distribution. In addition, HGLM Studentized residuals may exhibit truncation, particularly in the lower tail of the distribution, and the PN provides a suitable model.

The power normal distribution was shown to satisfy the skewness, truncation, and degeneracy requirements. Thus,

$$d_{Ri} \sim gamma(\zeta, 2\zeta^2) \mathrel{\dot\sim} \mathcal{PN}(\lambda, 2\zeta^3, 2\zeta^5). \tag{3.12}$$

It remains to equate the expected value of the random effects deviance residuals to the first moment of the power normal distribution. This equation will permit estimation of the power function exponent ψ.

First, the random effects deviance residuals parameter notation needs to be used in the first moment in Equation 3.10 such that:

$$\mathcal{E}(d_{Ri}) = 2\zeta^3 = (\lambda d_{\lambda i} + 1)^{\frac{1}{\lambda}}(1 - \lambda^2), \tag{3.13}$$

where $d_{\lambda i} = (d_{Ri}^{\lambda} - 1)/\lambda,\ \lambda \neq 0$. Here, $d_{\lambda i}$ is substituted for Y and d_{Ri} is substituted for X in the power normal distribution development in Chapter 2.

Next, the expected value of the random effects deviance residuals is equated to the first moment expression. Therefore,

$$2\left[\frac{\mathcal{E}(\xi^{2-\psi}) - (2-\psi)\mu_{Ri}^{1-\psi}\mathcal{E}(\xi_i) + (1-\psi)\mu_{Ri}^{2-\psi}}{(1-\psi)(2-\psi)}\right] = (\lambda d_{\lambda i} + 1)^{\frac{1}{\lambda}}(1 - \lambda^2). \tag{3.14}$$

With the exception of the parameter ψ, all variables and parameters are either observed or estimated. For example, $\mathcal{E}(\xi_i)$ may be estimated from the IWLS pseudo-response values by averaging, i.e., $\bar{\xi}$, as can μ_{Ri}. As such, iterative methods can be employed to estimate ψ. Bootstrap confidence intervals then can be obtained. Estimating ψ and the corresponding confidence intervals are the subject of the next chapter.

To obtain an estimate of ψ, it is clear from Equation 3.14 that an estimate for λ is required. The practical method to estimate λ is to use the Box-Cox transformation (BCT) on the first iteration of the random effects model deviance residuals d_{Ri} such that $d_{\lambda i} = d_{Ri}^{\lambda}$ as used in Equations 3.13 and 3.14. An additional feature of the BCT is the generation of a confidence interval (CI) about $\hat{\lambda}$, which is commonly found in statistics packages such as the R statistical package (R Core Team, 2012). This CI may be used to simplify ψ in cases for,

$$0 \oplus \frac{1}{2} \oplus 1 \oplus 2 \in [CI_{low}, CI_{high}] \tag{3.15}$$

where $\oplus$ is the symbol for exclusive or. Then,

$$\psi \sim \begin{cases} \mathcal{N}(\cdot,\cdot) & ,\psi = 0 \\ Beta(\cdot,\cdot) & ,\psi = \frac{1}{2} \\ gamma(\cdot,\cdot) & ,\psi = 1, 2. \end{cases} \tag{3.16}$$

These cases are in common use, and shall be discussed no further than has already appeared in Chapter 2.

3.5 IWLS for Hierarchical Generalized Least Squares

The implementation of the DEQL IWLS in Chapter 2 iteratively solves for the parameters of the four submodels of the HGLM. A conditional response mean model provides estimates of the conditional response parameters β, and provides weights to the conditional response dispersion model. The conditional response mean model is dependent on the random effects mean model parameter estimates and hence, solutions to the random effects mean model affect the conditional response mean and dispersion model parameter estimates. Current implementations of the random effects mean model parameter estimates are determined independently of the conditional response mean and dispersion models. The random effects mean model supplies weights to the random effects dispersion model for estimating the dispersion model parameters.

Therefore, IWLS is used to approximate MLEs for the parameters of all four submodels, which thereby give solutions for the HGLM overall. The IWLS requires initial values to begin the iteration process. The initial values germane to solutions of the random effects mean and dispersion models are problematic for random effects with a power function mean-variance relationship. Current implementations such as the R *hglm* package (Ronnegard, Shen & Alam, 2010) assign a value of either 0 or 1 to the pseudo-response ξ_i, and these values are specific to the normal distribution and the gamma distribution, respectively. An initial value of $\frac{1}{2}$ is used for random effects that follow a beta distribution, but this generally is applicable when the conditional response model response follows the conjugate binomial distribution, which is not treated here. SAS/STAT® software (SAS, 2011) has an option in PROC GLIMMIX for a conditional response GLM to be fitted initially to obtain starting values for the fixed-effects parameters. Given the fixed-effects estimates, starting values for the covariance parameters are computed by a MIVQUE0 step (Goodnight, 1978). For METHOD=QUAD, pseudo-likelihood updates are generated to improve on the estimates and to obtain solutions for the random effects models. Chapter 4 develops of a method for choosing the initial values for the random effects pseudo-response, ξ_i, that utilize a similar self-starting capability for random effects that follow a power function mean-variance relationship, .

Briefly, to find viable initial values for the random effects pseudo-response ξ_i, recall that the conditional response model deviance, d_{ij}, is a function of the random effects predictors $\boldsymbol{Z}_i^T v_i$. Therefore, d_{ij} contains information pertaining to ξ_i, especially when $\phi > 1$, i.e., when the d_{ij} are overdispersed. Thereby, $\mathcal{E}(d_{ij})$ is connected to the variance multipliers in the mean models, whose variance involves ψ. Then the initial values of ξ_i can be derived from d_{ij}, namely, $\mathcal{E}(d_{ij})$ become the seed values for the ξ_i in the IWLS. After iteration convergence of the IWLS, the random effect deviance, d_{Ri}, provides information on ψ such that

$$\psi \propto f(u_i, \zeta, d_{Ri}), \tag{3.17}$$

from which estimates for ψ may be derived as indicated by Equation 3.14. Hence, using both d_{ij} and d_{Ri}, an estimate of ψ may be found. The algorithms to find ψ follow in Chapter 4.

Chapter 4

Exponent ψ Estimation And Efficacy

This chapter describes the methods of estimating the random effects power function mean=variance quasi distribution exponent ψ. This description is followed by a comparison of the normal, gamma, and quasi distributions. The comparison demonstrates the efficacy of the quasi distribution when used on three selected data sets.

4.1 Introduction

The construction of generalized linear models (GLMs) can include a response conditional on random effects which group the responses into clusters of homogeneous variance. This conditional response may exhibit overdispersion that can result from misspecification of the explanatory variables or from misspecification of the random effects distributions. Additionally, overdispersion on the random effects, independent of the conditional response overdispersion, is possible. Thus, four submodels, a conditional response sub-model, a conditional response dispersion submodel, a random effects submodel, and a random effects dispersion submodel, may be assembled to form a joint generalized linear model adequate to characterize the conditional responses.

The conditional response of interest in this dissertation is counts data that are assumed to follow a Poisson distribution, though these data may be overdispersed. Overdispersed conditional responses may be modeled by a linearizing transformation on the parameters of the explanatory variables, as may the random effects which cluster the conditional response. This research focuses particularly on modeling the overdispersion of the conditional response.

The random effects model deviance residuals are assumed to be estimated by the first moment of a power normal distribution parameterized by a truncation parameter, λ, a mean parameter, μ_{Ri}, such that the random effects pseudo-response expected value is $\mathcal{E}(\xi_i) = \mu_{Ri}$. The d_{Ri} is the i^{th} random effects model deviance residual. The expected value of d_{Ri} was shown in Chapter 3 to be

$$\mathcal{E}(d_{Ri}) = 2\left[\frac{\mathcal{E}(\xi_i^{2-\psi}) - (2-\psi)\mu_{Ri}^{1-\psi}\mathcal{E}(\xi_i) + (1-\psi)\mu_{Ri}^{2-\psi}}{(1-\psi)(2-\psi)}\right], \quad \psi \neq 1, 2, \tag{4.1}$$

where ξ_i is the pseudo-response in the random effects model. The power normal first moment approximation of the mean parameter also was shown in Chapter 3 to be

$$\mu_{Ri} = (1-\lambda^2)d_{Ri}, \tag{4.2}$$

where λ is obtained from a Box-Cox transformation on the deviance residuals from the HGLM random effects with conditional response following the Poisson distribution, and the random effects following the gamma distribution, which is the conjugate distribution to the Poisson distribution.

Recall that ψ is the exponent in the power function mean-variance relationship such that

$$V(\mu_{Ri}) = \mu_{Ri}^{\psi}, \tag{4.3}$$

where $V(\mu_{Ri})$ is the variance of the quasi distribution with mean μ_{Ri}. It is the exponent ψ that will be estimated, and thereby allow testing of the efficacy of a quasi distribution, $(\mu_{Ri}, \mu_{Ri}^{\psi})$, to describe the random effects.

Chapter 1 presented research questions regarding the estimation and efficacy of use of the quasi distribution for the random effects. The questions are whether the exponent in the quasi distribution may be estimated, and will the quasi distribution assumption for the random effects reduce the overdispersion of the GLM more than when assuming either a normal or gamma distribution for the random effects? Further, are the quasi distribution estimators for the random effects improved over the normal and gamma distribution estimators? Answers to these questions will be obtained in this chapter.

Before the quasi distribution may be evaluated against the normal and gamma distribution assumptions for the random effects, the exponent ψ must be estimated.

This is the subject of the Section "Solutions of the Quasi Distribution Exponent ψ". Once ψ is estimated, specific data sets are required to make the comparisons among the normal, gamma, and quasi distributions. The Section "Data Sets" describes the data used in the assessment. Following the data sets descriptions is the Section "Comparative Assessment of the Quasi Distribution", which provides the analyses comparing the normal, gamma, and quasi distribution random effects estimates. The final section, "Summary", gives an overview of the assessment findings.

4.2 Solutions of the Quasi Distribution Exponent ψ

The first research question to answer is whether the power function mean-variance relationship quasi distribution exponent can be estimated. This is the topic of this section.

To find a solution for the quasi distribution exponent ψ, the equation relating the expected value of the random effects deviance residuals to the first moment approximation of the power normal distribution mean is rewritten as:

$$2\left\{\frac{\mathcal{E}(\xi_i^{2-\psi}) - (2-\psi)[\mathcal{E}(\xi_i)]^{1-\psi} + (1-\psi)[\mathcal{E}(\xi_i)]^{2-\psi}}{(1-\psi)(2-\psi)}\right\} = (1-\lambda^2)d_{Ri},\ \psi \neq 1, 2, \tag{4.4}$$

where ξ_i is the i^{th} random effects pseudo-response in the iterated weighted least squares equations, $\mathcal{E}(\xi_i) = \mu_{Ri}$ for μ_{Ri} is the i^{th} random effects mean, d_{Ri} is the i^{th} deviance residual from a Poisson-gamma HGLM, λ is the truncation parameter in the power normal distribution which is estimated from the Box-Cox transformation on the random effects deviance residuals, and ψ is the quasi distribution exponent requiring a solution.

Equation 4.4 has one problematic term that needs attention before a solution for ψ is possible. This term is $\mathcal{E}(\xi_i^{2-\psi})$, and appears intractable even though values for ξ_i are available. Two methods are used in this work to find a solution for ψ: a closed form approximation which uses a Taylor approximation for $\mathcal{E}(\xi_i^{2-\psi})$, and a nonlinear, empirical estimation of ψ, which also uses the Taylor approximation for $\mathcal{E}(\xi_i^{2-\psi})$.

Closed Form Estimation of ψ

A closed form solution for the quasi distribution exponent ψ may be obtained by applying the Taylor series to $\mathcal{E}(\xi_i^{2-\psi})$ in Equation 4.4. If a function of a random variable X, say $H(X)$, can be expanded in a Taylor series, then an expression for the approximate mean and variance of $H(X)$ can be obtained in terms of the mean and variance of X. See, for example, Bain & Engelhardt (1992).

Suppose $H(X)$ is infinitely differentiable on an open interval containing $\mathcal{E}(X)$. The function $H(X)$ has a Taylor approximation about the mean μ of X as

$$H(X) \doteq H(\mu) + H'(\mu)(x-\mu) + \frac{1}{2}H''(\mu)(x-\mu)^2, \tag{4.5}$$

where $H'(\mu)$ is the first derivative of $H(\mu)$, and $H''(\mu)$ is the second derivative of $H(\mu)$. This suggests the approximations

$$\mathcal{E}[H(X)] \doteq H(\hat{\mu}) + \frac{1}{2}H''(\hat{\mu}), \tag{4.6}$$

and

$$\widehat{Var[H(X)]} \doteq [H''(\hat{\mu})]^2. \tag{4.7}$$

Note that the expected value of $(x-\mu)=0$, so the terms with $H'(\mu)$ are zero.

So, for $\xi_i = X$, $\xi_i^{2-\psi} = H(X)$, $\mathcal{E}(\xi) = \bar{\xi}$ which is the average of the pseudo-response values ξ_i, and using the Taylor approximation about $\hat{\mu}$,

$$\mathcal{E}(\xi_i^{2-\psi}) \doteq \bar{\xi}^{2-\psi} + \frac{1}{2}(\bar{\xi}^{2-\psi})''Var(\xi_i). \tag{4.8}$$

The second derivative of $\xi_i^{2-\psi}$ is

$$\begin{aligned}(\bar{\xi}_i^{2-\psi})'' &= [(\bar{\xi}^{2-\psi})']' \\ &= [(2-\psi)\bar{\xi}^{1-\psi}]' \\ &= (1-\psi)(2-\psi)\bar{\xi}^{-\psi}.\end{aligned} \tag{4.9}$$

Then

$$\mathcal{E}(\xi_i^{2-\psi}) = \bar{\xi}^{2-\psi} + \frac{1}{2}[(1-\psi)(2-\psi)\bar{\xi}^{-\psi}]Var(\xi_i). \tag{4.10}$$

With an expression for the problematic term, $\mathcal{E}(\xi_i^{2-\psi})$, Equation 4.4 may be simplified. Let $A = \mathcal{E}(\xi_i)$, $a = 1 - \psi$, and $b = 2 - \psi$, then

$$
\begin{aligned}
2\left\{\frac{\mathcal{E}(\xi_i^b) - b \cdot A^a + a \cdot A^b}{ab}\right\} &= 2\left\{\frac{\mathcal{E}(\xi_i^b) + (a-b)A^b}{ab}\right\} \\
&= 2\left\{\frac{\mathcal{E}(\xi_i^{2-\psi}) - [\mathcal{E}(\xi_i)]^{2-\psi}}{(1-\psi)(2-\psi)}\right\} \\
&= 2\left\{\frac{\mathcal{E}(\xi_i^{2-\psi}) - \bar{\xi}^{2-\psi}}{(1-\psi)(2-\psi)}\right\} \\
&\doteq 2\left\{\frac{[\bar{\xi}^{2-\psi}\frac{1}{2}(1-\psi)(2-\psi)\bar{\xi}^{-\psi}Var(\xi_i)] - \bar{\xi}^{2-\psi}}{(1-\psi)(2-\psi)}\right\} \\
&= 2\left\{\frac{\frac{1}{2}(1-\psi)(2-\psi)\bar{\xi}^{-\psi}Var(\xi_i)}{(1-\psi)(2-\psi)}\right\} \\
&= \bar{\xi}^{-\psi}Var(\xi_i).
\end{aligned}
\tag{4.11}
$$

Note that this closed form approximation no longer requires that $\psi \neq 1, 2$.

Using this approximation and defining $\bar{d}_R$ as the average of the random effects deviance residuals, Equation 4.4 may be rewritten as

$$\bar{\xi}^{-\psi}Var(\xi_i) = (1 - \lambda^2)\bar{d}_R, \tag{4.12}$$

where $\bar{d}_R$ may be substituted for d_{Ri} as the deviance residuals are considered independent.

Therefore, the closed form solution for the quasi distribution exponent ψ is

$$\hat{\psi} = \frac{\ln[Var(\xi_i)] - \ln(1 - \hat{\lambda}^2) - \ln(\bar{d}_R)}{\ln(\bar{\xi})}, \tag{4.13}$$

for $\bar{\xi} > 0$, $\bar{d}_R > 0$, and $-1 < \hat{\lambda} < 1$. So, with new restrictions on the independent factors for $\hat{\psi}$, a closed form approximate solution for the power function mean-variance relationship quasi distribution exponent is obtained. In addition, a confidence interval for this closed form solution may be found from the supporting data, ξ_i, by using bootstrap resampling.

Empirical Estimation of ψ

The implementation of an empirical method to find a solution for the quasi distribution exponent ψ utilizes the Taylor approximation described in the previous subsection. Equation 4.4 is modified to

$$\xi_i^{-\psi} Var(\xi_i) = (1 - \hat{\lambda}^2) d_{Ri}, \tag{4.14}$$

where the means $\bar{\xi}$ and $\bar{d}_R$ are substituted with ξ_i and d_{Ri}, respectively, again as these data are assumed independent.

Equation 4.14 is algebraically manipulated to obtain a form suitable for nonlinear regression:

$$d_{Ri} = \frac{Var(\xi_i)}{1 - \hat{\lambda}^2} \xi_i^{-\psi}, \quad -1 < \hat{\lambda} < 1. \tag{4.15}$$

Note that in this form, the random effects deviance residuals, d_{Ri}, are treated as the response variables, and the pseudo-responses, ξ_i are treated as the predictors. Thus, the empirical nonlinear solution for ψ is algorithmically represented as

$$d_{Ri} = a\xi_i^{-\psi}, \tag{4.16}$$

where $a = Var(\xi_i)/(1 - \hat{\lambda}^2)$. The commonly used algorithms in current statistical packages provide a confidence interval for the nonlinear regression estimate of ψ.

Algorithms for Estimating ψ

Iterated weighted least squares (IWLS), as described in Chapter 2, is used to find solutions to GLMs with random effects, known as hierarchical generalized linear models. IWLS solutions are needed to find estimates of the quasi distribution exponent ψ, namely, the pseudo-response ξ_i and the random effects deviance residuals d_{Ri} used in the closed form and empirical methods above. However, the IWLS implementations of the various available statistical packages do not allow the specification of the quasi distribution for the random effects. As such, either an original IWLS implementation must be created, or an existing implementation must be modified.

The package *hglm* (Alam, Ronnegard & Shen, 2010), available from the Comprehensive R Archive Network (R Core Team, 2012), implements the IWLS for hierarchical generalized linear models described by Lee & Nelder (1996). The IWLS fits GLMs with random effects where the random effects must follow a normal, gamma, beta, or inverse-gamma distribution. The *hglm* package produces estimates of any fixed effects, random effects, and variance components, as well as their standard errors. The *hglm* output includes model diagnostics such as deviance components leverages and diagnostic plots.

The fitting algorithm (Lee et al., 2006) is summarized as follows:

1. Initialize starting values.
2. Construct an augmented response vector such that

$$\boldsymbol{y}_a = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{\xi} \end{pmatrix}, \tag{4.17}$$

   where $\boldsymbol{y}_a$ is the augmented response vector, $\boldsymbol{y}$ is the conditional response vector, and $\boldsymbol{\xi}$ is the pseudo-response for the random effects as described and used in the earlier subsections.
3. Use a GLM to estimate initial values of the conditional response model parameters $\boldsymbol{\beta}$ and the random effects parameters $\boldsymbol{v}$, given the conditional response dispersion parameter $\boldsymbol{\phi}$, and the random effects dispersion parameter $\boldsymbol{\zeta}$. Save the deviance components and leverages from this initial fitted model.
4. Use a $d_i \sim gamma(\cdot, \cdot)$ GLM to estimate $\boldsymbol{\beta}_d$ from the conditional response deviance residuals and their associated leverages $\boldsymbol{q}_i$ (IWLS leverages from Chapter 2), where these deviance components are from step 3). The gamma GLM is a random intercept model. Update the dispersion parameter by setting $\boldsymbol{\phi}$ equal to the predicted response vector from this step 4 model.
5. Use a similar GLM to step 4 to estimate $\boldsymbol{\zeta}$ from the random effects deviance residuals again obtained from the step 3 GLM.
6. Iterate steps 3 to 5 until convergence.

While the necessary components for the IWLS are present in the *hglm* package, two important elements are not made available or not implemented, so solutions for ψ may not be found without modifying the package. The first element that is not made available is the pseudo-response ξ_i of the random effects from the augmented response vector shown in step 2. The *hglm* scripts were modified to return ξ_i in the output as a variable named xi. This variable then is used to estimate both the closed form and empirical estimates of ψ.

The second element that needs implementation in the *hglm* scripts is the recognition and execution of quasi distributions for the random effects. The R package has provision to construct a quasi distribution according to a user's specifications: in this case, a quasi distribution with a power function variance, $V(\mu) = \mu^{\psi}$ and a log link function. With the random effect distribution family defined as a quasi distribution, the *hglm* scripts were modified to recognize and initialize the power function mean-variance relationship quasi distribution family. Execution of the *hglm* scripts then proceeds normally.

Testing of the modified *hglm*, named *ghglm*, was on several existing data sets, used by statistics packages as examples procedure or function use, to assure the *ghglm* output matched exactly the output of *hglm*. The tests and final execution of *ghglm* using the random effects quasi family definition used $\hat{\lambda}$ as the initial value of ψ. Convergence was realized in the same number of iterations as those of the usual normal and gamma distributions when these distributions were used for the random effects.

The efficacy of the closed form and empirical solutions for ψ, as implemented in *ghglm*, is examined by subjecting them to three data sets. What follows is a description of these three data sets.

4.3 Data Sets

Chapter 1 introduced three data sets that were chosen to assess the efficacy of using a power function mean-variance relationship quasi distribution for random effects in hierarchical generalized linear models. These data sets were subjected to the *ghglm* scripts that were modified from the original Alam et al. (2010) *hglm* package to handle quasi distributions.

Two of the data sets have been used in the literature for model comparisons between competing analyses. They are the fabric data of Bissell (1972), and the rats data of Myers & Montgomery (2002). Both these data sets have random effects that require estimation.

The third data set is sunspot counts data supplied by the American Association of Variable Star Observers Solar Section, and is the motivating data set for this quasi distribution research.

In this dissertation, the data sets are analyzed in three hierarchical GLMs that assume the random effects first are normally distributed, then gamma-distributed, and then distributed as a power function quasi distribution. The outcomes are reported in the next section, while descriptions of the three data sets follow.

Fabric Data

Bissell (1972) presents data for which the number of faults, Y_{ij}, in a bolt of fabric is dependent on the random effect of the distance between faults. The conditional response Y_{ij} is the number of faults in a bolt of fabric. There are no fixed effects other than a random intercept, β. The random effects design matrix is $\boldsymbol{z}_i$, which is a design matrix of distances, u_i between faults, transformed as $v_i = \ln(u_i)$. The random effect u_i was modeled by Lee & Nelder (2000) as a normal distribution, but is shown in Figure 4.1 as a power function quasi distribution.

Effect	Mean Model Components	Dispersion Model Components
Fixed	$Y_{ij} \mid u_i \sim (\mu_{ij}, \phi_{ij}\mu_{ij})$	$d_{ij} \sim (\phi_{ij}, 2\phi_{ij}^2)$
	$\eta_{ij} = \beta + \boldsymbol{z}_i^T \boldsymbol{v}$	$\eta_{dij} = \gamma_0$
	$\eta_{ij} = \ln(\mu_{ij})$	$\eta_{dij} = d)ij$
Random	$u_i \sim (\mu_{Ri}, \zeta_i \mu_{Ri}^{\psi})$	$d_{R_i} \sim (\zeta_i, 2\zeta_i^2)$
		$\eta_{dRi} = \delta_0$
	$v_i = \ln(u_i)$	$\eta_{dRi} = d_{Ri}$

Figure 4.1 Fabric data HGLM with a power function mean-variance random effects model is comprised of a conditional response mean model, a conditional response dispersion model, a power function random effects mean model, and a random effects dispersion model.

Rat Data

Three chemotherapy drugs were applied to thirty rats that had induced leukemia. White and red blood cell counts were collected as covariates and the conditional response, Y_{ij}, is the number of cancer cell colonies. The data were collected on each rat at four different times. Myers & Montgomery (2002) analyzed these data first under the assumption of no within-rat white cell count correlation, and then with correlated white cell count. This dissertation considers only the first, uncorrelated case.

Among the three covariates, drug, white blood cell count, and red blood cell count, drug is a between-rat covariate, while the white and red blood cell counts are within-rat covariates. Figure 4.2 shows that $\boldsymbol{x}_{ij}$ is the design matrix of drug treatment, white blood cell count, and red blood cell count. The matrix $\boldsymbol{z}_i$ is the random effect design matrix of a function of the white blood cell count. The random effects was modeled by Lee et al. (2006) as both a normally distributed effect and as a gamma-distributed effect where v_i is a polynomial function in white blood cell count with intercept γ_0. The figure depicts the random effects function as a power function quasi distribution.

Effect	Mean Model Components	Dispersion Model Components
Fixed	$Y_{ij} \mid u_i \sim (\mu_{ij}, \phi_{ij}\mu_{ij})$ $\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{v}$ $\eta_{ij} = \ln(\mu_{ij})$	$d_{ij} \sim (\phi_{ij}, 2\phi_{ij}^2)$ $\eta_{dij} = \gamma_0$ $\eta_{dij} = d)ij$
Random	$u_i \sim (\mu_{Ri}, \zeta_i\mu_{Ri}^{\psi})$ $v_i = \ln(\gamma_0 + W_i + W_i^2)$	$d_{R_i} \sim (\zeta_i, 2\zeta_i^2)$ $\eta_{dRi} = \delta_0$ $\eta_{dRi} = d_{Ri}$

Figure 4.2 Rat data HGLM with a power function mean-variance random effects model is comprised of a conditional response mean model, a conditional response dispersion model, a power function random effects mean model, and a random effects dispersion model.

Sunspot Data

The sunspot data analysis was the motivation for this research on power function quasi distributions. Obtaining an accurate sunspot number with observer as a random effect proved challenging, leading to modeling observer with a quasi distribution.

The American sunspot number was developed during World War II when United States scientists could no longer access the historically-used Zurich sunspot numbers (Shapley, 1949). The American sunspot number is a relative index of sunspot activity, and has been recorded by the American Association of Variable Star Observers (AAVSO) Solar Section since its inception in 1944. Shapley (1949), Taylor (1985), and Schaefer (1993) provide descriptions of the traditional method of sunspot counts data reduction. AAVSO observer-supplied daily sunspot counts arrive at the Solar Section as sets of date- and time-stamped values which are converted to a sunspot number known as the Wolf number after its developer R. Wolf (Shapley, 1949) such that:

$$w = 10g + s, \tag{4.18}$$

where w is the Wolf number, g is the count of sunspot groups, and s is the total count of sunspots on a given day. Taylor (1985) states that the sunspot grouping (g) scheme is the evolutionary classification system outlined by Waldmeier (1961). The traditional data reduction method involves weighted least-squares regression parameter adjustments to a logarithmically transformed w to account for variations in the observers' counts. The variations in observer counts is due to such factors as seeing condition and observer experience. These factors are handled implicitly in the regression. The log-transformed w often still is right-skewed, and analysts arbitrarily truncate this skewness in an attempt to emulate normally-distributed data. This leaves the sunspot number accuracy questionable.

The GLM approach, displayed in Figure 4.3, uses a conditional response model where Y_{ij} is the Wolf number w from Equation 4.18. The fixed effects design matrix, $\boldsymbol{x}_{ij}$, is the matrix of seeing condition, observer experience, year and month of the observation, and $\boldsymbol{z}_i$ is the random effect design matrix of observer identifiers. The random effects model is shown in the figure as a power function quasi distribution.

Effect	Mean Model Components	Dispersion Model Components
Fixed	$Y_{ij} \mid u_i \sim (\mu_{ij}, \phi_{ij}\mu_{ij})$ $\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{v}$ $\eta_{ij} = \ln(\mu_{ij})$	$d_{ij} \sim (\phi_{ij}, 2\phi_{ij}^2)$ $\eta_{dij} = \gamma_0$ $\eta_{dij} = d)ij$
Random	$u_i \sim (\mu_{Ri}, \zeta_i\mu_{Ri}^{\psi})$ $v_i = \ln(u_i)$	$d_{R_i} \sim (\zeta_i, 2\zeta_i^2)$ $\eta_{dRi} = \delta_0$ $\eta_{dRi} = d_{Ri}$

Figure 4.3 Sunspot data HGLM with a power function mean-variance random effects model is comprised of a conditional response mean model, a conditional response dispersion model, a power function random effects mean model, and a random effects dispersion model.

Data Sets Summary

Each of these three data sets, fabric, rats, and sunspots, has a conditional response consisting of counts, and a random effect which influences the GLM overdispersion both in the conditional response and in the random effect to varying degrees depending on the data set. The next section is an analysis of the changes in overdispersion, random effects standard errors, and model diagnostic plots when the random effects in each data set are analyzed as following a normal distribution, a gamma distribution, and a power function mean-variance relationship quasi distribution.

4.4 Comparative Analysis of the Quasi Distribution

The fabric, rats, and sunspots data sets, described above, each are modeled with their respective random effects assuming each of three distributions. These three distributions are the normal distribution, the gamma distribution, and the power function mean-variance relationship quasi distribution. These distributions will often be referred to as the three distributions going forward. For each data set, the power normal distribution truncation parameter, λ, and its 95% confidence interval is given. Then the power function quasi distribution exponent, ψ, and its 95% confidence interval is given.

The sequence for estimating λ and ψ is that first a baseline Poisson-normal GLM is run. Poisson-normal refers to the conditional response following a Poisson distribution, while the random effects follow a normal distribution. Similar terminology follows for the Poisson-gamma and Poisson-quasi GLMs. Once the Poisson-normal GLM baseline is established, a Poisson-gamma GLM is run. The deviance residuals from this model are used to estimate the power normal distribution truncation parameter λ using a Box-Cox transformation. The pseudo-response $\boldsymbol{\xi}$ from the Poisson-gamma HGLM is used with $\hat{\lambda}$ to estimate ψ using both the closed form method and the empirical method. A X-Y plot of $\hat{\psi}$ versus $\hat{\lambda}$ is given for each of the three data sets. This plot shows the nonlinear response of the closed form of $\hat{\psi}$ to $\hat{\lambda}$.

With the determination of ψ, a Poisson-quasi GLM is run. Then each of the three random effects GLM outcomes are compared. The comparison consists of analyzing the overdispersion using the variance-to-mean ratio, reduction in the estimated average random effects standard errors, and the respective normal Q-Q diagnostic plots. The variance-to-mean ratio is calculated from the three GLM fitted values. The variance and mean of the fitted values form the variance-to-mean ratio. The random effects standard errors are generated for each level of the random effect by the GLMs. These individual standard errors then are averaged to form a metric for comparison. Both the variance-to-mean ratios and the estimated average random effects standard errors comparisons are graphically represented with bar plots. The bar plots represent the three distributions efficacy measure by the height of the bars, and the estimation method of ψ is represented by cross-hatching in the bars: positive slope hatching represents the closed form method, and negative slope hatching represents the empirical method. Finally, the normal Quantile-Quantile (normal Q-Q) diagnostic plots are examined. Normal Q-Q plots are graphical methods for comparing the standardized deviance residuals probability distribution to the standardized normal probability distribution by plotting their respective quantiles against each other. First, the set of intervals for the quantiles are chosen. A point (x,y) on the plot corresponds to one of the quantiles of the second distribution (ordinate) plotted against the same quantile of the first distribution (abscissa). Thus the normal line is compared to a parametric curve with each parameter being the number of the interval for the quantile. The normal Q-Q plots show if the standardized deviance residuals of the random effects can be considered to be normally distributed, and the three distributions can be compared to see which standardized deviance residuals are normally distributed and which are not.

The analysis for the fabric data is given first, followed by the rats data analysis, and then the sunspots data outcomes are given. An overall comparison summary closes this chapter.

Fabric Data

The estimate of $\hat{\lambda}$ and its associated 95% confidence interval (CI) is given in Table 4.1, and the associated closed form and empirical estimates of $\hat{\psi}$ along their respective 95% CI are in Table 4.2.

Table 4.1 *Fabric Data $\hat{\lambda}$ and 95% Confidence Interval*

Estimate	95% CI
0.1818	(0.0335, 0.3007)

Table 4.2 *Fabric Data $\hat{\psi}$ and 95% Confidence Interval*

Solution	$\hat{\psi}$	95% CI
Closed form	0.9102	(0.5785, 1.1738)
Empirical	0.8286	(0.6397, 1.0086)

Note that the empirical solution 95% CI is contained by the 95% CI of the closed form solution. This indicates there is no statistical difference between the two values of $\hat{\psi}$. The closed form estimate of ψ is 8.9651% larger than the empirical estimate. Both CI's contain the value one, which implies that a one-parameter gamma (exponential) distribution is a viable model for the random effects deviance residuals. As with the rat data and the sunspot data, a normal model and a gamma model for the fabric data did converge, and produce two HGLMs. However, with the quasi distribution, the variance function, μ_{Ri}^{ψ} iteratively drives to zero. Recall from the description on IWLS in Chapter 2, the random effects weight, W_{Ri}, is calculated as:

$$W_{Ri} = \frac{\partial\mu_{Ri}/\partial\xi_i}{\mu_{Ri}^{\psi}}, \tag{4.19}$$

$$s.e.(v_i) = \sqrt{diag\{f(W_{Ri}^{-1})\}} \tag{4.20}$$

where μ_{Ri} is the mean of the random effects, ξ_i is the pseudo-response of the random effects in the IWLS, and the quasi distribution variance function is μ_{Ri}^{ψ}. On each IWLS iteration, the variance function collapses to toward zero, which, as the variance function in the IWLS weight function appears in the denominator of the random effects standard error, forces the random effects variance-covariance matrix elements and thereby the standard errors of the random effects estimates to plus infinity. Hence, convergence of the HGLM is not obtained, and no model parameters are calculated.

Figure 4.4 is the fabric data plot of the power function exponent $\hat{\psi}$ versus the power normal distribution truncation parameter $\hat{\lambda}$. The plot is a downward-opening parabola with the abscissa having domain $-1 < \hat{\lambda} < 1$, and ordinate $0 \leq \hat{\psi} \leq 1$, with the parabola maximum of $\hat{\psi} = 0.9236$. The range restricts $\hat{\psi}$ to being a normal distribution ($\hat{\psi} = 0$), a gamma (exponential) distribution ($\hat{\psi} = 1$), or a quasi distribution ($0 < \hat{\psi} < 1$).

The inability of the power function quasi distribution to model the random effects of the fabric suggest that further work is needed to find suitable power function exponents. For example, a profile method may provide useful estimates of ψ. The power normal truncation parameter as a predictor of ψ may still be viable if another estimation method such as a maximum likelihood estimate of the expected value of the random effects deviance residuals is used.

Rat Data

The estimate of $\hat{\lambda}$ and its associated 95% confidence interval (CI) is given in Table 4.3, and the associated closed form and empirical estimates of $\hat{\psi}$ and their respective 95% CI are in Table 4.4.

Table 4.3 *Rat data* $\hat{\lambda}$ *and 95% confidence interval*

Estimate	95% CI
0.1010	(-0.0288, 0.2718)

Table 4.4 *Rat data* $\hat{\psi}$ *and 95% confidence interval*

Solution	$\hat{\psi}$	95% CI
Closed form	0.9639	(0.8686, 1.0478)
Empirical	0.9701	(0.9001, 1.0396)
$\Delta\ \hat{\psi}$	0.6432%	

The 95% CI for the empirical estimate is contained by the 95% CI of the closed form estimate, which is evidence that there is no statistically significant difference between the closed form and empirical estimates of ψ. The closed form estimate of ψ is 0.6432% smaller than the empirical estimate.

A plot of the closed form estimate of $\hat{\psi}$ versus $\hat{\lambda}$ is shown in Figure 4.5. The abscissa is $-1 < \hat{\lambda} < 1$, and ordinate ranges $0 \leq \hat{\psi} \leq 1$, with the downward-opening curve maximum of $\hat{\psi} = 0.9655$. The range of values of $\hat{\psi}$ indicate the random effects may be normally-distributed ($\hat{\psi} = 0$) gamma- (exponentially-) distributed ($\hat{\psi} = 1$), or quasi-distributed ($0 < \hat{\psi} < 1$).

The quasi distributions, $u_i \sim (\mu_{Ri}, \mu_{Ri}^{0.9639})$ or $u_i \sim (\mu_{Ri}, \mu_{Ri}^{0.9701})$, are used to make the performance comparisons. The first comparison is among the variance-to-mean ratios (conditional response overdispersion measure) for the normal, gamma, and quasi distributions for the two estimates of ψ. Figure 4.6 is a bar plot of the variance-to-mean ratios of the three random effects distributions.

The difference between the mean-to-variance ratios of the closed form and the empirical estimates of ψ is 0.0127%, with the empirical estimate having the smaller value. Figure 4.6 gives these ratio values rounded to 4 significant figures, for which there is no difference. A 2.4297% difference exists between the quasi distribution with the closed form estimate of ψ and the normal distribution, with the normal distribution having the smaller value. A 2.3460% difference exists between the quasi distributions and the gamma distribution, again with the gamma distribution having the smaller value.

The second comparison, depicted in Figure 4.7, is a bar plot of the random effects estimated average standard errors for the three distributions. There is a difference of 0.1898% between the closed form and empirical estimates of ψ, with the empirical estimate being the smaller. The normal distribution is 32.009% smaller than the standard errors of the closed form quasi distribution. The normal distribution is 31.807% smaller than the standard error of the empirically-based quasi distribution. The gamma distribution is 32.127% smaller than the ratio of the closed form quasi distribution, and the gamma distribution is 31.925% smaller than the ratio of the empirical quasi distribution.

Table 4.5 gives the closed form and empirical random effects standard errors along with their respective 95% CIs. The CI of the empirically derived standard error is contained by the CI of the closed form, so there is no statistically significant difference between the closed form and the empirical quasi distribution random effects standard errors.

A best goodness-of-fit for fitted HGLMs occurs when standard deviance resid-

Table 4.5 *Rat Data Standard Errors and 95% Confidence Intervals*

Solution	s.e.(RE)	95% CI	
Closed form	0.1054	(0.0773,	0.1334)
Empirical	0.1052	(0.0771,	0.1333)
$\Delta s.e(RE)$	0.1898%		

uals follow a normal distribution. The normal Q-Q diagnostic plots in Figure 4.8 suggest that none of the distribution's random effects standardized deviance residuals are normally distributed. This is clarified and supported by the Shapiro-Wilk tests for normality. The Shapiro-Wilk test parameters are given in Table 4.6. The normal Q-Q plots and Shapiro-Wilk tests indicate neither the normal distribution, nor the gamma distribution, nor the two quasi distributions are normally distributed, suggesting the fits to these distributions may be improved. The quasi distributions have slightly better goodness-of-fit outcomes than do the normal and gamma distributions.

Table 4.6 *Rat Data Standardized Deviance Residuals Shapiro-Wilk Tests*

Distribution	W	p-value	Normal
Normal	0.7020	1.245e-06	No
Gamma	0.6972	1.058e-06	No
Quasi, Closed form	0.7532	7.842e-06	No
Quasi, Empirical	0.7528	7.749e-06	No

The variance-to-mean ratio measure suggests that no distribution is preferred. The random effects standard errors measure suggests that the normal or gamma distributions are preferred over the two quasi distributions unless the normal and gamma distributions underestimate the Type I error rate. The normality diagnostics suggest that no distribution is preferred even the thought the quasi distributions have slightly better fit outcomes. Therefore the choice comes to using that distribution which provides the best interpretation of the random effects.

Sunspot Data

The estimate of $\hat{\lambda}$ and its associated 95% confidence interval (CI) is given in Table 4.7. The CI includes zero which implies an appropriate distribution to model the random effects deviance residuals is a normal distribution. The associated closed form and empirical estimates of $\hat{\psi}$ and their respective 95% CI are in Table 4.8.

Table 4.7 *Sunspot Data* $\hat{\lambda}$ *and 95% Confidence Interval*

Estimate	95% CI
-0.0202	(-0.0826, 0.0017)

Table 4.8 *Sunspot Data* $\hat{\psi}$ *and 95% Confidence Interval*

Solution	$\hat{\psi}$	95% CI
Closed form	0.7734	(0.6077, 0.8882)
Empirical	0.9930	(0.9662, 1.0197)
$\Delta\ \hat{\psi}$	16.132%	

Unlike the fabric data and the rat data, the 95% CIs for the closed form and empirical estimates do not overlap. This suggests there is a statistically significant difference between the closed form and empirical estimates of ψ. The closed form estimate is 16.132% smaller than the empirical estimate. The CI for the empirical estimate of ψ includes the value 1 which indicates the appropriate distribution to model the random effects is a one-parameter gamma distribution.

The plot of the closed form estimate of ψ versus $\hat{\lambda}$ is given in Figure 4.9. The abscissa is $-1 < \hat{\lambda} < 1$, and ordinate is $0 \leq \hat{\psi} \leq 1$, with the maximum of the downward-opening curve of $\hat{\psi} = 0.7734$. This constraint on $\hat{\psi}$ does allow the random effects to be normal when $\hat{\psi} = 0$, a one-parameter gamma when $\hat{\psi} = 1$, or a quasi when $0 < \hat{\psi} < 1$.

The closed form and nonlinear estimates for ψ give the distribution of the random effects as $u_i \sim (\mu_{Ri}, \mu_{Ri}^{0.7734})$ or $u_i \sim (\mu_{Ri}, \mu_{Ri}^{0.9662})$. These distinct estimates affect of the performance comparisons. Beginning with the variance-to-mean ratios in Figure 4.10, the empirical estimate is 0.0192% smaller than that of the closed from estimate. No best random effects model, normal, gamma, or quasi, is indicated. However, the normal distribution is 0.2766% smaller than that of the closed form quasi distribution. The normal distribution is 0.2573% smaller than that of the nonlinear quasi distribution. The gamma distribution is 0.2798% smaller than that of the closed form quasi distribution. Similarly, the gamma distribution is 0.2605% smaller than that of the nonlinear quasi distribution. These small differences suggest no one distribution need be preferred.

Table 4.9 gives the closed form and empirical random effects standard errors along with their respective 95% CIs. The CI of the empirically derived standard error is contained by the CI of the closed form, so there is no statistically significant difference between the closed form and the empirical quasi distribution random effects standard errors.

Table 4.9 *Sunspot Data Standard Errors and 95% Confidence Intervals*

Solution	s.e.(RE)	95% CI
Closed form	0.2312	(0.2024, 0.2599)
Empirical	0.2227	(0.1944, 0.2515)
$\Delta s.e(RE)$	3.6765%	

The estimated average of the random effects standard errors performance comparison in Figure 4.11 show a quite different situation. Firstly, the statistically significant difference in the estimation methods of ψ result in the nonlinear quasi distribution for the random effects average standard errors is 3.6541% smaller than the closed form quasi distribution. The plot and the associated differences reveals that the two quasi distributions are less desirable with regards to the average standard errors. That is, the normal distribution is 48.264% smaller than that of the closed form quasi distribution, and 42.847% smaller than that of the nonlinear quasi distribution. The gamma distribution is 49.774% smaller than that of the closed form quasi distribution, and 44.301% smaller than that of the nonlinear quasi distribution. These differences suggest the normal and gamma distributions may be preferred over the quasi distributions, but a check of the Type I error rates is appropriate before discounting the quasi distribution estimates.

The normal Q-Q diagnostic plots in Figure 4.12 suggest that none of the distribution's random effects standardized deviance residuals are normally distributed. This is supported by the Shapiro-Wilk tests for normality in Table 4.10. The normal Q-Q plots and Shapiro-Wilk tests indicate none of the distributions is normal, suggesting the fits to these distributions may be improved. Note that the two quasi distributions do not have the possible outlier of the normal and gamma distribution random effects standardized deviance residuals, and that the goodness-of-fit outcomes are very slightly better than the normal and gamma distributions.

Table 4.10 *Sunspot Data Standardized Deviance Residuals Shapiro-Wilk Tests*

Distribution	W	p-value	Normal
Normal	0.5412	3.555e-17	No
Gamma	0.5570	6.879e-17	No
Quasi, closed form	0.5841	2.217e-16	No
Quasi, empirical	0.5820	2.021e-16	No

The three efficacy measures of the variance-to-mean ratios, the estimated average random effects standard errors, and the normal Q-Q diagnostic plots give no preference to the two quasi distributions. The empirical quasi distribution exponent confidence interval does include the value one, implying a one-parameter gamma distribution may be appropriate to model the random effects. The mean-to-variance ratio values and the average standard errors may lead to choosing the one-parameter gamma distribution as a random effects model for the sunspot data, however, the Type I error rates must be examined for liberal interpretation.

Comparison Summary

A summary of the efficacy measures comparisons for each data set now is presented. The overall efficacy of the power function mean-variance relationship quasi distribution will be given in the next chapter.

For each of the data sets, the Poisson-gamma HGLM produced deviance residuals from which the power normal truncation parameter λ and the quasi distribution exponent ψ were estimated. The truncation parameter domain was restricted as the result of using the first moment approximation of the power normal distribution, which contains the constraining term $1 - \hat{\lambda}^2$. However, while $-1 < \hat{\lambda} < 1$, none of the thee data sets Box-Cox estimates of λ from the random effects deviance residuals were outside this domain. A range of quasi distribution exponents results as $0 \leq \hat{\psi} \leq 1$. If $\hat{\psi} = 0$, then the random effects follow a normal distribution. If $\hat{\psi} = 1$, then the random effects follow a one-parameter gamma (exponential) distribution. For $0 < \hat{\psi} < 1$, then the random effects follow a power function mean-variance relationship quasi distribution.

The fabric data analysis shows no statistically significant difference between ψ estimated from the closed form versus ψ estimated empirically. Non-convergence of the HGLM IWLS due to the weights approaching zero and the lack of fixed effects in the model, didn't produce the necessary performance measures by which to compare the quasi distributions to the normal and gamma distributions. When values of $\psi < 0.5$ that were not selected from the deviance residuals, convergence was achieved. Hence another method for selecting ψ for at least these fabric data may be useful.

The rat data also show no statistically significant difference between the closed form and empirical estimation methods in the distribution comparisons. The efficacy measures all show that neither the normal, gamma, or quasi distribution are preferred to model the random effects, though the larger average standard errors of the two quasi distributions suggest either the normal or the gamma distributions Type I errors should be evaluated for accurate representation. The confidence intervals of the two ψ estimation methods include the value one, which implies a one-parameter gamma should be considered. The best choice then becomes the one that provides the most reasonable interpretation of the random effects.

The sunspot data set is the only set that shows a statistically significant difference between the estimates of ψ from the closed form method and the empirical method. However, the efficacy measures all show that neither the normal, gamma, nor quasi distributions are preferred to model the random effects. The larger average standard errors of the two quasi distributions suggest either the normal or the gamma distributions may be considered to model the random effects, Type I error rates allowing.

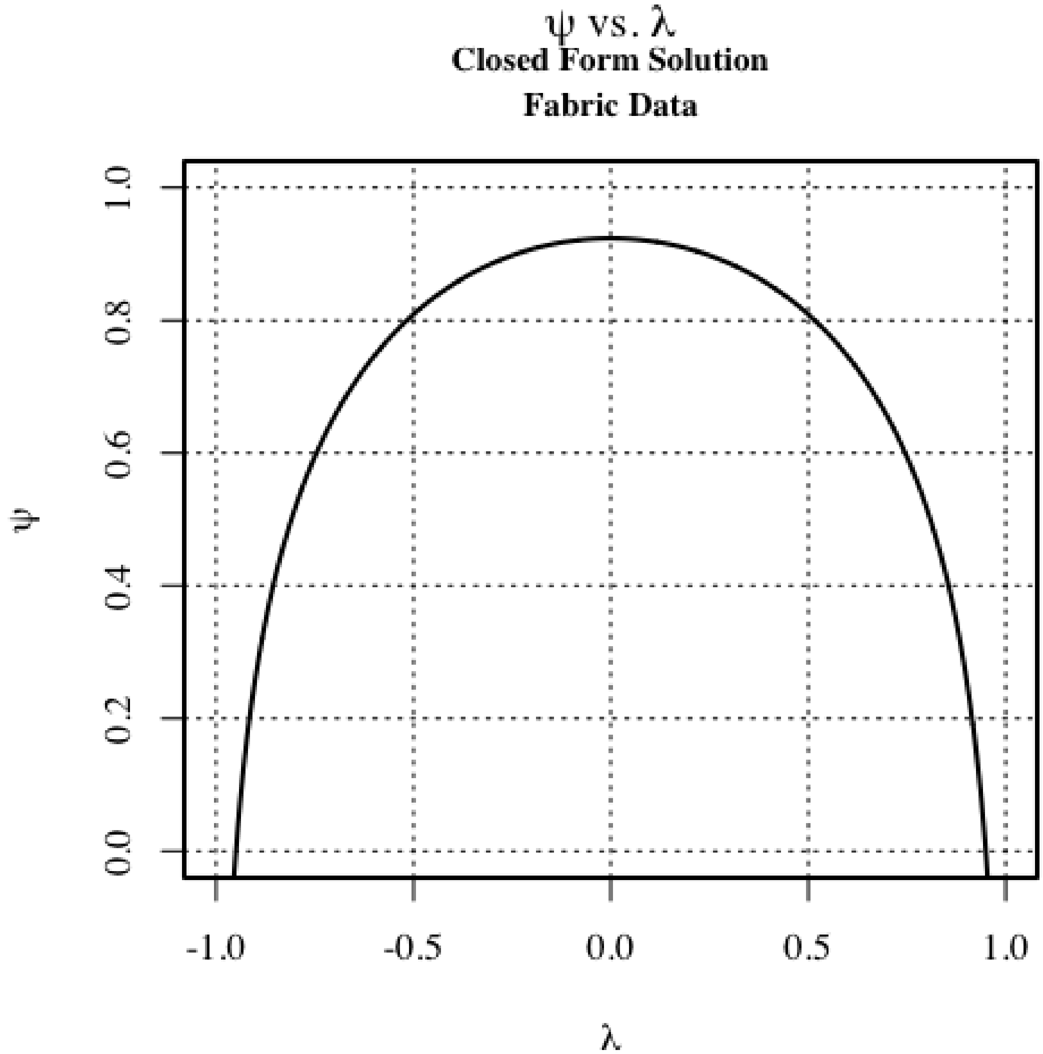

Figure 4.4 Power function exponent ψ versus the power normal truncation parameter λ for the fabric data.

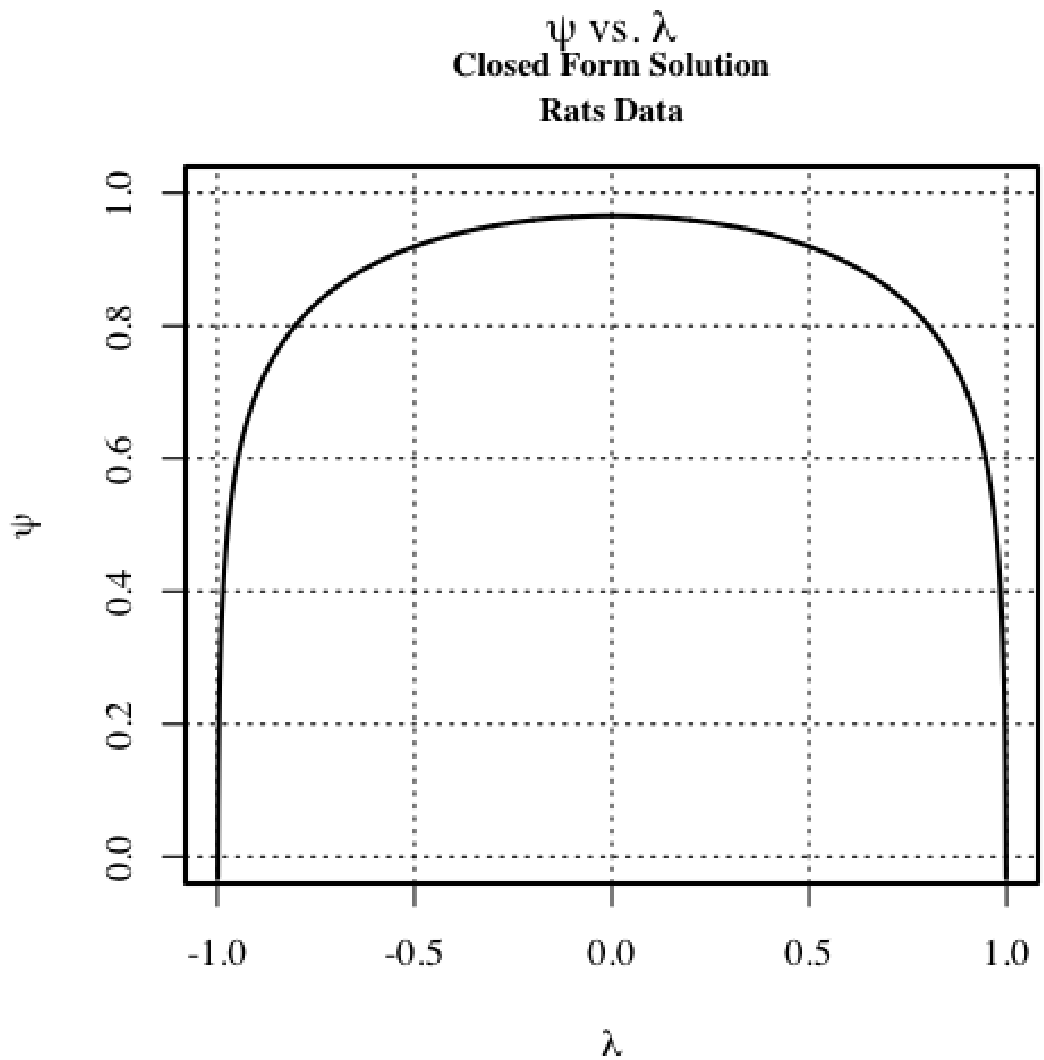

Figure 4.5 Power function exponent ψ versus the power normal truncation parameter λ for the rat data.

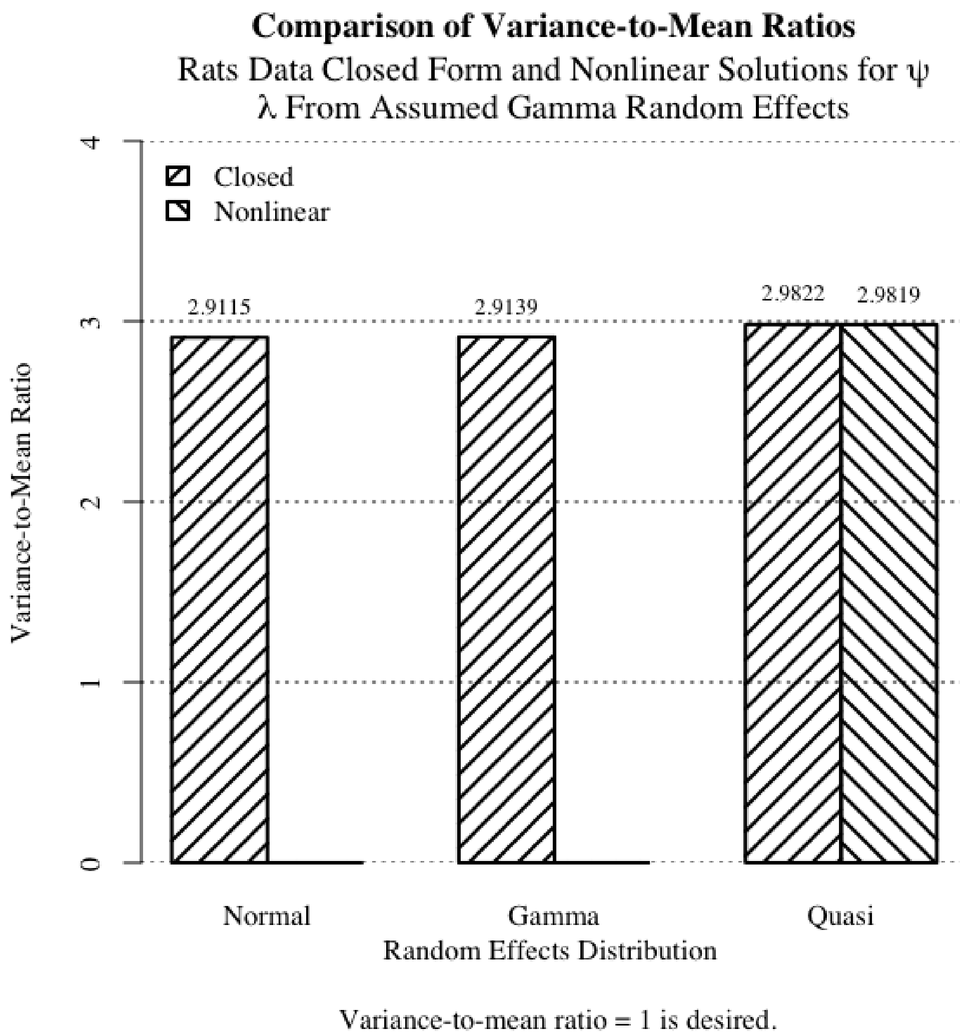

Figure 4.6 Variance-to-mean ratios for the normal, gamma, and quasi distributions for the rat data.

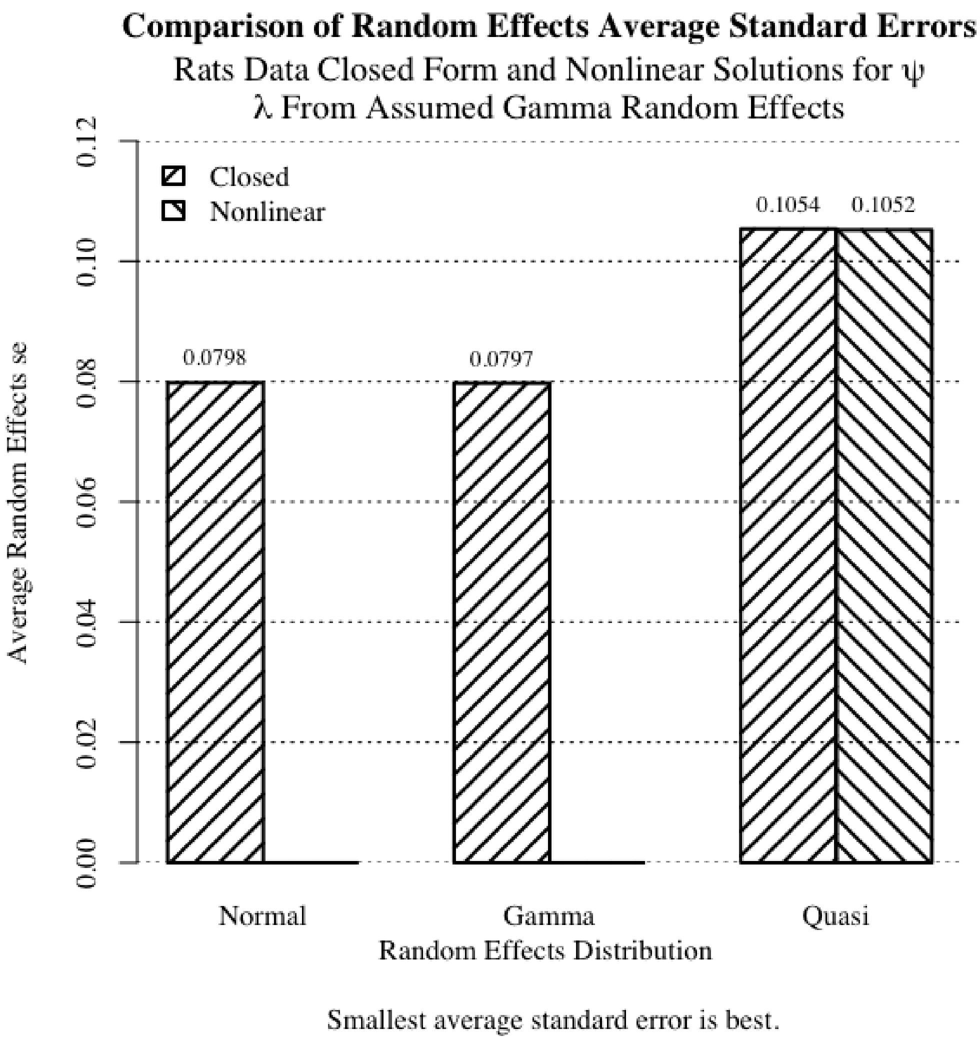

Figure 4.7 Average standard errors for the normal, gamma, and quasi distributions for the rat data.

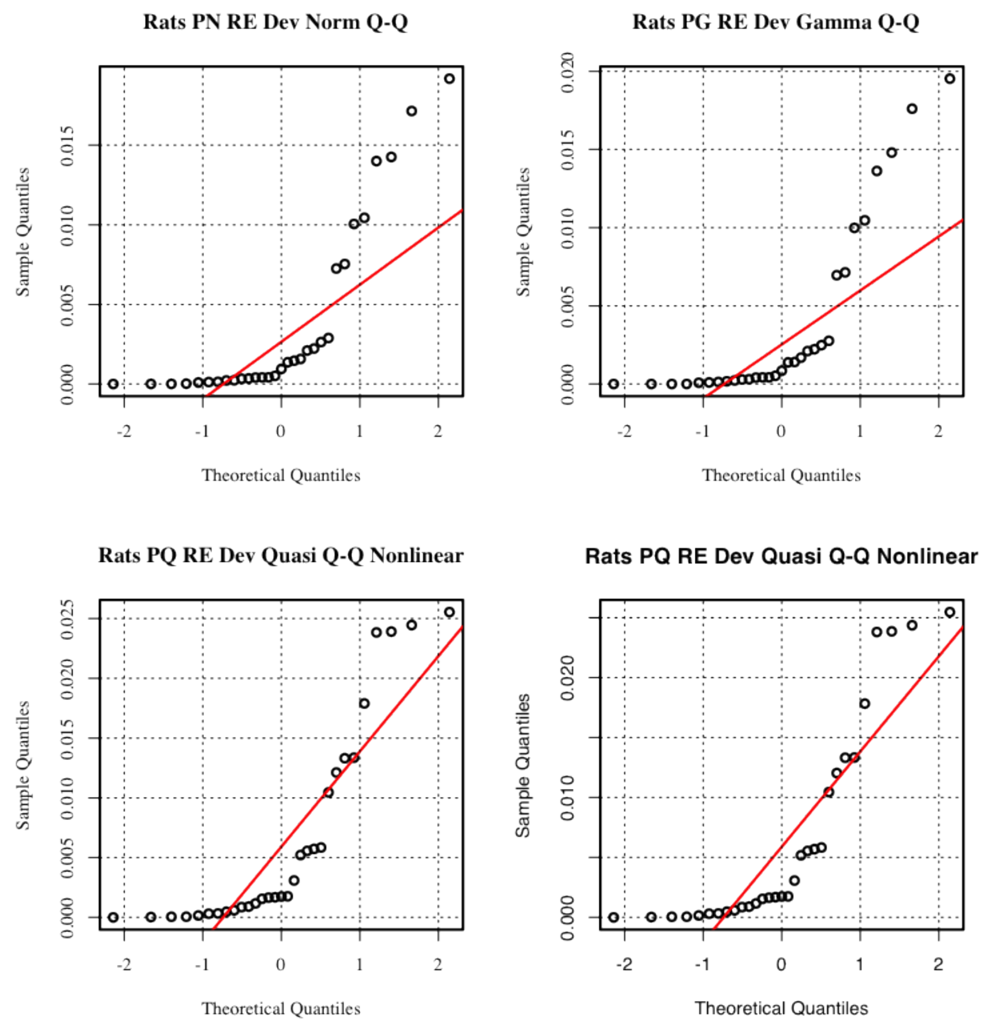

Figure 4.8 Q-Q normal plots of the rat data closed form and nonlinear solutions for the normal, gamma, and the two quasi distributions. The normal Q-Q plots indicate neither the normal distribution, nor the gamma distribution, nor the two quasi distributions are normally distributed, suggesting the fits using these distributions may be improved. PN is Poisson-normal, PG is Poisson-gamma, RE is random effects, Dev is deviance, and Norm is normal.

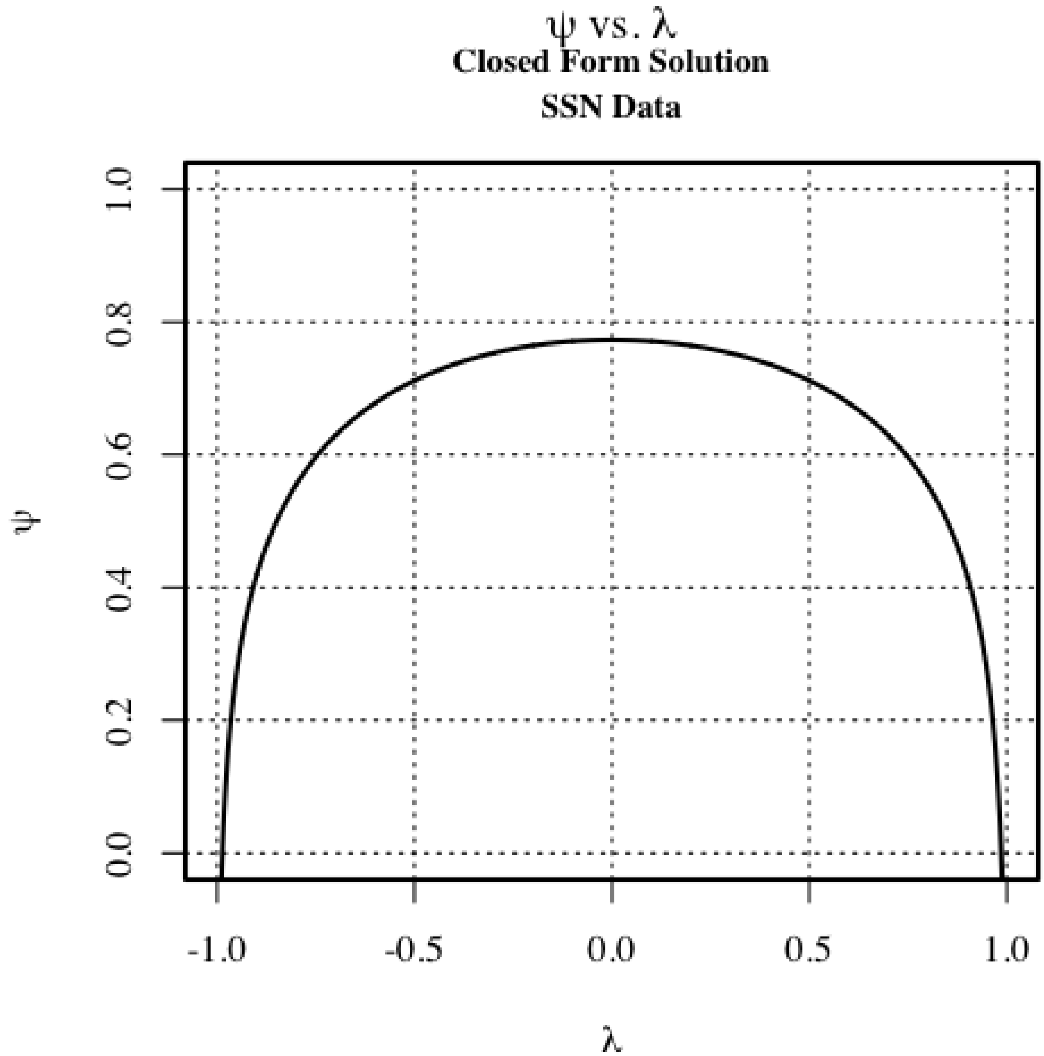

Figure 4.9 Power function exponent ψ versus the power normal truncation parameter λ for the sunspot data (SSN). The maximum of the downward-opening curve of $\hat{\psi} = 0.7734$.

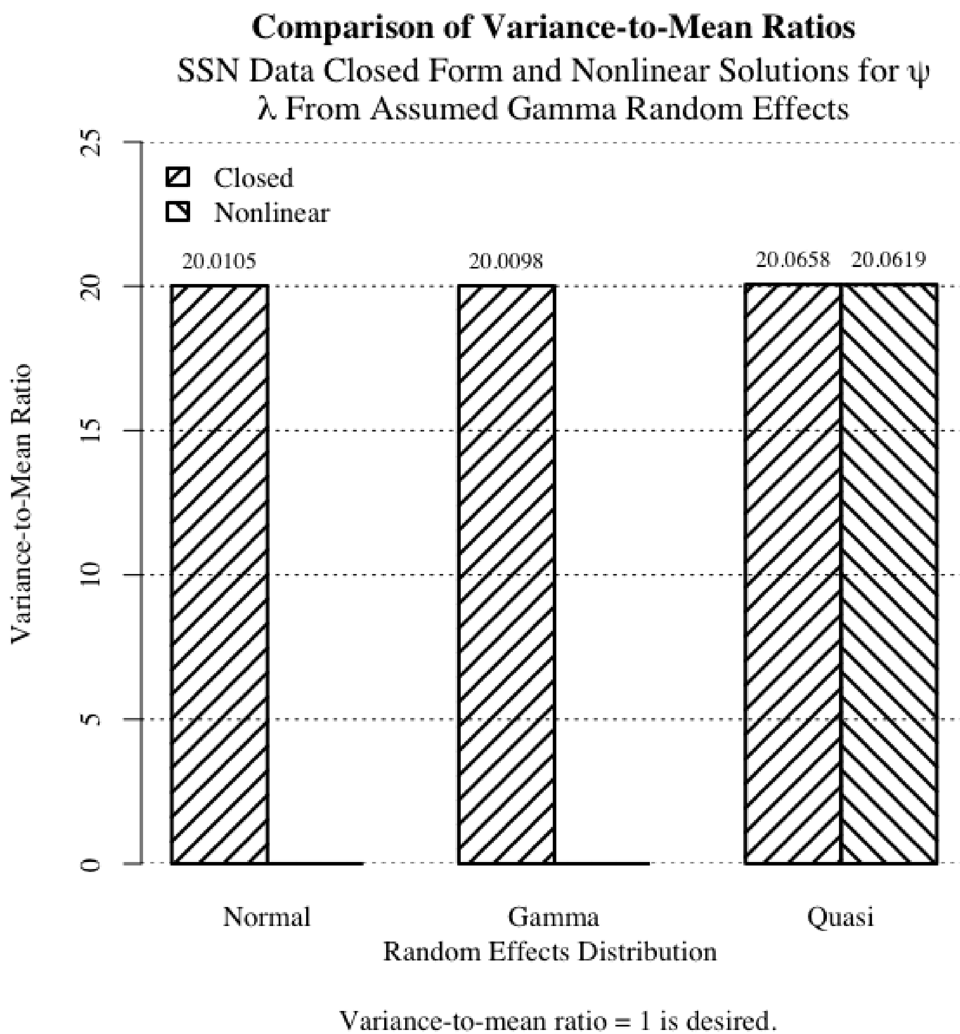

Figure 4.10 Variance-to-mean ratios for the normal, gamma, and the closed form and empirical quasi distributions for the sunspot data (SSN). No one distribution is suggested as the best one for modeling the sunspot random effects.

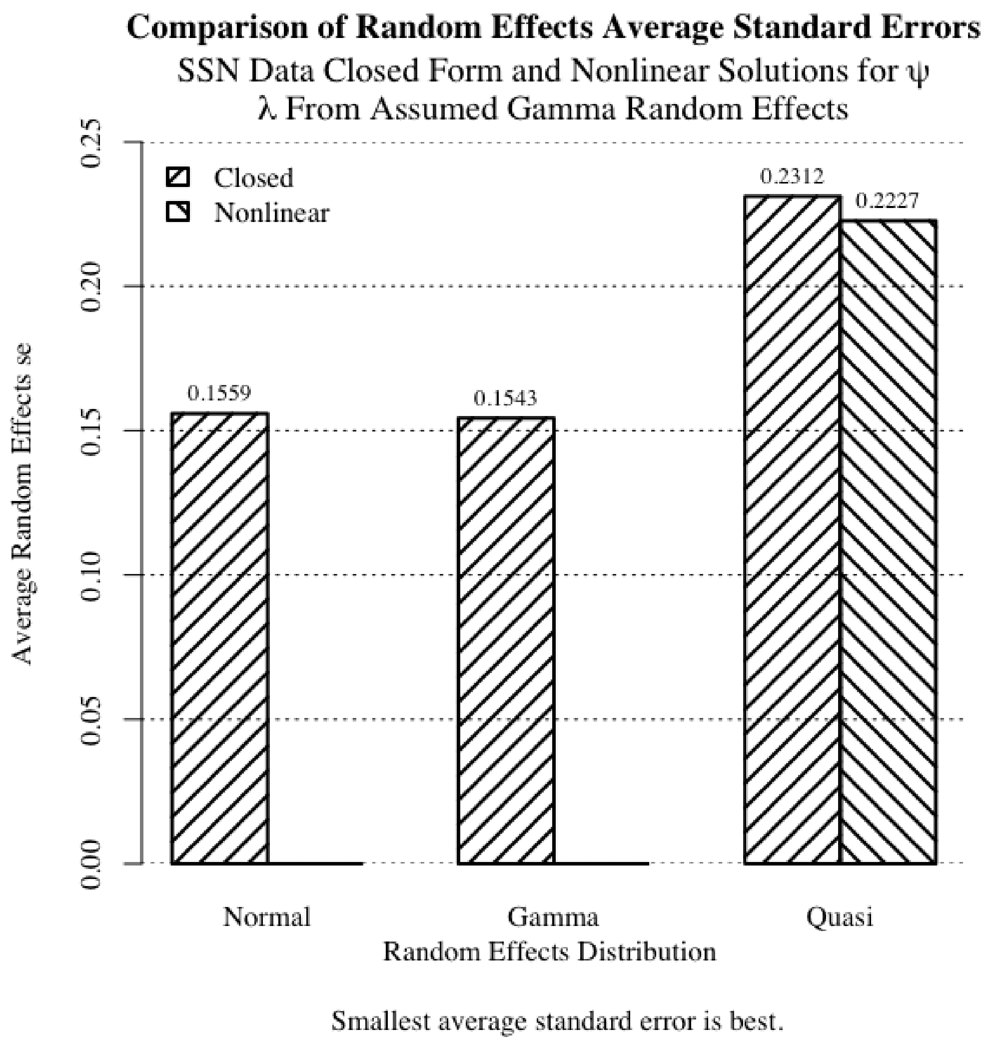

Figure 4.11 Average standard errors for the normal, gamma, and quasi distributions for the sunspot data (SSN). The plot suggests the two quasi distributions are less desirable estimators of the random effects.

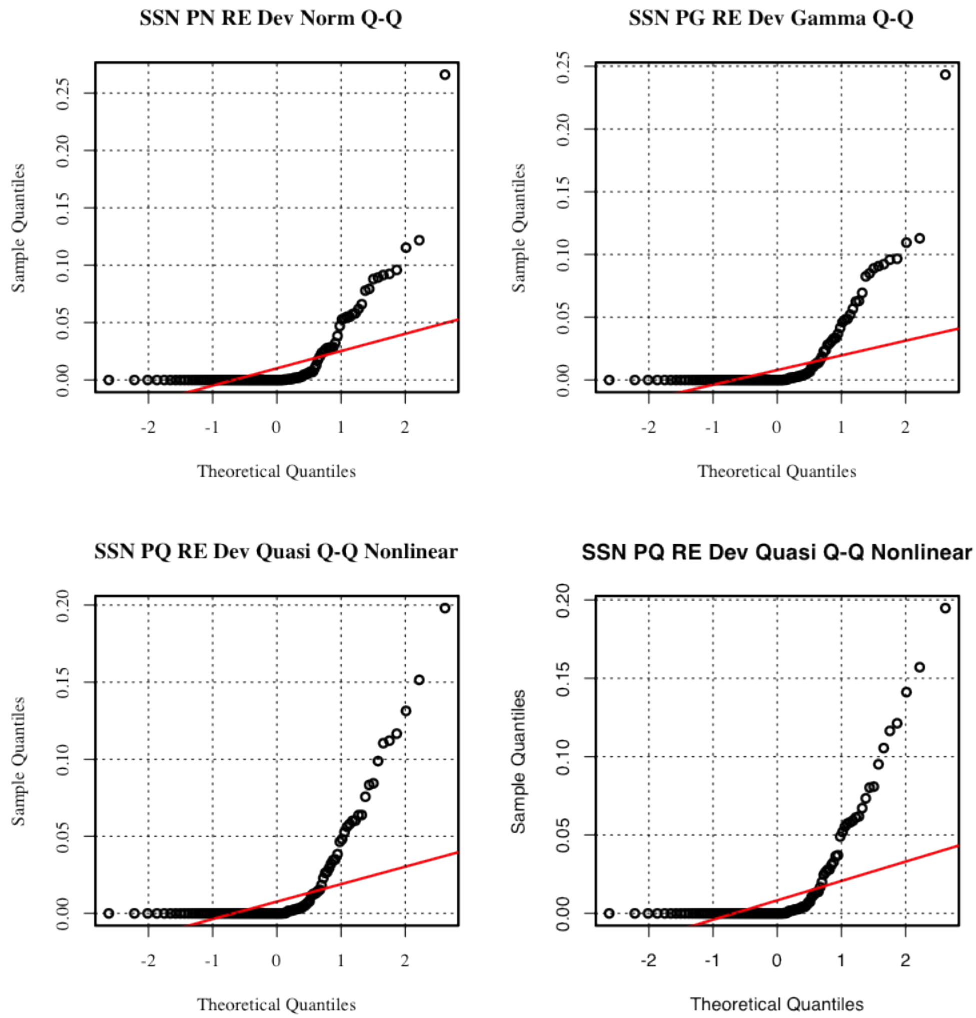

Figure 4.12 Q-Q normal plots of the sunspot data closed form and nonlinear solutions for the normal, gamma, and the two quasi distributions. The plots indicate none of the distributions is normal, suggesting the fits using these distributions may be improved. Note that the two quasi distributions do not have the possible outlier of the normal and gamma distribution random effects standardized deviance residuals. SSN is sunspot number, PN is Poisson-normal, PG is Poisson-gamma, RE is random effects, Dev is deviance, and Norm is normal.

Chapter 5

Conclusions

Selecting random effects distributions using a quasi-likelihood approach in this dissertation utilized components of generalized linear models (GLMs), creative incorporation of the power normal distribution in conjunction with the Box-Cox transformation, finding closed form and empirical solutions for a power function mean-variance relationship quasi distribution, and examination of efficacy measures for ascertaining the viability of the quasi distribution relative to the normal and gamma distributions. The selection methodology considers whether the exponent in the power function quasi distribution can be estimated, whether the power function quasi distribution affects overdispersion, random effects standard errors, and the the random effects model fit demonstrated by goodness-of-fit from the deviance residuals.

A review of the literature on generalized linear models showed that no work had been done on using a power function mean-variance relationship quasi distribution to characterize the random effects in hierarchical generalized linear models. The need to explore the performance properties of the power function quasi distribution for random effects came from an analysis of sunspot counts data. This research thereby was restricted to Poisson-distributed responses conditional upon random effects that group these responses into homoscedastic clusters (Figure 5.1). Distributions of random effects that best describe this clustering is the subject of this research, where the distributions of interest are the normal, gamma, and power function mean-variance relationship quasi distributions.

The properties of the normal and gamma distributions as applied to random effects are well known, unlike those of the power function quasi distribution. The power function mean-variance relationship quasi distribution for random effects u_i

Effect	Mean Model Components	Dispersion Model Components
Fixed	$Y_{ij} \mid u_i \sim (\mu_{ij}, a(\phi_{ij})V(\mu_{ij}))$	$d_{ij} \sim (\phi_{ij}, 2\phi_{ij}^2)$
	$\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{v}$	$\eta_{dij} = \gamma_0$
	$\eta_{ij} = g(\mu_{ij})$	$\eta_{dij} = g_d(\phi_{ij})$
Random	$u_i \sim (\mu_{Ri}, \zeta_i\mu_{Ri}^{\psi})$	$d_{R_i} \sim (\zeta_i, 2\zeta_i^2)$
		$\eta_{dRi} = \delta_0$
	$v_i = g_R(u_i)$	$\eta_{dRi} = g_{dR}(\zeta)$

Figure 5.1 The HGLM for parameter estimation HGLMs with random effects power function quasi distribution.

is

$$u_i \sim (\mu_{Ri}, \mu_{Ri}^{\psi}), \tag{5.1}$$

where u_i is the i^{th} random effect, μ_{Ri} is the i^{th} random effect mean, and ψ is the exponent of the power function on μ_{Ri} that defines the quasi distribution variance. An unknown, ψ must be estimated before the quasi distribution may be applied to the random effects for performance comparisons with the normal and gamma distributions.

The estimation of ψ was derived from the expected value of the power function random effects deviance residuals,

$$\mathcal{E}(d_{Ri}) = 2\left[\frac{\mathcal{E}(\xi^{2-\psi}) - (2-\psi)\mathcal{E}(\xi^{1-\psi})\mathcal{E}(\xi_i) + (1-\psi)\mathcal{E}(\xi^{2-\psi})}{(1-\psi)(2-\psi)}\right]. \tag{5.2}$$

Notice that ψ appears on the pseudo-response ξ_i that is used in the iterated weighted least squares (IWLS) parameters, used by Lee & Nelder (1996) to solve hierarchical GLMs, as the dependent variable for the random effects submodel. The pseudo-response also appears on the mean of random effects which is estimated by the expected value of the pseudo-response ξ_i. In addition, ψ appears in the denominator of Equation 5.2 as $(1-\psi)(2-\psi)$. While values of ξ_i are available from the IWLS algorithm, there remains finding a way to solve for ψ.

This research proposed that ψ can be estimated when the $\mathcal{E}(d_{Ri})$ is equated to the first moment approximation of the mean of the power normal distribution. The power normal distribution was proposed as it shares two properties often manifested by random effects deviance residuals: left truncation and right skewness.

The first moment approximation of the mean of the power normal distribution is a function of the power normal truncation parameter, λ, and the random effects deviance residuals as

$$\mu_{PNi} \doteq (1 - \lambda^2) d_{Ri}, \tag{5.3}$$

where μ_{PNi} is the mean of the power normal distribution. The truncation parameter is estimated from the random effects deviance residuals, d_{Ri}, and is the λ parameter from the Box-Cox transformation on these residuals (Box & Cox, 1964).

A solution for the power function quasi distribution exponent ψ is now possible when the expected value of the random effects deviance residuals are equated with the first moment approximation of the mean of the power normal distribution, thus,

$$2\left[\frac{\mathcal{E}(\xi^{2-\psi}) - (2-\psi)\mathcal{E}(\xi^{2-\psi}) + (1-\psi)\mathcal{E}(\xi^{2-\psi})}{(1-\psi)(2-\psi)}\right] = (1-\lambda^2) d_{Ri}. \tag{5.4}$$

Excepting ψ, all the terms in this equation have values resulting from the application of the IWLS algorithm. However, even though values are available, the term $\mathcal{E}(\xi_i^{2-\psi})$ is intractable.

By applying the Taylor approximation to $\mathcal{E}(\xi_i^{2-\psi})$ about the estimated mean of the random effects, μ_{Ri} , a closed form solution for ψ is obtained. Thus,

$$\hat{\psi} = \frac{\ln[Var(\xi_i)] - \ln(1 - \hat{\lambda}^2) - \ln(\bar{d}_R)}{\ln(\bar{\xi})}, \tag{5.5}$$

where $\bar{\xi}^{-\psi}$ is the mean of the pseudo-response ξ_i of the random effects model from the IWLS algorithm, $\bar{d}_R$ is the mean of the random effects deviance residuals, and $\hat{\lambda}$ is the power normal truncation parameter. This closed form may be algebraically manipulated to obtain an empirical, nonlinear regression equation to solve for ψ, assuming the ξ_i and d_{Ri} are independent. Thus,

$$d_{Ri} = \frac{Var(\xi_i)}{1 - \hat{\lambda}^2} \xi_i^{-\psi}. \tag{5.6}$$

Therefore, two estimates for ψ are obtained, and the answer to the first research question is that ψ can be estimated.

These two estimates for ψ now give specific values to the power function exponent, and the random effects quasi distribution becomes

$$u_i \sim (\mu_{Ri}, \mu_{Ri}^{\hat{\psi}}). \tag{5.7}$$

Random effects models by the normal distribution, the gamma distribution, and the quasi distribution may now be compared. To answer the second and third research questions, comparisons among these three distributions were made using three data sets assuming Poisson-distributed conditional responses, and random effects that were modeled by the three distributions.

The three data sets used are the fabric data of Bissell (1972), the rats data of Myers & Montgomery (2002), and sunspot counts data provided by the American Association of Variable Star Observers (AAVSO) Solar Section. The fabric data and the rats data appear in the literature as example data for model comparisons. The sunspot data motivated this power function quasi distribution research.

These three data sets each compared the normal, gamma, and quasi distributions using the efficacy measures from research questions 2 and 3; namely, the variance-to-mean ratios (overdispersion measure) and the estimated average random effects standard errors. In addition, a third goodness-of-fit criterion was used: the normal Q-Q diagnostic plots of the standardized random effects deviance residuals, and the correlative Shapiro-Wilk statistics for assessing normality. The comparisons outcomes varied by data set. The fabric data HGLM completely failed to converge using the power normal distribution, power function exponent estimating method, thereby giving no model for the comparisons. The sunspot data and the rat data gave ambivalent results on the three distributions. The answer to research question 2 is that the quasi distribution did not reduce overdispersion over the normal and gamma distributions. The answer to research question 3 is that the estimated average random effects standard errors were not reduced over those of the normal and gamma distributions.

Even though the performance comparisons did not show the quasi distributions to be efficacious over the normal and gamma distributions, they did demonstrate that the power function quasi distribution does influence random effects model estimates and the model goodness-of-fit plots and Shapiro-Wilk statistics. Hence it is reasonable to include the quasi distribution as a random effects model candidate. If the quasi distribution is included as a candidate, it has the capability of suggesting either a normal or a gamma distribution may be preferred.

While the power function implementation of the quasi distribution in this research on the chosen data sets was no more efficacious than either the normal and gamma distributions as random effects models, additional research is needed to allow an extended range of the power function exponent estimates of ψ below 0 and above 1. Recall that the expected value of the random effects deviance residuals was equated to the first moment approximation of the mean for the power normal distribution. This approximation includes the term $1 - \hat{\lambda}^2$, which constrains the power function exponent as $0 \leq \hat{\psi} \leq 1$. This precludes $\hat{\psi}$ from values outside 0 and 1. A possible remedy is to use the maximum likelihood estimates for the power normal distribution, though consideration must be made to account the lack of consistency in these estimates (Maruo et al., 2011).

Overdispersion in the conditional response mean model and the random effects mean model are, by McCullagh & Nelder (1989) and Sakate & Kashid (2012), independent of each other. Further work may be applied to understanding the influence of the power normal distribution power function exponent on random effects overdispersion.

Another area for investigation is to try other distributions to equate to the expected value of the random effects deviance residuals when they are from the power function mean-variance relationship. These candidate distributions must account for the frequent appearance of left-truncation and the right-skewness in the random effects deviance residuals.

A wholly different approach to estimating ψ is to use an iterative methodology that bypasses distributional estimation of the expected value of the random effects deviance residuals from the power function, namely using a profile estimation method. For example, increment $\hat{\psi}$ from -3 to +3 in increments of 0.1, and at each increment of $\hat{\psi}$, test $u_i \sim (\mu_{Ri}, \mu_{Ri}^{\hat{\psi}})$ in the Poisson-quasi hierarchical GLM. Select the quasi distributions with the smallest estimated average random effects standard errors, and compare these to the Poisson-normal and Poisson-gamma estimates. When the fabric data power normal distribution power function exponent estimating method failed to converge, hence resulting in no model, an arbitrary choice of ψ values less than 0.5 did allow the fabric data HGLM to converge. This is evidence for using the profile method.

Consideration may be given to using the power function quasi distribution to model functions of random effects predictors. A polynomial function was used in the random effects model of the rat data. The predictor was the white blood cell count, and a second order polynomial was used. Such a functional relationship may be extended to random effects with continuous data; that is, investigate continuous random effects as an analysis of covariance.

Though this dissertation's investigation into a power function mean-variance relationship quasi distribution covered a specific methodology for selecting the quasi distribution, the outcomes of the Poisson-quasi models encourage further work in this area. Even so, the specific methodology used in this research for quasi distribution selection clearly indicates that the power normal distribution estimates for the expected value of the power function deviance residuals is a data set-specific method for use in Poisson-quasi generalized linear models, but can suggest the random effects may be modeled by either a normal distribution, a gamma distribution, or a Poisson distribution. The power function quasi distribution thus adds options in modeling counts-specific hierarchical generalized linear model random effects.

Bibliography

Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using glim. *Applied Statistics*, *36*, 332–339.

Alam, M., Ronnegard, L., & Shen, X. (2010). hglm: Hierarchical generalized linear models. World Wide Web.

Anderson, D. (1988). Some models for overdispersed binomial data. *Australian Journal of Statistics*, *30*, 125–148.

Bain, L. J. & Engelhardt, M. (1992). *Inroduction to Probability and Mathematical Statistics* (2nd ed.). Pacific Grove, California: Duxbury.

Belsley, D., Kuh, E., & Welsch, R. (1980). *Regression Diagnostics*. New York: John Wiley & Sons, Inc.

Bissell, A. (1972). A negative binomial model with varying element sizes. *Biometrics*, *59*, 435–441.

Box, G. & Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society B*, *26*(2), 211–252.

Box, G., Hunter, W., & Hunter, J. (1978). *Statistics for Experimenters*. New York: John Wiley & Sons, Inc.

Breslow, N. (1984). Extra-poisson variation in log-linear models. *Applied Statistics*, *33*, 38–44.

Carroll, R. J. & Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. *Annals of Statistics*, *10*(2), 429–441.

Cox, D. (1983). Some remarks on overdispersion. *Biometrika*, *70*(1), 269–274.

Crowder, M. (1978). Beta-binomial anova for proportions. *Applied Statistics*, *27*, 34–37.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, *81*(395), 709–721.

Freeman, J. & Modarres, R. (2006a). Inverse box-cox: The power-normal distribution. *Statistics & Probability Letters*, *76*, 764–772.

Freeman, J. & Modarres, R. (2006b). Properties of the power-normal distribution. Technical report, Department of Statistics, George Washington University.

Goodnight, J. H. (1978). Computing mivque0 estimates of variance components. SAS Technical Report R-105 R-105, SAS Institute Inc., Cary, NC.

Griffiths, D. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. *Biometrics*, *25*, 489–504.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, *31*(2), pp. 423–447.

Jorgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society B*, *49*(2), 127–162.

Koehler, K. & Wilson, J. (1986). Chi-square tests for comparing vectors or proportions for several cluster samples. *Communications in Statistics*, *15*, 2977–2990.

Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society B*, *58*(4), 619–678.

Lee, Y. & Nelder, J. A. (2000). Two ways of modelling overdispersion in non-normal data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *49*(4), 591–598.

Lee, Y. & Nelder, J. A. (2001). Hierarchical generalised linear models : A synthesis of generalised linear models, random-effect model and structured dispersion. *Biometrika*, *88*, 987–1006.

Lee, Y. & Nelder, J. A. (2006). Double hierarchical generalized linear models (with discussion). *Applied Statistics*, *55*, 139–185.

Lee, Y., Nelder, J. A., & Pawitan, Y. (2006). *Generalised linear models with random effects: unified analysis via h-likelihood.* London: Chapman & Hall.

Maruo, K., Shirahata, S., & Goto, M. (2011). Underlying assumptions of the power-normal distribution. *Behaviormetrika*, *38*(1), 85–95.

McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics*, *11*(1), pp. 59–67.

McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models*. Boca Raton, Florida: Chapman & Hall/CRC.

Milliken, G. A. & Johnson, D. E. (1992). *Analysis of Messy Data Volume I: Designed Experiments*, volume 1. Boca Raton, Florida: Chapman & Hall/CRC.

Myers, P. & Montgomery, D. (2002). *Generalized Linear Models with Applications in Engineering and the Sciences*. New York: John Wiley & Sons, Inc.

Nelder, J. (1990). Nearly parallel lines in residual plots. *The American Statistician*, *44*(3), 221–222.

Nelder, J. & Lee, Y. (1991). Generalised linear models for the analysis of taguchi-type experiments. *Applied Stochastic Models and Data Analysis*, *7*, 107–120.

Nelder, J. & Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, *74*(2), 221–232.

Nelder, J. & Wedderburn, R. (1972). Generalised linear models. *Journal of the Royal Statistical Society A*, *135*, 370–384.

Pierce, D. A. & Schafer, D. W. (1986). Residuals in generalized linear models. *Journal of the American Statistical Association*, *81*(396), 977–986.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, *9*, 705–724.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Ronnegard, L., Shen, X., & Alam, M. (2010). hglm: A package for fitting hierarchical generalized linear models. *The R Journal*, *2*(2).

Ronnegard, L., Shen, X., & Alam, M. (2013). Package 'hglm'. World Wide Web.

Sakate, D. M. & Kashid, D. N. (2012). A deviance-based criterion for model selection in glm. *Statistics*, *0*(0), 1–15.

SAS (2011). *SAS Institute Inc. 2011. SAS/STAT® 9.3 User's Guide* (SAS/STAT® 9.3 ed.). Cary, NC: SAS Institute Inc.

Schaefer, B. E. (1993). Visibility of sunspots. *Astrophysical Journal*, *411*, 909–919.

Shapley, A. (1949). Reduction of sunspot-number observations. *Publication of the Astronomical Society of the Pacific*, *61*(358), 13–21.

Smyth, G. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society B*, *51*, 47–60.

Taylor, P. O. (1985). The Comptation of American Relative Sunspot Numbers. *Journal of the American Association of Variable Star Observers*, *14*, 28 – 32.

Waldmeier, M. (1961). *The sunspot-activity in the years 1610-1960*. Zurich: Verlag Schulthess u. Co. AG.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika*, *61*(3), 439–447.

Williams, D. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics*, *31*, 949–952.

Williams, D. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics*, *31*, 144–148.

Wilson, J. (1989). Chi-square tests for overdispersion with multiparameter estimates. *Applied Statistics*, *38*, 441–453.

Wilson, J. & Koehler, K. (1991). Hierarchical models for cross-classified overdispersed multinomial data. *Journal of Business & Economic Statistics*, *9*(1), 103–110.

Yanez, N. (1993). *Modeling the Effects of Overdispersion in Generalized Linear Models*. PhD thesis, Arizona State University.

Yanez, N. & Wilson, J. (1995). Comparison of quasi-likelihood models for overdispersion. *Australian Journal of Statistics*, *37*(2), 217–231.

Appendices

Appendix A

Derivation of the Power Normal First Moment

In Chapter 2, a moment generating function for the Power Normal (PN) distribution was obtained from Freeman & Modarres (2006a) and Freeman & Modarres (2006b). Here, the series approximation is expanded to give the result in Equation 3.13:

$$\mu = \mathcal{E}X^1 = \begin{cases} (\lambda y+1)^{\frac{1}{\lambda}}(1-\lambda^2) & ,\lambda \neq 0 \\ \exp(\mu+\sigma^2/2) & ,\lambda = 0, \end{cases} \tag{A.1}$$

where μ is as defined in Chapter 2. The moment generating function series approximation is

$$\mathcal{E}X^r = \begin{cases} \sum_{\text{even } k\geq 0} \frac{\sigma^k k!}{s^{\frac{k}{2}}\left(\frac{k}{2}\right)!}(\lambda y+1)^{\frac{r}{\lambda}-k}\prod_{l=1}^{k-1}(r-l\lambda) & ,\lambda \neq 0 \\ \exp(r\mu+r^2\sigma^2/2) & ,\lambda = 0. \end{cases} \tag{A.2}$$

The following are the calculations for obtaining Equation A.1.

Firstly, calculate the $\prod_{l=1}^{k-1}(r-l\lambda)$ term in Table A.1; secondly, calculate the $(\lambda y+1)^{\frac{r}{\lambda}-k}$ term in Table A.2, and thirdly calculate $\frac{\sigma^k k!}{2^{k/2}(k/2)!}$ term in Table A.3.

Table A.1 *Power Normal First Moment* $\prod_{l=1}^{k-1}(r-l\lambda)$, $\lambda \neq 0$ *Term for* $\mathcal{E}X^r$.

r	k	k-1	$\prod_{l=1}^{k-1}(r-l\lambda),\ l=1,2,\ldots,k-1$
1	0	-1	$(1-\lambda)(1-0)(1+\lambda)=1-\lambda^2$
	2	1	$1-\lambda$
	4	3	$(1-\lambda)(1-2\lambda)(1-3\lambda)=1-6\lambda+11\lambda^2-6\lambda^3$
2	0	-1	$(2-\lambda)(2-0)(2+\lambda)=2(4-\lambda^2)$
	2	1	$2-\lambda$
	4	3	$(2-\lambda)(2-2\lambda)(2-3\lambda)=8-36\lambda-8\lambda^2-12\lambda^3$
	6	5	$(2-\lambda)(2-2\lambda)(2-3\lambda)(2-4\lambda)(2-5\lambda)$
			$=(8-36\lambda-8\lambda^2-12\lambda^3)(4-18\lambda+20\lambda^2)$

Table A.2 *Power Normal First Moment* $(\lambda y+1)^{\frac{r}{\lambda}-k}, \quad \lambda \neq 0$ *erm for* $\mathcal{E}X^r$.

r	k	$(\lambda y+1)^{\frac{r}{\lambda}-k},\ k=0,2,4,\ldots$
1	0	$(\lambda y+1)^{\frac{r}{\lambda}}$
	2	$(\lambda y+1)^{\frac{r}{\lambda}-2}$
	4	$(\lambda y+1)^{\frac{r}{\lambda}-4}$
2	0	$(\lambda y+1)^{\frac{r}{\lambda}}$
	2	$(\lambda y+1)^{\frac{r}{\lambda}-2}$
	4	$(\lambda y+1)^{\frac{r}{\lambda}-4}$
	6	$(\lambda y+1)^{\frac{r}{\lambda}-6}$

Table A.3 *Power Normal First Moment* $\frac{\sigma^k k!}{2^{k/2}(k/2)!}, \quad \lambda \neq 0$ *Term for* $\mathcal{E}X^r$.

k	$\frac{\sigma^k k!}{2^{k/2}(k/2)!},\ k=0,2,4,\ldots$
0	1
2	$\frac{\sigma^2\cdot 2\cdot 1}{2^1\cdot 1}=\sigma^2$
4	$\frac{\sigma^4 4\cdot 3\cdot 2\cdot 1}{2^2 2\cdot 1}=2\sigma^4$
6	$\frac{\sigma^6 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{2^3 3\cdot 2\cdot 1}=15\sigma^6$

Table A.4 *Power Normal First Moment*
$\sum_{\text{ven } k\geq 0} \frac{\sigma^k k!}{s^{\frac{k}{2}}\left(\frac{k}{2}\right)!}(\lambda y+1)^{\frac{r}{\lambda}-k}\prod_{l=1}^{k-1}(r-l\lambda), \quad \lambda \neq 0$

r	k	$\frac{\sigma^k k!}{2^{k/2}(k/2)!}$	$(\lambda y+1)^{\frac{r}{\lambda}-k}$	$\prod_{l=1}^{k-1}(r-l\lambda)$
1	0	1	$(\lambda y+1)^{\frac{r}{\lambda}}$	$1-\lambda^2$
	2	σ^2	$(\lambda y+1)^{\frac{r}{\lambda}-2}$	$1-\lambda$
	4	$2\sigma^4$	$(\lambda y+1)^{\frac{r}{\lambda}-4}$	$1-6\lambda+11\lambda^2-6\lambda^3$
2	0	1	$(\lambda y+1)^{\frac{r}{\lambda}}$	$1-\lambda^2$
	2	σ^2	$(\lambda y+1)^{\frac{r}{\lambda}-2}$	$1-\lambda$
	4	$2\sigma^4$	$(\lambda y+1)^{\frac{r}{\lambda}-4}$	$1-6\lambda+11\lambda^2-6\lambda^3$
	6	$15\sigma^6$	$(\lambda y+1)^{\frac{r}{\lambda}-6}$	$(8-36\lambda-8\lambda^2-12\lambda^3)(4-18\lambda+20\lambda^2)$

Appendix B

HGLM for power function quasi distribution

This HGLM implementation derives from Ronnegard, Shen & Alam (2013). The basic change to Ronnegard et al. is the use of the R family such that $family = quasi(link = "log", variance = "mu^pwr", pwr = psi)$, the power normal truncation parameter λ derived from the Box-Cox transformation the random effects d_{Ri} to specify the power function mean-variance relationship quasi distribution. Also, the modification outputs the random effects pseudo-response from the augmented response vector.

```
######################################################################
#####     hglm allowing power fct mean-var relation     ##############
######################################################################

# This hglm implementation derives from Ronnegard, et.al. 2011 (see
# ref below).
# The basic change to Ronnegard, et.al. is the use of the R family as
# family=quasi(link=power(boxcox)), the Box-Cox x4m lambda on dri to
# specifiy the random effects mean-variance relationship.

# @url{ronnegard11,
# Author = {Ronnegard, Lars and Shen, Xia and Alam, Moudud},
# Title = {The hglm Package (version 1.2)},
# Urldate = {20110704}}

# Created  28 May 2011, Jamie Riggs
# Modified
#          05 Sep 2013, Jamie Riggs,
#               mvr <- quasipf(link="log", variance="mu^pwr", pwr=psi)
#          18 Aug 2013, Jamie Riggs, bar plots, consolidate model
#               output
#          05 Apr 2013, Jamie Riggs, modified for PhD research
#          21 Mar 2013, Jamie Riggs, clean-up, matches lmer
#          28 Jan 2012, Jamie Riggs, added variance by mean plot
#          01 Oct 2011, Jamie Riggs, Parametric statistics only
#          22 Aug 2011, Jamie Riggs, modified for regression trend
#               removal
#          07 Jun 2011, Jamie Riggs, random effects (Month) model

# Model designations
# m00 - m09: normal-normal
# m10 - m19: Poisson-normal
# m20 - m29: Poisson-gamma
# m30 - m39: Poisson-quasi


######################################################################
#####     Libraries     ##############################################
######################################################################
```

```
library(MASS)
library(stats)
library(lattice)
library(Matrix)
library(hglm)    # use for support functions that needn't be modified
library(nlme)
library(nlstools)
library(boot)
library(xtable)


######################################################################
#####     Functions     ##############################################
######################################################################

setwd("/Users/Jamie/Desktop/SPES/SSN/R/")
WD <- getwd()
fn <- paste(WD,"functions.R", sep="/")
source(fn)
setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()


setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm/R/hglm")
WD <- getwd()
fn <- paste(WD,"bct.R", sep="/")
source(fn)
setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()


#####     ghglm     ##################################################

 sourceDir <- function(path, trace = TRUE, ...) {
    for (nm in list.files(path, pattern = "[.][RrSsQq]$")) {
       if(trace) cat(nm,":")
       source(file.path(path, nm), ...)
       if(trace) cat("\n")
    }
 }

setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm/R/ghglm")
WD <- getwd()
fn <- paste(WD,"GLM.MME.R", sep="/")
source(fn)
fn <- paste(WD,"ghglm.formula.R", sep="/")
source(fn)
fn <- paste(WD,"ghglm.default.R", sep="/")
```

```
source(fn)
fn <- paste(WD,"ghglm.R", sep="/")
source(fn)
#fn <- paste(WD,"summary.ghglm.R", sep="/")
#source(fn)
#fn <- paste(WD,"print.summary.ghglm.R", sep="/")
#source(fn)
#fn <- paste(WD,"plot.ghglm.R", sep="/")
#source(fn)
setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()

##-----------------------------------------------##
## redefine the quasi function. this is just R's    ##
## quasi function with one section added by me.     ##
##-----------------------------------------------##
quasipf <- function ( link = "identity",
                       variance = "constant",
                       pwr = NULL
                       ) {
    linktemp <- substitute(link)
    if (is.expression(linktemp) || is.call(linktemp))
        linktemp <- link
    else if (!is.character(linktemp))
        linktemp <- deparse(linktemp)
    if (is.character(linktemp))
        stats <- make.link(linktemp)
    else stats <- linktemp
    variancetemp <- substitute(variance)
    if (!is.character(variancetemp)) {
        variancetemp <- deparse(variancetemp)
        if (linktemp == "variance")
            variancetemp <- eval(variance)
    }
    switch(variancetemp, constant = {
        variance <- function(mu) rep.int(1, length(mu))
        dev.resids <- function(y, mu, wt) wt * ((y - mu)^2)
        validmu <- function(mu) TRUE
      },
      ##---added didn't change the deviance residuals---##
        "mu^pwr" = {
        variance <- function(mu) mu^pwr
        validmu <- function(mu) all(mu > 0)
        dev.resids <- function(y, mu, wt)
            2 * wt * (y * log(ifelse(y ==
            0, 1, y/mu)) + (1 - y) * log(ifelse(y == 1, 1, (1 -
            y)/(1 - mu))))
      },
      ##---end part added -------------------------------##
```

```
        "mu(1-mu)" = {
        variance <- function(mu) mu * (1 - mu)
        validmu <- function(mu) all(mu > 0) && all(mu < 1)
        dev.resids <- function(y, mu, wt)
            2 * wt * (y * log(ifelse(y ==
            0, 1, y/mu)) + (1 - y) * log(ifelse(y == 1, 1, (1 -
            y)/(1 - mu))))
      }, mu = {
        variance <- function(mu) mu
        validmu <- function(mu) all(mu > 0)
        dev.resids <- function(y, mu, wt)
           2 * wt * (y * log(ifelse(y ==
            0, 1, y/mu)) - (y - mu))
      }, "mu^2" = {
        variance <- function(mu) mu^2
        validmu <- function(mu) all(mu > 0)
        dev.resids <- function(y, mu, wt)
           pmax(-2 * wt * (log(ifelse(y ==
            0, 1, y)/mu) - (y - mu)/mu), 0)
      }, "mu^3" = {
        variance <- function(mu) mu^3
        validmu <- function(mu) all(mu > 0)
        dev.resids <- function(y, mu, wt)
            wt * ((y - mu)^2)/(y *
            mu^2)
      }, stop(
           paste(variancetemp, "not recognised, possible variances",
           "are \"mu(1-mu)\", \"mu\", \"mu^2\", \"mu^3\
           "and \"constant\"")))
    initialize <- expression({
        n <- rep.int(1, nobs)
        mustart <- y + 0.1 * (y == 0)
    })
    aic <- function(y, n, mu, wt, dev) NA
    structure(list(family = "quasi", link = linktemp,
        linkfun = stats$linkfun, linkinv = stats$linkinv,
        variance = variance, dev.resids = dev.resids,
        aic = aic, mu.eta = stats$mu.eta, initialize = initialize,
        validmu = validmu, valideta = stats$valideta,
        varfun = variancetemp),
        class = "family")
   }

######################################################################
#####     Initialization     ########################################
######################################################################

#####     Sunspot data     ##########################################
```

```
 Ex <- "201204"
ver <- "00"

setwd("/Users/Jamie/Desktop/SPES/SSN/Data/")
WD <- getwd()
load(paste(WD, "/", "ymtd", Ex, ver, ".RData", sep=""))  # loads as X
summary(X)
nrow(X)
SSN <- X
setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()


#####     Fabric data     ############################################

Ex <- "Fabric"

setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm/R/data/")
WD <- getwd()
X <- fetch(Ex)
summary(X)
nrow(X)
Fab <- X
setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()


#####     Rats data     ##############################################

Ex <- "Rats"

setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm/R/data/")
WD <- getwd()
X <- fetch(Ex)
X$Subject <- factor(X$Subject)
X$Drug <- factor(X$Drug)
X$Period <- factor(X$Period)
summary(X)
nrow(X)
Rat <- X
setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()


######################################################################
     part <- "W"     #################################################
######################################################################

     y <- X$w ; y <- x$cal
```

```
    yn <- part
     n <- length(y)
median <- median(y,na.rm=TRUE)
   P25 <- summary(y)[[2]]
   P75 <- summary(y)[[5]]
   min <- summary(y)[[1]]
   max <- summary(y)[[6]]
  mean <- mean(y,na.rm=TRUE)
   std <- sd(y,na.rm=TRUE)
   var <- var(y,na.rm=TRUE)
  rate <- mean / var
 shape <- rate * mean

  rspl <- rgamma(seq(0,max(y),0.1), shape=shape, rate=rate)
   spl <- sample(y,length(rspl))
  (kst <- ks.test(spl,rspl))
  ksgs <- round(kst[[1]],4)
  ksgp <- round(kst[[2]],4)

  rspl <- rexp(seq(0,max(y),0.1), rate=1/mean)
   spl <- sample(y,length(rspl))
  (kst <- ks.test(spl,rspl))
  kses <- round(kst[[1]],4)
  ksep <- round(kst[[2]],4)


(loc <- paste("Plots/", Ex, ver, part, "Hist", ".png", sep=""))
quartz(w=7, h=7, bg="white")
plot(y,type="l")
hp <- hist(y,prob=T, ylim=c(0,max(dgamma(seq(0,max(y),0.1),
      shape=shape, rate=rate), dexp(seq(0,max(y),0.1),
      rate=1/mean))),col="red",
      main=paste(part, "Wolf Number Distribution", sep=" "),
      sub=paste("Fitted  gam(",round(rate,3),", ", round(shape,3),")
      (solid black), exp(",round(1/mean,3),") (dahsed green), n=", n,
      sep=""), xlab=paste(part, "Counts", sep=" "))
gl <- lines(seq(0,max(y),0.1), dgamma(seq(0,max(y),0.1), shape=shape,
      rate=rate), lwd=3)
el <- lines(seq(0,max(y),0.1), dexp(seq(0,max(y),0.1), rate=1/mean),
      lwd=3, lty="dashed", col="green")
text(max(hp$mid), max(hp$density), pos=2,
     paste("Kolmogorov-Smirnov Test: Gamma"))
text(max(hp$mid), max(hp$density)-max(hp$density)/10, pos=2,
     paste("D=",max(ksgs,kses)))
text(max(hp$mid), max(hp$density)-max(hp$density)/5, pos=2,
     paste("p=",max(ksgp,ksep),"        "))
grid()
#quartz.save(loc, type="png")
```

```
y  <- X$w
x1 <- X$obs
x2 <- factor(X$year)
x3 <- factor(X$mon)
x4 <- X$see
x5 <- X$r
ym <- aggregate(X$w,by=list(year=X$year,mon=X$mon),se)
x6 <- as.numeric(factor(as.numeric(X$year)*100+X$mon,
      labels=1:(nrow(ym))))


######################################################################
     part <- "ghglm"     #############################################
######################################################################

# Model designations
# First digit  = distribution (normal=1, gamma=2,
#       quasi (closed, nonlinear)=3)
# Second digit = data set (Fabric=2, Rat=3, Sunspot=0)
# m00 - m09: normal-normal
# m10 - m19: Poisson-normal
# m20 - m29: Poisson-gamma
# m30 - m39: Poisson-quasi


#####     normal-normal     ##########################################

X <- SSN
m00 <- hglm(fixed = w ~ see + r + as.factor(year) + as.factor(mon),
#        disp = ~ as.factor(year) + as.factor(mon),
random = ~ 1|obs,
family = quasipoisson(link = log),
rand.family = gaussian(link = identity),
method = "EQL",
data = X)

m <- m00
mm <- "m00"
(ms <- summary(m))
(m00vm <-var(m$fv)/mean(m$fv))
(SeRem <- mean(m$SeRe))
(SeRes <- sd(m$SeRe))


#####     Poisson-normal     #########################################

#####     fabric model
  X <- Fab
```

```
 Ex <- "Fab"
ver <- "PN"
m12 <- ghglm(fixed = Faults ~ 1,
random = ~ 1|Length,
family = quasipoisson(link = "log"),
rand.family = gaussian(link = "identity"),
method = "EQL",
data = X)

m <- m12
mm <- "m12"
(ms <- summary(m))
(m$vmratio)
(SeRem <- mean(m$SeRe))
(SeRes <- sd(m$SeRe))
(SeRev <- var(m$SeRe))


#####     rats model
  X <- Rat
 Ex <- "Rat"
ver <- "PN"
m13 <- ghglm(fixed = Y ~ W + R + Drug,
# random = ~ 1|(W + W*W),
random = ~ 1|W,
family = poisson(link = "log"),
rand.family = gaussian(link = "identity"),
method = "EQL",
data = X)

m <- m13
mm <- "m13"
(ms <- summary(m))
(m$vmratio)
(SeRem <- mean(m$SeRe))
(SeRes <- sd(m$SeRe))
(SeRev <- var(m$SeRe))


#####     ssn model
  X <- SSN
 Ex <- "SSN"
ver <- "PN"   # PoiNorm
m10 <- ghglm(fixed = w ~ see + r + as.factor(year) + as.factor(mon),
random = ~ 1|obs,
family = quasipoisson(link = "log"),
rand.family = gaussian(link = "identity"),
method = "EQL",
data = X)
```

```
m <- m10
mm <- "m10"
(ms <- summary(m))
(m$vmratio)
(SeRem <- mean(m$SeRe))
(SeRes <- sd(m$SeRe))
(SeRev <- var(m$SeRe))


######################################################################
#####     Save model     #############################################
######################################################################

#setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
part <- "ghglm"
WD <- getwd()
outfile <- paste(WD, "/model", Ex, ver, mm, part, ".RData", sep="")
save(m, file=outfile, ascii=FALSE)


#####     Load model     #############################################

# Model designations
# m00 - m09: normal-normal
# m10 - m19: Poisson-normal
# m20 - m29: Poisson-gamma
# m30 - m39: Poisson-quasi

mm <- "m10"  # SSN
mm <- "m12"  # Fab
mm <- "m13"  # Rat

WD <- getwd()
outfile <- paste(WD, "/model", Ex, ver, mm, part, ".RData", sep="")
load(outfile)
ms <- summary(m)

#####     End load model     ########################################


######################################################################
     part <- "ghglm"     #############################################
######################################################################

#####     Poisson-gamma     ##########################################

#####     fabric model
  X <- Fab
```

```
 Ex <- "Fab"
ver <- "PG"
m22 <- ghglm(fixed = Faults ~ 1,
random = ~ 1|Length,
family = quasipoisson(link = "log"),
rand.family = Gamma(link = "inverse"),
method = "EQL",
data = X)

m <- m22
mm <- "m22"
(ms <- summary(m))
(m$vmratio)
(SeRem <- mean(m$SeRe))
(SeRes <- sd(m$SeRe))
(SeRev <- var(m$SeRe))


#####     rats model
  X <- Rat
 Ex <- "Rat"
ver <- "PG"
m23 <- ghglm(fixed = Y ~ W + R + Drug,
random = ~ 1|(W + W*W),
family = quasipoisson(link = "log"),
rand.family = Gamma(link = "inverse"),
method = "EQL",
data = X)

m <- m23
mm <- "m23"
(ms <- summary(m))
(m$vmratio)
(SeRem <- mean(m$SeRe))
(SeRes <- sd(m$SeRe))
(SeRev <- var(m$SeRe))


#####     ssn model
  X <- SSN
 Ex <- "SSN"
ver <- "PG"   # PoiGamma
m20 <- ghglm(fixed = w ~ see + r + as.factor(year) + as.factor(mon),
# disp = ~ as.factor(year) + as.factor(mon),
random = ~ 1|obs,
family = quasipoisson(link = "log"),
rand.family = Gamma(link = "inverse"),
method = "EQL",
data = X)
```

```
m <- m20
mm <- "m20"
(ms <- summary(m))
(m$vmratio)
(SeRem <- mean(m$SeRe))
(SeRes <- sd(m$SeRe))
(SeRev <- var(m$SeRe))


######################################################################
#####     Save model     ##########################################
######################################################################

#setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
part <- "ghglm"
WD <- getwd()
outfile <- paste(WD, "/model", Ex, ver, mm, part, ".RData", sep="")
save(m, file=outfile, ascii=FALSE)


#####     Load model     #############################################

# Model designations
# m00 - m09: normal-normal
# m10 - m19: Poisson-normal
# m20 - m29: Poisson-gamma
# m30 - m39: Poisson-quasi

mm <- "m20"  # SSN
mm <- "m22"  # Fab
mm <- "m23"  # Rat

WD <- getwd()
outfile <- paste(WD, "/model", Ex, ver, mm, part, ".RData", sep="")
load(outfile)
ms <- summary(m)
print(ms,corr=FALSE)

#####     End load model     ########################################


######################################################################
#####     calc lambda     ############################################
######################################################################

devid <- (length(m$dev) - max(m$RandC) + 1):
         (length(m$dev) - max(m$RandC) + m$RandC[1])
dr <- m$dev[devid]
```

```
bc <- bct(dr~1,plotit=F)
(lvec <- c(bc$cil95,bc$lambda,bc$ciu95))


######################################################################
#####     Calc psi with magic    #####################################
#####     closed form            #####################################
######################################################################

Solution <- "Closed"
n.rand <- (length(m$dev) - max(m$RandC) + 1):
          (length(m$dev) - max(m$RandC) + m$RandC[1])
A <- data.frame(m$xi, m$dev[n.rand])
names(A) <- c("xi","dR")

lambda <- lvec[2]
lnvxi <- log(var(A$xi))
lnxi <- log(mean(A$xi))
ln1l <- log(1-lambda*lambda)
lndR <- log(mean(A$dR))
(psi <- (lnvxi - ln1l - lndR)/lnxi)

psifun <- function(A,i) {
lnvxi <- log(var(A$xi[i]))
lnxi <- log(mean(A$xi[i]))
ln1l <- log(1-lambda*lambda)
lndR <- log(mean(A$dR[i]))
psifun <- (lnvxi - ln1l - lndR)/lnxi
}
(psi.i <- boot(A, psifun, R=1000))
(psi.ci <- boot.ci(psi.i, conf = 0.95, type = "norm"))
cis <- c(psi.ci$normal[2],psi,psi.ci$normal[3])
names(cis) <- c("CI95L","hat{psi}","CI95H")
cis

#####     one-off only plot of psi by lambda     #####################

dset <- "Fabric"
dset <- "Rats"
dset <- "SSN"

lambda <- seq(-.999,.999,.001)
lnvxi <- log(var(A$xi))
lnxi <- log(mean(A$xi))
ln1l <- log(1-lambda*lambda)
lndR <- log(mean(A$dR))
psi <- (lnvxi - ln1l - lndR)/lnxi
max(psi)
```

```
(loc <- paste0("Plots/XYpsiBYlambda", dset, ".png"))
quartz(w=5, h=5)
par(lwd=2,family="serif")
plot(lambda, psi, col="black", type="l", xlim=c(-1,1), ylim=c(0,1),
     xlab=expression(lambda), ylab=expression(psi))
title(expression(paste(psi, " vs. ", lambda)), line=3)
title("Closed Form Solution", line=2, cex.main=0.9)
title(paste0(dset, " Data"), line=1, cex.main=0.9)
par(lwd=1,family="sans")
grid(col="black")
quartz.save(loc, type="png")


######################################################################
#####     Calc psi with magic      ###################################
#####     nonlinear regression     ###################################
######################################################################

Solution <- "Nonlinear"
n.rand <- (length(m$dev) - max(m$RandC) + 1):
          (length(m$dev) - max(m$RandC) + m$RandC[1])
A <- data.frame(m$xi, m$dev[n.rand])
names(A) <- c("xi","dR")
summary(A)
#preview(dR ~ xi^(psi), data=A, start=list(psi=1))

xi <- as.numeric(scale(m$xi, center=T, scale=T)) + 4
      #1 + abs(min(c(m$dev[n.rand],m$xi))) + m$xi
varxi <- var(xi)
lambda <- lvec[2]
dR <- as.numeric(scale(m$dev[n.rand] * ((1 - lambda*lambda)/varxi),
      center=T, scale=T)) + 4 #1 + abs(min(c(m$dev[n.rand],m$xi)))
      + m$dev[n.rand]
A <- data.frame(dR,xi)

mo <- nls(dR ~ a*xi^(psi), data=A,
      start=list(psi=lambda,a=1), trace=T)
(mos <- summary(mo))
#confint(mo,"psi")
psi <- mos$parameters[1]
cis <- c(confint(mo,"psi")[1],psi,confint(mo,"psi")[2])
names(cis) <- c("CI95L","hat{psi}","CI95H")
cis


######################################################################
     part <- "ghglm"     #############################################
######################################################################
```

```
#####     Poisson-quasi     ##########################################

#####     fabric model
  X <- Fab
 Ex <- "Fab"
ver <- "PQ"
mvr <- quasipf(link="log", variance="mu^pwr", pwr=psi)
m32 <- ghglm(fixed = Faults ~ 1,
random = ~ 1|length,
family = quasipoisson(link = "log"),
rand.family = mvr,
method = "EQL",
data = X)

m <- m32

m32Closed <- m32
m32Nonlinear <- m32

mm <- paste0("m32",Solution)
(ms <- summary(m))
(m$vmratio)
(SeRem <- mean(m$SeRe))
(SeRes <- sd(m$SeRe))
(SeRev <- var(m$SeRe))


#####     rats model
  X <- Rat
 Ex <- "Rat"
ver <- "PQ"
mvr <- quasipf(link="log", variance="mu^pwr", pwr=psi)
m33 <- ghglm(fixed = Y ~ W + R + Drug,
random = ~ 1|(W + W*W),
family = quasipoisson(link = "log"),
rand.family = mvr,
method = "EQL",
data = X)

m <- m33

m33Closed <- m33
m33Nonlinear <- m33

mm <- paste0("m33",Solution)
(ms <- summary(m))
(m$vmratio)
(SeRem <- mean(m$SeRe))
(SeRes <- sd(m$SeRe))
```

```
(SeRev <- var(m$SeRe))


#####     ssn model
  X <- SSN
 Ex <- "SSN"
ver <- "PQ"   # PoiQuasi
mvr <- quasipf(link="log", variance="mu^pwr", pwr=psi)
#mvr <- quasi(link=power(psi))
#mvr$variance function(mu) mu^psi
#mvr$varfun <- paste0("mu^",as.character(psi))
m30 <- ghglm(fixed = w~ see + r + as.factor(year) + as.factor(mon),
# disp = ~ as.factor(year) + as.factor(mon),
random = ~ 1|obs,
family = quasipoisson(link = "log"),
rand.family = mvr,
method = "EQL",
data = X)

m <- m30

m30Closed <- m30
m30Nonlinear <- m30

mm <- paste0("m30",Solution)
(ms <- summary(m))
(m$vmratio)
(SeRem <- mean(m$SeRe))
(SeRes <- sd(m$SeRe))
(SeRev <- var(m$SeRe))


######################################################################
#####     Save model     #############################################
######################################################################

#setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
part <- "ghglm"
WD <- getwd()
outfile <- paste(WD, "/model", Ex, ver, mm, part, ".RData", sep="")
save(m, file=outfile, ascii=FALSE)


#####     Load model     #############################################

# Model designations
# m00 - m09: normal-normal
# m10 - m19: Poisson-normal
# m20 - m29: Poisson-gamma
```

```
# m30 - m39: Poisson-quasi

mm <- "m30Closed"  # SSN
mm <- "m30Nonlinear"  # SSN
mm <- "m32Closed"  # Fab
mm <- "m32Nonlinear"
mm <- "m33Closed"  # Rat
mm <- "m33Nonlinear"

WD <- getwd()
outfile <- paste(WD, "/model", Ex, ver, mm, part, ".RData", sep="")
load(outfile)
ms <- summary(m)
print(ms,corr=FALSE)

#####     End load model     ########################################


######################################################################
#####     Construct data frame of model parameters     ###############
######################################################################

rm(cmp)
Ex <- "Rat"
cmd <- rbind(c(m13$vm,m23$vm,m33$vm),
c(m13$varFix,m23$varFix,m33$varFix),
c(m13$varRanef,m23$varRanef,m33$varRanef),
c(mean(m13$SeRe),mean(m23$SeRe),mean(m33$SeRe)),
c(sd(m13$SeRe),sd(m23$SeRe),sd(m33$SeRe)),
c(var(m13$SeRe),var(m23$SeRe),var(m33$SeRe)),
c(NA,lvec[1],NA),
c(NA,lvec[2],NA),
c(NA,lvec[3],NA),
c(NA,cis[1], NA),
c(NA,cis[2], NA),
c(NA,cis[3], NA)
)
cmd <- data.frame(rep(Ex, nrow(cmd)), c("vmr", "phi", "zeta",
       "ReSeAvg", "ReSeSe", "ReSeV", "lambdaL", "lambda", "lambdaU",
       "psiL", "psi", "psiU"), cmd)
cmd$Solution <- Solution
colnames(cmd) <- c("DataSet", "Parameter", "Normal", "Gamma",
                   "Quasi", "Solution")
(cmp <- cmd)
(cmp <- data.frame(rbind(cmp,cmd)))

Ex <- "SSN"
cmd <- rbind(c(m10$vm,m20$vm,m30$vm),
c(m10$varFix,m20$varFix,m30$varFix),
```

```
c(m10$varRanef,m20$varRanef,m30$varRanef),
c(mean(m10$SeRe),mean(m20$SeRe),mean(m30$SeRe)),
c(sd(m10$SeRe),sd(m20$SeRe),sd(m30$SeRe)),
c(var(m10$SeRe),var(m20$SeRe),var(m30$SeRe)),
c(NA,lvec[1],NA),
c(NA,lvec[2],NA),
c(NA,lvec[3],NA),
c(NA,cis[1], NA),
c(NA,cis[2], NA),
c(NA,cis[3], NA)
)
cmd <- data.frame(rep(Ex,nrow(cmd)), c("vmr", "phi", "zeta",
       "ReSeAvg", "ReSeSe", "ReSeV", "lambdaL", "lambda", "lambdaU",
       "psiL", "psi", "psiU"), cmd)
cmd$Solution <- Solution
colnames(cmd) <- c("DataSet", "Parameter", "Normal", "Gamma",
                  "Quasi", "Solution")
(cmp <- data.frame(rbind(cmp,cmd)))

Ex <- "Fab"
cmd <- rbind(c(m12$vm,m22$vm,m32$vm),
c(m10$varFix,m20$varFix,m30$varFix),
c(m10$varRanef,m20$varRanef,m30$varRanef),
c(mean(m10$SeRe),mean(m20$SeRe),mean(m30$SeRe)),
c(sd(m10$SeRe),sd(m20$SeRe),sd(m30$SeRe)),
c(var(m10$SeRe),var(m20$SeRe),var(m30$SeRe)),
c(NA,lvec[1],NA),
c(NA,lvec[2],NA),
c(NA,lvec[3],NA),
c(NA,cis[1], NA),
c(NA,cis[2], NA),
c(NA,cis[3], NA)
)
cmd <- data.frame(rep(Ex,nrow(cmd)), c("vmr", "phi", "zeta",
      "ReSeAvg", "ReSeSe", "ReSeV", "lambdaL", "lambda", "lambdaU",
      "psiL", "psi", "psiU"), cmd)
cmd$Solution <- Solution
colnames(cmd) <- c("DataSet", "Parameter", "Normal", "Gamma",
                   "Quasi", "Solution")
(cmp <- data.frame(rbind(cmp,cmd)))

PG <- cmp
rm(cmp,cmd)


######################################################################
#####     Diagnostics          ######################################
#####     uses model in "m"    ######################################
          part <- "diags"      ######################################
```

```
####################################################################

(loc <- paste("Plots/", Ex, part, "diag", mm, ".png", sep=""))
quartz(w=6,h=6)
plot(m)
quartz.save(loc, type="png")

#####     Fab     ###################################################
dset <- "Fabric"
#m <- m12
reres <- (length(m$dev) - max(m$RandC) + 1):
          (length(m$dev) - max(m$RandC) + m$RandC[1])

drq1 <- m12$dev[reres]
drq2 <- m22$dev[reres]
drq3 <- m32Closed$dev[reres]
drq4 <- m32Nonlinear$dev[reres]

#####     Construct data frame of S-W stats     #####################
rm(sw)
dist <- "Normal"
sh <- data.frame(dset, dist, shapiro.test(drq1)$statistic[[1]],
       shapiro.test(drq1)$p.value[[1]], " ")
names(sh) <- c("Data","Dist","W","p.value","Solution")
sw <- sh
dist <- "Gamma"
sh <- data.frame(dset, dist, shapiro.test(drq2)$statistic[[1]],
      shapiro.test(drq2)$p.value[[1]], " ")
names(sh) <- c("Data","Dist","W","p.value","Solution")
sw <- rbind(sw,sh)
dist <- "Quasi"
sh <- data.frame(dset, dist, shapiro.test(drq3)$statistic[[1]],
      shapiro.test(drq3)$p.value[[1]], "Closed")
names(sh) <- c("Data","Dist","W","p.value","Solution")
sw <- rbind(sw,sh)
dist <- "Quasi"
sh <- data.frame(dset, dist, shapiro.test(drq4)$statistic[[1]],
       shapiro.test(drq4)$p.value[[1]], "Nonlinear")
names(sh) <- c("Data","Dist","W","p.value","Solution")
(sw <- rbind(sw,sh))

#####     Normal Q-Q plots     ######################################
(loc <- paste0("Plots/", part, "QQ", Ex, ".png"))
quartz(w=7.5, h=8)
par(mfrow=c(2,2), lwd=2, family="serif")
qqnorm(drq1, main=paste0(dset, " PN RE Dev Norm Q-Q ", " "))
qqline(drq1,lwd=2,col="red")
par(lwd=1)
grid(col="black")
```

```
par(lwd=2)
qqnorm(drq2, main=paste0(dset, " PG RE Dev Gamma Q-Q ", " "))
qqline(drq2,lwd=2,col="red")
par(lwd=1)
grid(col="black")
par(lwd=2)
qqnorm(drq3, main=paste0(dset, " PQ RE Dev Quasi Q-Q ", Solution))
qqline(drq3,lwd=2,col="red")
par(lwd=1, family="sans")
grid(col="black")
par(lwd=2)
qqnorm(drq4, main=paste0(dset, " PQ RE Dev Quasi Q-Q ", Solution))
qqline(drq4,lwd=2,col="red")
par(lwd=1, family="sans")
grid(col="black")
quartz.save(loc, type="png")


#####     Rat     ####################################################
#devid <- 1:(length(m$dev) - max(m$RandC))
dset <- "Rats"
#m <- m13
reres <- (length(m$dev) - max(m$RandC) + 1):
          (length(m$dev) - max(m$RandC) + m$RandC[1])

drq1 <- m13$dev[reres]
drq2 <- m23$dev[reres]
drq3 <- m33Closed$dev[reres]
drq4 <- m33Nonlinear$dev[reres]

#####     Construct data frame of S-W stats     ######################
rm(sw)
dist <- "Normal"
sh <- data.frame(dset, dist, shapiro.test(drq1)$statistic[[1]],
           shapiro.test(drq1)$p.value[[1]], " ")
names(sh) <- c("Data","Dist","W","p.value","Solution")
sw <- sh
sw <- rbind(sw,sh)
dist <- "Gamma"
sh <- data.frame(dset, dist, shapiro.test(drq2)$statistic[[1]],
           shapiro.test(drq2)$p.value[[1]], " ")
names(sh) <- c("Data","Dist","W","p.value","Solution")
sw <- rbind(sw,sh)
dist <- "Quasi"
sh <- data.frame(dset, dist, shapiro.test(drq3)$statistic[[1]],
           shapiro.test(drq3)$p.value[[1]], "Closed")
names(sh) <- c("Data","Dist","W","p.value","Solution")
sw <- rbind(sw,sh)
dist <- "Quasi"
```

```
sh <- data.frame(dset, dist, shapiro.test(drq4)$statistic[[1]],
           shapiro.test(drq4)$p.value[[1]], "Nonlinear")
names(sh) <- c("Data","Dist","W","p.value","Solution")
(sw <- rbind(sw,sh))


#####     Normal Q-Q plots     #####################################
(loc <- paste0("Plots/", part, "QQ", Ex, ".png"))
quartz(w=7.5, h=8)
par(mfrow=c(2,2), lwd=2, family="serif")
qqnorm(drq1, main=paste0(dset, " PN RE Dev Norm Q-Q ", " "))
qqline(drq1,lwd=2,col="red")
par(lwd=1)
grid(col="black")
par(lwd=2)
qqnorm(drq2, main=paste0(dset, " PG RE Dev Gamma Q-Q ", " "))
qqline(drq2,lwd=2,col="red")
par(lwd=1)
grid(col="black")
par(lwd=2)
qqnorm(drq3, main=paste0(dset, " PQ RE Dev Quasi Q-Q ", Solution))
qqline(drq3,lwd=2,col="red")
par(lwd=1, family="sans")
grid(col="black")
par(lwd=2)
qqnorm(drq4, main=paste0(dset, " PQ RE Dev Quasi Q-Q ", Solution))
qqline(drq4,lwd=2,col="red")
par(lwd=1, family="sans")
grid(col="black")
quartz.save(loc, type="png")


#####     SSN     ##################################################
#devid <- 1:(length(m$dev) - max(m$RandC))
dset <- "SSN"
#m <- m10
reres <- (length(m$dev) - max(m$RandC) + 1):
         (length(m$dev) - max(m$RandC) + m$RandC[1])

drq1 <- m10$dev[reres]
drq2 <- m20$dev[reres]
drq3 <- m30Closed$dev[reres]
drq4 <- m30Nonlinear$dev[reres]

#####     Construct data frame of S-W stats     ######################
dist <- "Normal"
sh <- data.frame(dset, dist, shapiro.test(drq1)$statistic[[1]],
           shapiro.test(drq1)$p.value[[1]], " ")
names(sh) <- c("Data","Dist","W","p.value","Solution")
```

```
sw <- rbind(sw,sh)
dist <- "Gamma"
sh <- data.frame(dset, dist, shapiro.test(drq2)$statistic[[1]],
           shapiro.test(drq2)$p.value[[1]], " ")
names(sh) <- c("Data","Dist","W","p.value","Solution")
sw <- rbind(sw,sh)
dist <- "Quasi"
sh <- data.frame(dset, dist, shapiro.test(drq3)$statistic[[1]],
           shapiro.test(drq3)$p.value[[1]], "Closed")
names(sh) <- c("Data","Dist","W","p.value","Solution")
sw <- rbind(sw,sh)
dist <- "Quasi"
sh <- data.frame(dset, dist, shapiro.test(drq4)$statistic[[1]],
           shapiro.test(drq4)$p.value[[1]], "Nonlinear")
names(sh) <- c("Data","Dist","W","p.value","Solution")
(sw <- rbind(sw,sh))
rm(sh)


######################################################################
#####     Normal Q-Q plots     #######################################
(loc <- paste0("Plots/", part, "QQ", Ex, ".png"))
quartz(w=7.5, h=8)
par(mfrow=c(2,2), lwd=2, family="serif")
qqnorm(drq1, main=paste0(dset, " PN RE Dev Norm Q-Q ", " "))
qqline(drq1,lwd=2,col="red")
par(lwd=1)
grid(col="black")
par(lwd=2)
qqnorm(drq2, main=paste0(dset, " PG RE Dev Gamma Q-Q ", " "))
qqline(drq2,lwd=2,col="red")
par(lwd=1)
grid(col="black")
par(lwd=2)
qqnorm(drq3, main=paste0(dset, " PQ RE Dev Quasi Q-Q ", Solution))
qqline(drq3,lwd=2,col="red")
par(lwd=1, family="sans")
grid(col="black")
par(lwd=2)
qqnorm(drq4, main=paste0(dset, " PQ RE Dev Quasi Q-Q ", Solution))
qqline(drq4,lwd=2,col="red")
par(lwd=1, family="sans")
grid(col="black")
quartz.save(loc, type="png")


#####     Save data set     ##########################################

setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
```

```
WD <- getwd()
outfile <- paste0(WD, "/", "sw", ".RData")
save(sw, file=outfile, ascii=FALSE)

#####     Load data set     #########################################

setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()
outfile <- paste(WD, "/", "sw", ".RData", sep="")
load(outfile)

#####     End load model     ########################################

#####     Save data set     ########################################

setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()
outfile <- paste0(WD, "/", "PG", ".RData")
save(PG, file=outfile, ascii=FALSE)

#####     Load data set     #########################################

setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()
outfile <- paste(WD, "/", "PG", ".RData", sep="")
load(outfile)

#####     End load model     ########################################


######################################################################
#####     bar plots, cf     #########################################
######################################################################

# Model designations
# m00 - m09: normal-normal
# m10 - m19: Poisson-normal
# m20 - m29: Poisson-gamma
# m30 - m39: Poisson-quasi

#####     fabric data     ###########################################

dset <- "Fab"
 (ds <- subset(PG, DataSet == dset))

dist <- "Gamma"
  pd <- "PG"
   m <- m12
```

```
#####     VMR % change of quasi over normal or gamma     #############

pltt <- "VMR"
rows <- c(1,13)
  yl <- c(0,4)
  os <- 0.18
 sub <- "Random Effects Distribution \n \n
         Variance-to-mean ratio = 1 is desired."
ylab="Variance-to-Mean Ratio"

#####     Construct data frame model parameter % difs     ############
rm(deltas)
deltas <- data.frame(dset, pltt, "Q-N", 100*(ds[rows[1],
               "Quasi"] - ds[rows[1], "Normal"])/ds[rows[1],
               "Normal"], "Closed")
names(deltas) <- c("Data","Parameter","Delta","Percent","Solution")
dg <- data.frame(dset, pltt, "Q-G", 100*(ds[rows[1],
         "Quasi"]-ds[rows[1], "Gamma"])/ds[rows[1],
         "Gamma"], "Closed")
names(dg) <- c("Data","Parameter","Delta","Percent","Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-N", 100*(ds[rows[2],
         "Quasi"]-ds[rows[2], "Normal"])/ds[rows[2],
         "Normal"], "Nonlinear")
names(dg) <- c("Data","Parameter","Delta","Percent","Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-G", 100*(ds[rows[2],
         "Quasi"]-ds[rows[2], "Gamma"])/ds[rows[2],
         "Gamma"], "Nonlinear")
names(dg) <- c("Data","Parameter","Delta","Percent","Solution")
deltas <- rbind(deltas,dg)

#####     Avg RE SE % change of quasi over normal or gamma     #######

pltt <- "SEavg"
rows <- c(4,16)
  yl <- c(0,0.35)
  os <- 0.026
 sub <- "Random Effects Distribution \n \n
         Smallest average standard error is best."
ylab="Average Random Effects se"

#####     Construct data frame model parameter % difs     ############
dg <- data.frame(dset, pltt, "Q-N", 100*(ds[rows[1],
           "Quasi"]-ds[rows[1],"Normal"])/[rows[1],
           "Normal"], "Closed")
names(dg) <- c("Data","Parameter","Delta","Percent","Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-G", 100*(ds[rows[1],
```

```
           "Quasi"]-ds[rows[1], "Gamma"])/ds[rows[1],
           "Gamma"], "Closed")
names(dg) <- c("Data","Parameter","Delta","Percent","Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-N", 100*(ds[rows[2],
           "Quasi"]-ds[rows[2],"Normal"])/ds[rows[2],
           "Normal"], "Nonlinear")
names(dg) <- c("Data","Parameter","Delta","Percent","Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-G", 100*(ds[rows[2],
           "Quasi"]-ds[rows[2], "Gamma"])/ds[rows[2],
           "Gamma"], "Nonlinear")
names(dg) <- c("Data","Parameter","Delta","Percent","Solution")
deltas <- rbind(deltas,dg)


ci <- data.frame(dset,"lambda",lvec[1],lvec[2],lvec[3]," "," ")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist","Solution")
CIs <- rbind(CIs,ci)
ci <- data.frame(dset, "psi", CIs[[1]], CIs[[2]], CIs[[3]],
                " ", "Closed")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist","Solution")
CIs <- rbind(CIs,ci)
ci <- data.frame(dset, "psi", CIs[[1]], CIs[[2]], CIs[[3]],
                 " ", "Nonlinear")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist", "Solution")
(CIs <- rbind(CIs,ci))


rm(CIs)
CIs <- data.frame(dset, "lambda", ds[7,"Gamma"], ds[8,"Gamma"],
            ds[9,"Gamma"]," "," ")
names(CIs) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
                "Dist", "Solution")
ci <- data.frame(dset, "psi", ds[10,"Gamma"], ds[11,"Gamma"],
            ds[12,"Gamma"], " ", "Closed")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist", "Solution")
CIs <- rbind(CIs,ci)
ci <- data.frame(dset, "psi", ds[22,"Gamma"], ds[23,"Gamma"],
            ds[24,"Gamma"], " ", "Nonlinear")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist", "Solution")
CIs <- rbind(CIs,ci)
se <- qnorm(0.975)*ds[rows[1],"Normal"]/sqrt(m$RandC)
ci <- data.frame(dset, pltt, ds[rows[1],"Normal"] - se,
```

```
            ds[rows[1],"Normal"], ds[rows[1],"Normal"] + se,
            "Normal", " ")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist","Solution")
CIs <- rbind(CIs,ci)
se <- qnorm(0.975)*ds[rows[2],"Normal"]/sqrt(m$RandC)
ci <- data.frame(dset, pltt,ds[rows[1],"Gamma"] - se,
           ds[rows[1],"Gamma"], ds[rows[1],"Gamma"] + se,
           "Gamma", " ")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist", "Solution")
CIs <- rbind(CIs,ci)
ci <- data.frame(dset, pltt, ds[rows[1],"Quasi"] - se,
           ds[rows[1],"Quasi"], ds[rows[1],"Quasi"] + se,
           "Quasi", "Closed")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
              "Dist","Solution")
CIs <- rbind(CIs,ci)
ci <- data.frame(dset, pltt, ds[rows[2],"Quasi"] - se,
            ds[rows[2],"Quasi"], ds[rows[2],"Quasi"] + se,
            "Quasi", "Nonlinear")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist","Solution")
CIs <- rbind(CIs,ci)


#####     Bar plots     ############################################
(loc <- paste("Plots/Bar", pltt, pd, dset, ".png", sep=""))
mds <- as.matrix(ds[rows,3:5])
mds[4,5] <- 0.5
mds[16,5] <- 0.5
quartz(w=6, h=6, bg="white")
par(lwd=2,family="serif")
bp <- barplot(as.matrix(ds[rows,3:5]), beside=T, density=10,
       angle=c(45,-45), col=c("black","black"), ylim=yl, sub=sub,
       ylab=ylab)
text(bp, as.matrix(ds[rows,3:5])+os,
      as.character(round(as.matrix(ds[rows,3:5]),4)), pos=1, cex=0.8)
ifelse(pltt == "VMR",
(title("Comparison of Variance-to-Mean Ratios", line=3)),
(title("Comparison of Random Effects Average Standard Errors",
       line=3))
)
title(expression(paste("Fabric Data Closed Form and Nonlinear
      Solutions for ", psi)), line=2)
ifelse(dist == "Normal",
(title(expression(paste(lambda,
    " From Assumed Normal Random Effects")), line=1)),
(title(expression(paste(lambda,
```

```
    " From Assumed Gamma Random Effects")), line=1))
)
legend("topleft", c("Closed","Nonlinear"), density=10,
     angle=c(45,-45), col=c("black","black"), bty="n")
par(lwd=1,family="sans")
grid(nx=NA, ny=NULL, col="black")
quartz.save(loc, type="png")


#####     rats data     #############################################

dset <- "Rat"
 (ds <- subset(PG, DataSet == dset))

dist <- "Gamma"
  pd <- "PG"
   m <- m13

#####     VMR % change of quasi over normal or gamma     #############

pltt <- "VMR"
rows <- c(1,13)
  yl <- c(0,4)
  os <- 0.3
 sub <- "Random Effects Distribution \n \n
         Variance-to-mean ratio = 1 is desired."
ylab="Variance-to-Mean Ratio"

#####     Construct data frame model parameter % difs     #############
rm(deltas)
dg <- data.frame(dset, pltt, "Q-N", 100*(ds[rows[1],"Quasi"]
       - ds[rows[1],"Normal"])/ds[rows[1], "Normal"], "Closed")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
(deltas <- dg)
dg <- data.frame(dset,pltt,"Q-G",100*(ds[rows[1],"Quasi"]
       - ds[rows[1],"Gamma"])/ds[rows[1],"Gamma"],"Closed")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-N", 100*(ds[rows[2],"Quasi"]
       - ds[rows[2],"Normal"])/ds[rows[2],"Normal"],"Nonlinear")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-G", 100*(ds[rows[2],"Quasi"]
       - ds[rows[2],"Gamma"])/ds[rows[2],"Gamma"], "Nonlinear")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
(deltas <- rbind(deltas,dg))

#####     Avg RE SE % change of quasi over normal or gamma     #######
```

```
pltt <- "SEavg"
rows <- c(4,16)
  yl <- c(0,0.12)
  os <- 0.008
 sub <- "Random Effects Distribution \n \n
         Smallest average standard error is best."
ylab="Average Random Effects se"

#####     Construct data frame model parameter % difs     ############
dg <- data.frame(dset, pltt, "Q-N", 100*(ds[rows[1],"Quasi"]
       -  ds[rows[1],"Normal"])/ds[rows[1],"Normal"], "Closed")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-G", 100*(ds[rows[1],"Quasi"]
       - ds[rows[1],"Gamma"])/ds[rows[1],"Gamma"], "Closed")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-N", 100*(ds[rows[2],"Quasi"]
       - ds[rows[2],"Normal"])/ds[rows[2],"Normal"], "Nonlinear")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-G", 100*(ds[rows[2],"Quasi"]
       - ds[rows[2],"Gamma"])/ds[rows[2],"Gamma"], "Nonlinear")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
(deltas <- rbind(deltas,dg))


rm(CIs)
ci <- data.frame(dset, "lambda", ds[7,"Gamma"], ds[8,"Gamma"],
           ds[9,"Gamma"], " ", " ")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist","Solution")
(CIs <- ci)
#CIs <- rbind(CIs,ci)
ci <- data.frame(dset, "psi", ds[10,"Gamma"], ds[11,"Gamma"],
           ds[12,"Gamma"], " ", "Closed")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist","Solution")
CIs <- rbind(CIs,ci)
ci <- data.frame(dset, "psi", ds[22,"Gamma"], ds[23,"Gamma"],
            ds[24,"Gamma"], " ", "Nonlinear")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist", "Solution")
CIs <- rbind(CIs,ci)
se <- qnorm(0.975)*ds[rows[1],"Normal"]/sqrt(m$RandC)
ci <- data.frame(dset, pltt, ds[rows[1],"Normal"] - se,
           ds[rows[1],"Normal"], ds[rows[1],"Normal"] + se,
           "Normal", " ")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
```

```
                "Dist","Solution")
CIs <- rbind(CIs,ci)
se <- qnorm(0.975)*ds[rows[2],"Normal"]/sqrt(m$RandC)
ci <- data.frame(dset, pltt, ds[rows[1],"Gamma"] - se,
           ds[rows[1],"Gamma"], ds[rows[1],"Gamma"] + se,
           "Gamma", " ")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
              "Dist", "Solution")
CIs <- rbind(CIs,ci)
ci <- data.frame(dset, pltt, ds[rows[1],"Quasi"] - se,
           ds[rows[1],"Quasi"], ds[rows[1],"Quasi"] + se,
           "Quasi", "Closed")
names(ci) <- c("Data","Parameter","CI95l","Estimate","CI95h",
              "Dist","Solution")
CIs <- rbind(CIs,ci)
ci <- data.frame(dset, pltt, ds[rows[2],"Quasi"] - se,
           ds[rows[2],"Quasi"], ds[rows[2],"Quasi"] + se, "Quasi",
           "Nonlinear")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist", "Solution")
(CIs <- rbind(CIs,ci))


#####     Bar plots     ##############################################
(loc <- paste("Plots/Bar", pltt, pd, dset, ".png", sep=""))
quartz(w=6, h=6, bg="white")
par(lwd=2,family="serif")
bp <- barplot(as.matrix(ds[rows,3:5]), beside=T, density=10,
     angle=c(45,-45), col=c("black","black"), ylim=yl, sub=sub,
     ylab=ylab)
text(bp, as.matrix(ds[rows,3:5])+os,
       as.character(round(as.matrix(ds[rows,3:5]),4)), pos=1,
       cex=0.8)
ifelse(pltt == "VMR",
(title("Comparison of Variance-to-Mean Ratios", line=3)),
(title("Comparison of Random Effects Average Standard Errors"
              line=3))
)
title(expression(
   paste("Rats Data Closed Form and Nonlinear Solutions for ",
       psi)),
   line=2)
ifelse(dist == "Normal",
(title(expression(paste(lambda,
    " From Assumed Normal Random Effects")), line=1)),
(title(expression(paste(lambda,
    " From Assumed Gamma Random Effects")), line=1))
)
legend("topleft", c("Closed","Nonlinear"), density=10,
```

```
    angle=c(45,-45), col=c("black","black"), bty="n")
par(lwd=1,family="sans")
grid(nx=NA, ny=NULL, col="black")
quartz.save(loc, type="png")


#####     SSN data     ##############################################

dset <- "SSN"
 (ds <- subset(PG, DataSet == dset))

dist <- "Gamma"
  pd <- "PG"
   m <- m10

#####     VMR % change of quasi over normal or gamma     #############

pltt <- "VMR"
rows <- c(1,13)
  yl <- c(0,25)
  os <- 1.7
 sub <- "Random Effects Distribution \n \n
         Variance-to-mean ratio = 1 is desired."
ylab="Variance-to-Mean Ratio"

#####     Construct data frame model parameter % difs     ############
dg <- data.frame(dset, pltt, "Q-N", 100*(ds[rows[1],"Quasi"]
       - ds[rows[1],"Normal"])/ds[rows[1],"Normal"], "Closed")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-G", 100*(ds[rows[1],"Quasi"]
        - ds[rows[1],"Gamma"])/ds[rows[1],"Gamma"], "Closed")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-N", 100*(ds[rows[2],"Quasi"]
       - ds[rows[2],"Normal"])/ds[rows[2],"Normal"], "Nonlinear")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-G", 100*(ds[rows[2],"Quasi"]
       - ds[rows[2],"Gamma"])/ds[rows[2],"Gamma"], "Nonlinear")
names(dg) <- c("Data", "Parameter", "Delta", "Percent",  "Solution")
(deltas <- rbind(deltas,dg))

#####     Avg RE SE % change of quasi over normal or gamma     #######

pltt <- "SEavg"
rows <- c(4,16)
  yl <- c(0,0.25)
  os <- 0.015
```

```
 sub <- "Random Effects Distribution \n \n
         Smallest average standard error is best."
ylab="Average Random Effects se"

#####     Construct data frame model parameter % difs     ############
dg <- data.frame(dset, pltt, "Q-N", 100*(ds[rows[1],"Quasi"]
       - ds[rows[1],"Normal"])/ds[rows[1],"Normal"], "Closed")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-G", 100*(ds[rows[1],"Quasi"]
       - ds[rows[1],"Gamma"])/ds[rows[1],"Gamma"], "Closed")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-N", 100*(ds[rows[2],"Quasi"]
       - ds[rows[2],"Normal"])/ds[rows[2],"Normal"], "Nonlinear")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
deltas <- rbind(deltas,dg)
dg <- data.frame(dset, pltt, "Q-G", 100*(ds[rows[2],"Quasi"]
       - ds[rows[2],"Gamma"])/ds[rows[2],"Gamma"], "Nonlinear")
names(dg) <- c("Data", "Parameter", "Delta", "Percent", "Solution")
(deltas <- rbind(deltas,dg))
rm(dg)


ci <- data.frame(dset, "lambda", ds[7,"Gamma"], ds[8,"Gamma"],
      ds[9,"Gamma"], " ", " ")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist", "Solution")
CIs <- rbind(CIs,ci)
ci <- data.frame(dset, "psi", ds[10,"Gamma"], ds[11,"Gamma"],
       ds[12,"Gamma"], " ", "Closed")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
              "Dist", "Solution")
CIs <- rbind(CIs,ci)
ci <- data.frame(dset, "psi", ds[22,"Gamma"], ds[23,"Gamma"],
       ds[24,"Gamma"], " ", "Nonlinear")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist", "Solution")
CIs <- rbind(CIs,ci)
se <- qnorm(0.975)*ds[rows[1],"Normal"]/sqrt(m$RandC)
ci <- data.frame(dset, pltt, ds[rows[1],"Normal"] - se,
       ds[rows[1],"Normal"], ds[rows[1],"Normal"] + se, "Normal", " ")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
              "Dist","Solution")
CIs <- rbind(CIs,ci)
se <- qnorm(0.975)*ds[rows[2],"Normal"]/sqrt(m$RandC)
ci <- data.frame(dset, pltt, ds[rows[1],"Gamma"] - se,
       ds[rows[1],"Gamma"], ds[rows[1],"Gamma"] + se,
       "Gamma", " ")
```

```
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist","Solution")
CIs <- rbind(CIs,ci)
ci <- data.frame(dset, pltt, ds[rows[1],"Quasi"] - se,
       ds[rows[1],"Quasi"], ds[rows[1],"Quasi"] + se,
       "Quasi", "Closed")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist", "Solution")
CIs <- rbind(CIs,ci)
ci <- data.frame(dset, pltt, ds[rows[2],"Quasi"] - se,
       ds[rows[2],"Quasi"], ds[rows[2],"Quasi"] + se,
       "Quasi", "Nonlinear")
names(ci) <- c("Data", "Parameter", "CI95l", "Estimate", "CI95h",
               "Dist", "Solution")
(CIs <- rbind(CIs,ci))
rm(ci)


#####     Bar plots     ##############################################
(loc <- paste("Plots/Bar", pltt, pd, dset, ".png", sep=""))
quartz(w=6, h=6, bg="white")
par(lwd=2,family="serif")
bp <- barplot(as.matrix(ds[rows,3:5]), beside=T, density=10,
        angle=c(45,-45), col=c("black","black"), ylim=yl, sub=sub,
        ylab=ylab)
text(bp, as.matrix(ds[rows,3:5])+os,
    as.character(round(as.matrix(ds[rows,3:5]),4)),
    pos=1, cex=0.8)
ifelse(pltt == "VMR",
(title("Comparison of Variance-to-Mean Ratios",
   line=3)),
(title("Comparison of Random Effects Average Standard Errors",
    line=3))
)
title(expression(
   paste("SSN Data Closed Form and Nonlinear Solutions for ", psi)),
   line=2)
ifelse(dist == "Normal",
(title(expression(paste(lambda,
     " From Assumed Normal Random Effects")), line=1)),
(title(expression(paste(lambda,
    " From Assumed Gamma Random Effects")), line=1))
)
legend("topleft", c("Closed","Nonlinear"), density=10,
     angle=c(45,-45), col=c("black","black"), bty="n")
par(lwd=1,family="sans")
grid(nx=NA, ny=NULL, col="black")
quartz.save(loc, type="png")
```

```
#####     Save data set     ##########################################

setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()
outfile <- paste0(WD, "/", "deltas", ".RData")
save(deltas, file=outfile, ascii=FALSE)

#####     Load data set     ##########################################

setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()
outfile <- paste(WD, "/", "deltas", ".RData", sep="")
load(outfile)

#####     End load model     #########################################


#####     Save data set     #########################################

setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()
outfile <- paste0(WD, "/", "CIs", ".RData")
save(CIs, file=outfile, ascii=FALSE)

#####     Load data set     ##########################################

setwd("/Users/Jamie/Desktop/UNC/Dissertation/ghglm")
WD <- getwd()
outfile <- paste(WD, "/", "CIs", ".RData", sep="")
load(outfile)

#####     End load model     #########################################

######################################################################
```