ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

PhD THESIS

Fikriye KURTOĞLU

ESTIMATION METHODS IN GENERALIZED LINEAR MODELS

DEPARTMENT OF STATISTICS

ADANA, 2017

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

ESTIMATION METHODS IN GENERALIZED LINEAR MODELS

Fikriye KURTOĞLU

PhD THESIS

DEPARTMENT OF STATISTICS

We certify that the thesis titled above was reviewed and approved for the award of degree of the Doctor of Philosophy by the board of jury on 10/11/2017.

.....	.....	.....
Prof. Dr. Mahmude Revan ÖZKALE	Prof. Dr. Selahattin KAÇIRANLAR	Prof. Dr. Müjgan TEZ
SUPERVISOR	MEMBER	MEMBER
.....	.....	
Prof. Dr. Esin FİRUZAN	Assoc. Prof. Dr. Hüseyin GÜLER	
MEMBER	MEMBER	

This PhD Thesis is written at the Department of Statistics of Institute of Natural and Applied Sciences at Çukurova University.
Registration Number:

Prof. Dr. Mustafa GÖK
Director
Institute of Natural and Applied Sciences

ABSTRACT

PhD THESIS

ESTIMATION METHODS IN GENERALIZED LINEAR MODELS

Fikriye KURTOĞLU

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF STATISTICS

Supervisor : Prof. Dr. Mahmude Revan ÖZKALE
Year: 2017, Pages: 145
Jury : Prof. Dr. Mahmude Revan ÖZKALE
: Prof. Dr. Selahattin KAÇIRANLAR
: Prof. Dr. Müjgan TEZ
: Prof. Dr. Esin FİRUZAN
: Assoc. Prof. Dr. Hüseyin GÜLER

Multicollinearity among the explanatory variables seriously effects the maximum likelihood estimator in linear regression models which results in large estimates in absolute value and in large variance-covariance matrix. The adverse effects of multicollinearity on parameter estimation in generalized linear models (GLMs) are also explored by various authors in the case of maximum likelihood estimator. Although ridge estimator and restricted estimator were proposed as an alternative to maximum likelihood estimator in GLMs, there are a few number of methods of the alternative estimators in GLMs.

In this study, restricted ridge, Liu and restricted Liu estimators are introduced to combat multicollinearity in GLMs and mean squared error comparisons are done in the context of first-order approximated estimators. The performance of these estimators is examined in detail for the gamma and Poisson response models. The results are illustrated by conducting numerical examples and simulation studies.

Key Words: Generalized linear models, Multicollinearity, Mean squared error, Gamma distribution, Poisson distribution.

ÖZ

DOKTORA TEZİ

GENELLEŞTİRİLMİŞ LİNEER MODELLERDE TAHMİN YÖNTEMLERİ

Fikriye KURTOĞLU

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK

Danışman : Prof. Dr. Mahmude Revan ÖZKALE
Yıl: 2017, Sayfa: 145
Jüri : Prof. Dr. Mahmude Revan ÖZKALE
: Prof. Dr. Selahattin KAÇIRANLAR
: Prof. Dr. Müjgan TEZ
: Prof. Dr. Esin FİRUZAN
: Doç. Dr. Hüseyin GÜLER

Lineer regresyon modellerinde, açıklayıcı değişkenler arasında çoklu iç ilişki, en çok olabilirlik tahmin ediciyi önemli derecede etkilemektedir. Bu durumda tahminlerin mutlak değerleri ve varyansları oldukça büyük olur. Genelleştirilmiş lineer modellerde (GLM) de çoklu iç ilişkinin parametre tahmini üzerindeki olumsuz etkileri, en çok olabilirlik tahmin edici için çeşitli yazarlar tarafından incelenmiştir. GLM'de en çok olabilirlik tahmin ediciye alternatif olarak ridge tahmin edici ve kısıtlı tahmin edici önerilmesine rağmen, alternatif tahmin yöntemleri ile ilgili çok az çalışma bulunmaktadır.

Bu çalışmada, GLM'de, çoklu iç ilişkiyi giderebilmek için kısıtlı ridge, Liu ve kısıtlı Liu tahmin ediciler elde edilmiş ve hata kareler ortalaması kriterine göre birinci sıra yaklaşık tahmin ediciler ile karşılaştırma yapılmıştır. Bu tahmin edicilerin performansı, gamma ve Poisson yanıt değişkenli modellerde detaylı bir şekilde incelenmiştir. Elde edilen sonuçlar sayısal örnekler ve simülasyon çalışmaları yapılarak örneklendirilmiştir.

Anahtar Kelimeler: Genelleştirilmiş lineer modeller, Çoklu iç ilişki, Hata kareler ortalaması, Gamma dağılımı, Poisson dağılımı.

GENİŞLETİLMİŞ ÖZET

GLM ortalamanın fonksiyonlarını modelleyen normal dağılımlı ve normal dağılımlı olmayan yanıt değişkenlerini kapsayan sıradan regresyon modellerinin uzantısıdır. Nelder ve Wedderburn (1972) tarafından tanıtılan GLM, üstel dağılım ailesi olarak adlandırılan ailenin üyelerinden birinin, bir yanıt değişkenin dağılımı olarak seçimine izin veren lineer ve lineer olmayan regresyon modellerinin bir birleşimidir. Yanıt değişken, üstel ailenin üyelerinden olan normal, binom, Poisson, üstel, gamma, ters normal ve negatif binom dağılımlarından birine sahiptir. GLM'de normallik ve sabit varyans varsayımı gerekmemektedir. Ancak tüm üstel aile üyelerinde varyans ortalamanın bir fonksiyonudur.

Lineer regresyon modellerinde, açıklayıcı değişkenler arasında çoklu iç ilişki, en çok olabilirlik tahmin ediciyi etkilemektedir. Bu durumda tahminlerin mutlak değerleri ve varyansları oldukça büyük olur. Çoklu iç ilişki problemi, uygulamalı çalışmalarda çok yaygındır ve regresyon katsayılarının istatistiksel olarak anlamsız çıkmasına ve de katsayıların tahminlerinin yorumlanmasında zorluklara yol açabilmektedir. GLM'de bilgi matrisi, en çok olabilirlik tahmin edicinin değişken ve yüksek varyansa sahip olmasına yol açan kötü koşullu durumda olabilir. Lineer regresyon modelleri ile benzerliğinden, Mackinnon (1986), Mackinnon ve Puterman (1989), Nyquist (1991) ve Segerstedt (1992) çalışmalarında GLM'de de çoklu iç ilişki olabileceğini belirtmişlerdir. Literatürde çoklu iç ilişkili lineer model için birçok araştırma yapılmıştır. Bunlardan en popüleri Hoerl ve Kennard'ın (1970) ridge regresyon tahmin edicidir. Lineer regresyon ile benzerliği düşünülerek, bu tahmin edici Segerstedt (1992) tarafından GLM'e taşınmıştır.

Lineer regresyon modellerinde karşılaşılan çoklu iç ilişki problemi GLM'de de ortaya çıkabilmektedir (Mackinnon, 1986). Lineer regresyon modellerinde, çoklu

iç ilişki probleminin etkisini azaltmada kullanılan yöntemlerin lineer kısıtlamalar kullanılarak yeni tahmin edicilerin önerilmesi ve yanlı tahmin edicilerin kullanılması olduğu bilinmektedir. Nyquist (1991), GLM'de kısıtlı tahmin edici olarak adlandırdığı yeni bir tip tahmin edici tanıtmıştır. Segerstedt (1992), GLM'de ridge tahmin ediciyi önermiştir.

Bu çalışmada, GLM'den, GLM'de çoklu iç ilişkiden, çoklu iç ilişki varlığında ortaya çıkan problemleri gideren yeni tahmin yöntemlerinden, yakınsama kriterlerinden, GLM ailesinin üyeleri olan binom, Poisson ve gamma regresyondan bahsedilmiştir. Elde edilen yeni tahmin ediciler hata kareler ortalaması matrisi kriterine göre teorik olarak karşılaştırılmıştır. Ayrıca önerilen tahmin edicilerin performansı nümerik örnek ve Monte Carlo simülasyon çalışması ile incelenmiştir.

Çalışmada ilk olarak lineer regresyon durumuna benzer bir şekilde, genelleştirilmiş lineer modeller için kısıtlı ridge tahmin edici tanımlanmıştır. Kısıtlı ridge tahmin edicinin elde edilmesinde kullanılan amaç fonksiyonuna $R\beta = r$ kısıtı eklenmesi ile

$$\begin{aligned}\hat{\beta}_r(k) &= (X^\top \hat{W} X + kI)^{-1} X^\top \hat{W} \hat{z} + (X^\top \hat{W} X + kI)^{-1} R^\top \\ &\quad \times \left\{ R(X^\top \hat{W} X + kI)^{-1} R^\top \right\}^{-1} \left\{ r - R(X^\top \hat{W} X + kI)^{-1} X^\top \hat{W} \hat{z} \right\}\end{aligned}$$

tahmin edici elde edilmiştir. Bu tahmin edici GLM'de kısıtlı ridge tahmin edici olarak adlandırılmıştır. Bu tahmin edici iteratif olarak yeniden ağırlıklandırılmış en küçük kareler, kısıtlı ve ridge tahmin edici ile hata kareler ortalaması matrisi kriterine göre teorik olarak karşılaştırılmıştır.

Çalışmada tanımlanan diğer tahmin edici, lineer regresyon modellerindeki Liu tahmin edicinin bir uzantısı olan ve GLM'deki çoklu iç ilişkiyi gidermek için önerilen birinci sıra yaklaşık Liu tahmin edici

$$\hat{\beta}_d^{(1)} = \beta^{(0)} + \{(X^\top W X + I)^{-1} [X^\top W D(y - \mu) + d\hat{\beta}^{(1)} - \beta]\}_{\beta = \beta^{(0)}}$$

dir. Bu tahmin edici için ayrıca Liu yanlılık parametresi d'nin seçimi için kullanılabilecek yöntemler detaylı bir şekilde incelenmiştir. Bu tahmin edici iteratif olarak yeniden ağırlıklandırılmış en küçük kareler tahmin edicisi ile hata kareler ortalaması matrisi kriterine göre teorik olarak karşılaştırılmıştır. Sonuçlar nümerik örnek ve simülasyon çalışmaları ile desteklenmiştir.

Çalışmada son olarak, GLM'de çoklu iç ilişki probleminin üstesinden gelebilmek için diğer bir alternatif yöntem olan kısıtlı Liu tahmin edicisi tanımlanmıştır. Liu tahmin edicinin elde edilmesinde kullanılan amaç fonksiyonuna $R\beta = r$ kısıtı eklenmesi ile

$$\hat{\beta}_r(d)^{(1)} = \hat{\beta}(d)^{(1)} - (X^\top W^{(0)} X + I)^{-1} R^\top [R(X^\top W^{(0)} X + I)^{-1} R^\top]^{-1} (R\hat{\beta}(d)^{(1)} - r)$$

birinci sıra yaklaşık kısıtlı Liu tahmin edici elde edilir.

Bu amaçla, bu tezde, GLM'de çoklu iç ilişki olması durumunda tahmin yöntemlerine üç yeni yaklaşım sunulmuştur. GLM'de çoklu iç ilişkinin üstesinden gelebilmek için elde edilen tüm bu tahmin yöntemlerinin performansları hata kareler ortalamalarına dayalı olarak, nümerik örnek ve Monte Carlo simülasyonu ile incelenmiştir. Simülasyon çalışmasında, bu tahmin edicilerin, farklı çoklu iç ilişki derecesi, farklı örneklem büyüklüğü, farklı kısıtlar ve farklı yanlılık parametreleri alınarak hata kareler ortalaması performansları karşılaştırılmıştır. Böylelikle birçok bilim dalında ihtiyaç olan GLM'in, çoklu iç ilişkiye sahip olmaları durumunda üç yeni tahmin yöntemi literatüre kazandırılmıştır.

ACKNOWLEDGEMENTS

I would like to put my deepest gratefulness into words to my advisor, Prof. Dr. M. Revan ÖZKALE, for her endless support. Her great support made me feel confident and encouraged. It is difficult for me to express my whole feelings in words, but I should point out that I am happy and lucky to have had an opportunity to work with her.

I am also thankful to Prof. Dr. Selahattin Kaçıranlar, Prof. Dr. Müjgan Tez, Prof. Dr. Esin Firuzan and Assoc. Prof. Dr. Hüseyin Güler for their relevant discussions, suggestions and comments. I would like to thank to all instructors and members who lectured me and made contribution to my skills at Çukurova University for their all support. My warm and sincere thanks to my best friends for being always with me.

I would like to express my appreciations to my family and my sister Özlem for entering our life and making it colorful for us. This thesis is the product of endless support of Kurtoğlu family.

CONTENTS **PAGE**

LIST OF TABLES **PAGE**

LIST OF FIGURES **PAGE**

LIST OF ABBREVIATONS

I	: Identity matrix
$a(\phi)$	: Dispersion parameter
λ_j	: Eigenvalue
λ_{max}	: Maximum eigenvalue
Λ^*	: Diagonal matrix of eigenvalues
T	: Matrix of eigenvectors
W	: Weight matrix
z	: Working-response
k	: Ridge parameter
d	: Liu-biasing parameter
η	: Linear predictor
$g(\cdot)$	: Link function
$H_l(.)$	: Hessian matrix
$\hat{\beta}$	: Iteratively reweighted least squares estimator
$\hat{\beta}_k$	: Ridge estimator
$\hat{\beta}_d$	: Liu estimator
$\hat{\beta}_r(k)$	: Restricted ridge estimator
$\hat{\beta}_r$	: Restricted maximum likelihood estimator
$\hat{\beta}_r(d)$	: Restricted Liu estimator
$\hat{\beta}^{(1)}$	: First-order approximated maximum likelihood estimator
$\hat{\beta}_d^{(1)}$	: First-order approximated Liu estimator
$\hat{\beta}_r(k)^{(1)}$	: First-order approximated restricted ridge estimator
$\hat{\beta}_r^{(1)}$	: First-order approximated restricted maximum likelihood estimator
$\hat{\beta}_r(d)^{(1)}$	: First-order approximated restricted Liu estimator

GLMs : Generalized linear models

OLS : Ordinary least squares

ML : Maximum likelihood

IRLS : Iteratively reweighted least squares

MSE : Mean squared error

sMSE : Scalar mean squared error

EMSE : Estimated mean squared error

RLS : Restricted least squares

tr : Trace

Var : Variance-covariance matrix

1. INTRODUCTION

Nelder and Wedderburn (1972) first introduced generalized linear models (GLMs) which are actually a uniting approach of both linear and nonlinear regression models. Under the framework of GLMs, discrete as well as continuous responses can be accomodated, and the usual asssumptions of normality and constant variance are not necessarily made on the response under consideration. Thus, GLMs provide a unified representation for a large class of models for discrete variables and continuous responses. They have, therefore, proven to be very effective in several areas of application (Khuri, 2009). GLMs include logistic and probit regression models for binary response data, Poisson regression models for count response data, linear regression models with known variance and models for lifetime data.

Collinearity has long been recognised as a potential source of problem in the estimation, computation and interpretation of linear model parameters (Belsley et al, 1980). As in the case in linear regression, collinearity causes difficulty in the interpretation of estimated regression coefficients in GLMs. Collinearity problem is very common in applied research and it may lead to some of regression coefficients being statistically insignificant and difficulties in interpreting the estimates of an individual coefficient. The information matrix in GLMs may be ill conditioned which leads to instability and high variance of the maximum likelihood (ML) estimator. By analogy with the linear regression models, Mackinnon (1986), Mackinnon and Puterman (1989), Nyquist (1991) and Segerstedt (1992) state that collinearity can be presented in the GLMs and it has similar effects in GLMs as in linear regression models. Mackinnon and Puterman (1989) study how to detect the collinearity in GLMs. Marx and Smith (1990) present a principle component estimator for GLMs, and show that it can be useful with the presence of an ill conditioned information matrix. Nyquist

(1991) considers the ML estimator in GLMs under linear restrictions on the parameters. Segerstedt (1992) introduces the ridge regression estimator for GLMs.

In the literature, many studies have been conducted for linear regression models under multicollinearity where the ridge regression estimator of Hoerl and Kennard (1970) is the most popular. There has been a substantial amount of interest in estimating the regression coefficients in linear regression models as a remedy for collinearity. The ridge estimator of Hoerl and Kennard (1970) is the most popular of the estimators. On the other hand, if the model of interest is connected with a system of linear restrictions, then the estimates are obtained based on the restrictions. Although restricted least squares (RLS) estimator is proposed in linear regression, the method of RLS may also produce poor estimates when multicollinearity is present and provide misleading information as in ordinary least squares (OLS) estimates. Hence, Sarkar (1992) introduced the restricted ridge estimator which is a linear transformation of RLS estimator. However, Groß (2003) stated the restricted ridge estimator of Sarkar (1992) does not satisfy the restrictions for every outcome of the response and therefore is not a restricted estimator with respect to the linear restrictions and take values outside the subspace generated by the linear restrictions. To this end, he proposed a restricted ridge estimator which is a restricted estimator to a given restriction. The Liu estimator which is proposed by Liu (1993) for linear regression model is an alternative estimator to overcome multicollinearity. As in the case in the linear regression models, Mackinnon (1986), Mackinnon and Puterman (1989) and Segerstedt (1992) stated that collinearity can be present in the GLMs which results in inflated asymptotic variance-covariance matrix of the ML estimator and large ML estimates in absolute value. Considering the similarities between the linear regression models and the GLMs, Segerstedt (1992) introduced the ridge estimator in GLMs. In the context of GLMs, restrictions can also appear. For example, restrictions appear in sum-to-zero

parameterizations when explanatory variables are qualitative and in quantal response problems with constant relative potency between drugs (see, Nyquist 1991), or in order to analyze respiratory cancer deaths among a cohort of smelter workers exposed to airborne arsenic trioxide, the death rates for each exposure might form a nondecreasing sequence which implies restrictions in Poisson regression model (see, Paula 1993, McDonald and Diamond 1990), or in a study that natural abortion rates are argued to increase with the degree of consanguinity of the parents, a restricted logistic model is adapted to explain the proportion of natural abortions (see, Paula 1993, McDonald and Diamond 1990). Hence, Nyquist (1991) considered the ML estimation in the GLMs under linear restrictions on parameters. Kibria and Saleh (2012) proposed the restricted ridge estimator in probit model. They expressed the restricted ridge estimator as a transformation of the ML estimator where the idea is followed by Sarkar (1992) in linear regression.

The main purpose of this thesis is to introduce three new estimators to cope with the effects of multicollinearity in GLMs. The superiorities of these three new estimators are examined theoretically. The numerical examples and simulation studies are conducted in order to see the effect of collinearity on the superiority of these estimators in GLMs with respect to MSE criterion.

This thesis is organized as follows: The first chapter comprises of the introduction of the research, research background. The second chapter of the thesis starts to give requirements of estimation methods in GLMs. Exponential family and link function types in the GLMs are presented. Basic principles of the GLMs based on ML methods are discussed. The purpose of that chapter is to provide an introduction to the subject of GLMs. The third chapter discusses multicollinearity in GLMs and proposes different biased estimation methods in GLMs such as restricted ridge, Liu, restricted Liu. The methodology and applications of each estimator are described and

presented step by step. Several numerical examples are performed to demonstrate the performance of proposed methods in the fourth chapter. Results are presented and proposed estimators are compared with the existing estimators. In chapter five, several simulation studies are performed to demonstrate the performance of proposed methods. The superiority of the proposed estimation methods are deeply analyzed during Monte Carlo simulation studies according to the mean squared error (MSE) criterion. The last chapter gives concluding remarks which summarizes the main contributions of the thesis and discusses directions for future work.

2. GENERALIZED LINEAR MODELS

There are many experimental situations where linearity of the postulated model and/or normality of the response Y may not be quite valid. For example, Y may be a discrete random variable, that is, it has values that can be counted or put in a one-to-one correspondence with the set of positive integers. A binary response having two possible outcomes, labeled as success or failure, is one example of a discrete random variable. In this case, one is usually interested in modeling the probability of success. In other situation, the response may have values in the form of counts (number of defects in a given lot, or the number of traffic accidents at a busy intersection in a given time period). Other examples of responses that are not normally distributed can be found in biomedical applications, clinical trials and quality engineering, to name just a few (Khuri, 2009). GLMs provide a unified approach to many of the most common statistical procedures used in applied statistics. They have applications in disciplines as widely varied as agriculture, demography, ecology, economics, education, engineering, environmental studies and pollution, geography, geology, history, medicine, political science, psychology and sociology (Lindsey, 1997).

2.1. Introduction

The method of least squares has been commonly used as the basis for estimation and statistical inference in linear models where the response is normally distributed. As an estimation method, least squares is a mathematical method for minimizing the sum of squared errors that does not depend on the probability distribution of the response (Gbur et al, 2012). Under the framework of GLMs, discrete as well as continuous responses can be accommodated, and the usual assumptions of normality and constant variance are not necessarily made on the response under con-

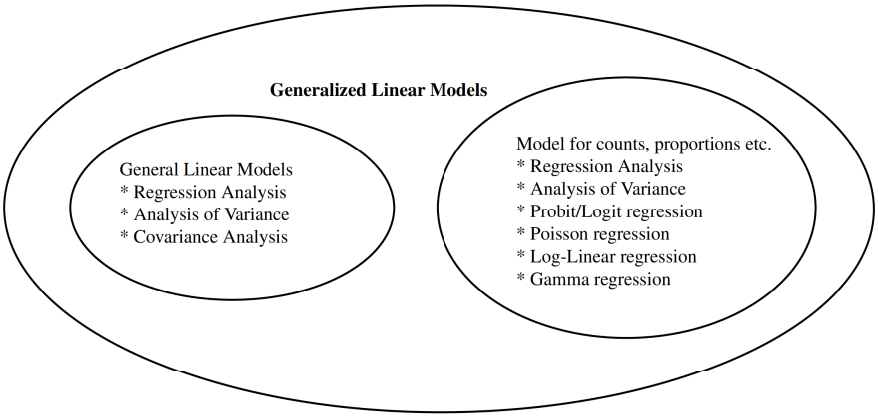

Figure 2.1. Family of generalized linear models (Olsson, 2002)

sideration. Thus, GLMs provide a unified representation for a large class of models for discrete and continuous responses. They have therefore proved to be very effective in several areas of application. For example, in biological assays, reliability and quality engineering, survival analysis and a variety of applied biomedical fields, the use of GLMs has become quite prevalent (Khuri, 2009). GLMs specify a relationship between the mean of the random variable Y and a function of the linear combination of the predictors. This generalization admits a model specification allowing for continuous or discrete outcomes and allows a description of the variance as a function of the mean (Hardin and Hilbe, 2012).

As Figure 2.1. shows, GLMs are a very general class of statistical models that includes many commonly used models as special cases. For example the class of GLMs that includes linear regression, analysis of variance and analysis of covariance, is a special case of GLMs. GLMs also include log-linear models for analysis of contingency tables, probit/logit regression, Poisson regression and much more (Olsson, 2002). The common linear model for a response Y has the form

$$Y = x^{\top}\beta + \varepsilon \tag{2.1.}$$

where x is a vector of known explanatory variables, β is a vector of unknown parameters and ε is an error term. Such models have found widespread application, and statistical theory for their analysis is well developed; see, for example, Draper and Smith (1981), Dobson and Barnett (2008). GLMs seek to extend the domain of applicability of Equation (2.1.) by relaxation of the assumption of additive error. In the linear model, the density of Y may be written as

$$f_Y(y) = f_\varepsilon(y - x^\top \beta),$$

the generalization yields a class of models of the form

$$f_Y(y) = f(y; x^\top \beta), \tag{2.2.}$$

in which x continues to appear only through the linear predictor, $\eta = x^\top \beta$. In many applications it is reasonable to assume the existence of the mean, say, $E(Y) = \mu$; then μ is determined by η and it is conventional to write

$$g(\mu) = \eta$$

and to call g the link function between mean and linear predictor. While f in Equation (2.2.) may be any suitable density function or probability function in the case of discrete data, the exponential families play a special role, relate to that of the normal distribution in classical linear models. In particular, parameter estimates based on the likelihood for an appropriate linear exponential family enjoy properties analogous to those of least squares estimates in the linear model (Firth, 1991).

A parametric model for the mean of a response Y may be represented as

$$E(Y) = \mu(\beta_1, \ldots, \beta_q)$$

where $\beta_1, \ldots, \beta_q$ are unknown parameters and $\mu(\cdot)$ is a known function. The model

is said to be linear if $\mu(\cdot)$ is a linear function of $\beta_1, \ldots, \beta_q$,

$$\mu = \sum_{j=1}^{q} x_j \beta_j$$

where $x_1, \ldots, x_q$ are known constants associated with the response Y. The x_j's may be quantitative variates, or they may be indicator variables representing the levels of a qualitative factor.

A GLM is one that is of the form

$$\mu = g^{-1}\left(\sum_{j=1}^{q} x_j \beta_j\right)$$

where $g(\cdot)$ is a one-to-one differentiable function. Thus GLMs are themselves non-linear, but in a restricted sense that permits strong parallels to be drawn with linear models. It is usual to write

$$g(\mu) = \eta = \sum_{j=1}^{q} x_j \beta_j.$$

The link function determines the scale on which the linearity is assumed. If the parameters $\beta_1, \ldots, \beta_q$ are not to be subject to restrictions, $g(\mu)$ must be able to achieve any value in $(-\infty, \infty)$ and so the form of a suitable link function is determined to some extent by the domain of variation of $\mu = E(Y)$. The choice of a link function from within the large class of functions satisfying this mild requirement is usually made on the grounds of (i) fit to the data; (ii) ease of interpretation of parameters in the linear predictor; and (iii) existence of simple sufficient statistics (Firth, 1991).

A GLM has two key features. The first is that the data have a distribution belonging to the exponential family. The second feature is that covariate x_i (associated with observation i) is included using a linear model, but this is a linear model for some appropriate transformation of $E(Y_i)$. For example, in the case of count data the standard model assumes that the data are Poisson distributed and that the log of $E(Y_i)$ follows a linear model (Millar, 2011).

2.2. The Exponential Family

The development of the theory of the GLMs is based upon the exponential family of distributions. This formalization simply recharacterizes familiar functions into a formula that is more useful theoretically and demonstrates similarity between seemingly disparate mathematical forms (Gill, 2001). GLMs are traditionally formulated within the framework of the exponential family of distributions. This family of distributions includes the Gaussian or normal, binomial, Poisson, gamma, inverse Gaussian, geometric and negative binomial (Hardin and Hilbe, 2012).

GLMs are characterized by three components given by the following (Hardin and Hilbe, 2012):

(i) A random component for the response Y which has the characteristic variance of a distribution that belongs to the exponential family.

(ii) A linear systematic component relating the linear predictor $\eta = X\beta$ to the product of the design matrix X and the parameters β.

(iii) A known monotonic, one-to-one, differentiable function called link function $g(\cdot)$ relating the linear predictor to the fitted values. Because the function is one-to-one, there is an inverse function relating the mean expected response, $E(Y) = \mu$, to the linear predictor such that $\mu = g^{-1}(\eta) = E(Y)$.

Suppose that $y_1, \ldots, y_n$ are observations of independent random variables $Y_1, \ldots, Y_n$, each of which has the probability density function such that the mean of Y_i is μ_i. The random component of a GLM consists of a response Y with independent observations from a distribution having probability density or mass function of the form for single

observation y

$$f(y,\theta,\phi) = L(\theta,\phi,y) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right] \tag{2.3.}$$

where $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are known functions, θ is a natural parameter and $\phi > 0$ is a dispersion parameter, which in some families takes on a fixed, known value, while in other families it is an unknown parameter to be estimated from the data.

Some elementary properties of linear exponential families follow from the familiar identities

$$\mathrm{E}(\partial l/\partial\theta) = 0,$$

$$\mathrm{E}(\partial^2 l/\partial\theta^2) + \mathrm{Var}(\partial l/\partial\theta) = 0$$

where $l = \log L(\theta,\phi,y)$. The first of these applied to Equation (2.1.) gives

$$\mathrm{E}(Y) = b'(\theta) = \mu$$

say, while the second implies that

$$\mathrm{Var}(Y) = a(\phi)b''(\theta) = a(\phi)\mathrm{Var}(\mu),$$

where $\mathrm{Var}(\mu)$ is the variance function associated with the family. Note that $\mathrm{Var}(Y)$ is the product of $a(\phi)$ which depends only on the dispersion parameter ϕ and $b''(\theta)$ which depends only on the canonical parameter θ and hence on the mean μ of the distribution of Y. Variance functions for the examples listed below are (Firth, 1991):

(i) Normal, $\mathrm{Var}(\mu) = 1$,

(ii) Poisson, $\mathrm{Var}(\mu) = \mu$,

(iii) Gamma $\mathrm{Var}(\mu) = \mu^2$,

(iv) Binomial, $\mathrm{Var}(\mu) = \mu(1-\mu)$.

The Link Function: The link function $g(\cdot)$ is a function relating the expected value of the response Y to the predictors $x_1, \ldots, x_q$. It has the general form $g(\mu) = \eta = X\beta$. The function $g(\cdot)$ must be monotone and differentiable. For a monotone function, we can define the inverse function $g^{-1}(\cdot)$ by the relation $g^{-1}[g(\mu)] = \mu$. The choice of link function depends on the type of data. For continuous normal theory data, an identity link may be appropriate. For data in the form of counts, the link function should restrict μ to be positive, while data in the form of proportions should use a link that restricts μ to the interval $[0,1]$. Some commonly used link functions and their inverses are as folllows:

- The identity link: $\eta = \mu$. The inverse is simply $\mu = \eta$.
- The logit link: $\eta = \log[\mu/(1-\mu)]$. The inverse $\mu = \exp(\eta)/(1+\exp(\eta))$ is restricted to the interval $[0,1]$.
- The probit link: $\eta = \Phi^{-1}(\mu)$, where Φ is the standard Normal distribution function. The inverse $\mu = \Phi(\eta)$ is restricted to the interval $[0,1]$.
- The complementary log-log link: $\eta = \log[-\log(1-\mu)]$. The inverse $\mu = 1-\exp[-\exp(\eta)]$ is restricted to the interval $[0,1]$.
- Power links: $\eta = (\mu^\lambda - 1)/\lambda$ where we take $\eta = \log(\mu)$ for $\lambda = 0$. Examples of power links are $\eta = \mu^2$; $\eta = 1/\mu$; $\eta = \sqrt{\mu}$ and $\eta = \log(\mu)$. These all belong to the Box-Cox family of transformations.

Canonical Links: Certain link functions are, in a sense, "natural" for certain distributions. These are called canonical links. The canonical link is that function which transforms the mean to a canonical location parameter of the exponential dispersion family member (Lindsey, 1997). This means that the canonical link is that function $g(\cdot)$ for which $g(\mu) = \theta$. It holds that (Olsson, 2002):

- Poisson: $\theta = \log(\mu)$ so the canonical link is log.
- Binomial: $\theta = \log[\mu/(1-\mu)]$ which is the logit link.
- Normal: $\theta = \mu$ so the canonical link is the identity link.
- Gamma: $\theta = 1/\mu$ which is the inverse link.

If $y_1, \ldots, y_n$ are independent, with each y_i from an exponential family with parameters (θ_i, ϕ_i), the log-likelihood function for the sample is given by

$$l(\beta) = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} \left[\frac{\theta_i y_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i) \right]. \tag{2.4.}$$

Suppose that structure is imposed on the θ_i through a GLM

$$g(\mu_i) = g[b'(\theta_i)] = \sum_{j=1}^{q} x_{ij}\beta_j, \quad i = 1, \ldots, n.$$

The likelihood for the regression parameters $\beta_1, \ldots, \beta_q$ is quite complicated in general; but in the special case where $g(\cdot)$ is the inverse of $b'(\theta_i)$, so that $g(\mu_i) = \theta_i$, the log-likelihood may be written in the form

$$\sum_{j=1}^{q} \beta_j \sum_{i=1}^{n} \frac{y_i x_{ij}}{\phi_i} - \sum_{i=1}^{n} \left\{ \frac{b(\theta_i)}{\phi_i} - c(y_i, \phi_i) \right\}.$$

If $\phi_1, \ldots, \phi_n$ are taken as known then the usual factorization criterion yields $\sum_{i=1}^{n} \frac{y_i x_{ij}}{\phi_i}$, $j = 1, \ldots, q$ as a set of minimal sufficient statistics for $\beta_1, \ldots, \beta_q$. The special link function $g = b'^{-1}$ which allows this simplifications is known as the canonical link associated with the linear exponential family. Canonical links for the most commonly used families are:

(i) Normal, $g(\mu) = \mu$;

(ii) Poisson, $g(\mu) = \log \mu$;

(iii) Gamma $g(\mu) = -1/\mu$;

(iv) Binomial, $g(\mu) = \log[\mu/(1-\mu)]$.

Note that the canonical link and the variance function are related by $\mathrm{Var}(\mu) = 1/g'(\mu)$ (Firth, 1991).

The link function relates the mean $\mu = \mathrm{E}(Y)$ to the linear predictor $X\beta$ and the variance function relates the variance as a function of the mean $\mathrm{Var}(Y) = a(\phi)\mathrm{Var}(\mu)$, where $a(\phi)$ is the dispersion parameter. For the Poisson, binomial, and negative binomial variance models, $a(\phi) = 1$. For GLMs, let us consider Table 2.1, which provides the probability structure, the appropriate canonical link function relating the mean to the predictors, and the variance for several distributions that are in the exponential family (Hardin and Hilbe, 2012).

Table 2.1. Distributions, variances and link functions for some distributions in the exponential family (Myers, 1990)

Distribution	$a(\phi)$	Canonical Link	Mean	Variance	Model (Can. Link)	θ_i	$g(\theta_i)$
Normal: $f(y_i,\mu_i,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-1/2((y_i-\mu_i)/\sigma)^2}$ $\sigma > 0$	σ^{-2}	$x_i^\top\beta = \mu_i$	μ_i	σ^2	$E(Y_i) = x_i^\top\beta$	μ_i	$\frac{\theta_i^2}{2}$
Poisson: $f(y_i,\mu_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}$ $y = 0,1,2,..,\ \mu_i > 0$	1	$x_i^\top\beta = \ln\mu_i$	μ_i	μ_i	$E(Y_i) = t_i e^{x_i^\top\beta}$	$\ln\mu_i$	e^{θ_i}
Binomial (Logistic): $f(y_i,n_i,P(x_i)) = \binom{n_i}{y_i} P(x_i)^{y_i}[1-P(x_i)]^{n_i-y_i}$ $y_i = 0,1,2,\ldots,\ 0 \le P(x_i) \le 1$	1	$x_i^\top\beta = \ln[\frac{\mu_i}{n-\mu_i}]$	$n_i P(x_i)$	$\mu_i[1-P(x_i)]$	$E(Y_i) = \frac{n_i}{1+e^{-x_i^\top\beta}}$	$\ln\frac{P(x_i)}{1-P(x_i)}$	$-n_i\ln[\frac{e^{-\theta_i}}{1+e^{-\theta_i}}]$
Gamma: $f(y_i,r,\lambda_i) = \frac{\lambda_i^r}{\Gamma(r)} e^{-y_i\lambda_i} y_i^{r-1}$ $y_i,\lambda_i,r > 0$	$-r$	$x_i^\top\beta = \frac{1}{\mu_i}$	$\frac{r}{\lambda_i}$	$\frac{\mu_i^2}{r}$	$E(Y_i) = \frac{1}{x_i^\top\beta}$	$\frac{\lambda_i}{r}$	$-\ln\left(\frac{1}{\theta_i}\right)$
Geometric: $f(y_i,P(x_i)) = P(x_i)[1-P(x_i)]^{y_i}$ $y_i = 0,1,2,\ldots,\ 0 \le P(x_i) \le 1$	1	$x_i^\top\beta = \ln[\frac{\mu_i}{1+\mu_i}]$	$[1-P(x_i)]/P(x_i)$	$\mu_i(1+\mu_i)$	$E(Y_i) = \frac{1}{e^{-x_i^\top\beta}-1}$	$\ln[1-P(x_i)]$	$-\ln(1-e^{\theta_i})$
Negative Binomial: $f(y_i,P(x_i),\alpha) = \binom{y_i+\alpha-1}{\alpha-1} P(x_i)^\alpha[1-P(x_i)]^{y_i}$ $y_i = 0,1,2,\ldots,\ \alpha > 0,\ 0 \le P(x_i) \le 1$	1	$x_i^\top\beta = -\ln[\frac{\alpha+\mu_i}{\mu_i}]$	$\alpha[1-P(x_i)]/P(x_i)$	$\mu_i(1+\frac{\mu_i}{\alpha})$	$E(Y_i) = \frac{\alpha}{e^{-x_i^\top\beta}-1}$	$\ln[1-P(x_i)]$	$-\alpha\ln(1-e^{\theta_i})$

2.3. Estimation

The method of ML is the theoretical basis for parameter estimation in the GLMs. However, the actual implementation of ML results in an algorithm based on iterative reweighted least squares (IRLS) (Montgomery et al, 2001).

2.3.1. Estimation of the Regression Parameters

In this section, we show how to estimate the parameter vector β in the model for the linear predictor η. To apply the method of ML, differentation of the log-likelihood in Equation (2.3.), with respect to each β_j in turn is needed which yields the likelihood estimation equations

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \left[\frac{\partial l_i}{\partial \beta_j} \right] = \sum_{i=1}^{n} \left[\frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j} \right] = 0$$

$$\frac{1}{a(\phi)} \sum_{i=1}^{n} \frac{y_i - \mu_i}{\mathrm{Var}(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0, \quad j = 1 \ldots, q.$$

In the case of a GLM, equation becomes

$$\frac{1}{a(\phi)} \sum_{i=1}^{n} \frac{y_i - \mu_i}{\mathrm{Var}(\mu_i)} \frac{x_{ij}}{g'(\mu_i)} = 0, \quad j = 1, \ldots, q. \tag{2.5.}$$

Notice that if $g(\cdot)$ is the canonical link, the equation simplifies to

$$\frac{1}{a(\phi)} \sum_{i=1}^{n} y_i x_{ij} = \frac{1}{a(\phi)} \sum_{i=1}^{n} \mu_i x_{ij}, \quad j = 1, \ldots, q$$

that is, the jointly sufficient statistics are equated with their expectations. In general, Equations (2.5.) are nonlinear in $\beta_1, \ldots, \beta_q$ and do not have an explicit solution.

The Newton Raphson algorithm: The Newton Raphson algorithm is based on quadratic approximation of the objective function (the log-likelihood), via linear approximation of its derivative. The Newton Raphson method iteratively solves non-linear equations, to determine the point at which a function takes its maximum. It

begins with an initial approximation by a second degree polynomial and then finds the location of that polynomial's maximum value. It then repeats this step to generate a sequence of approximations. This method converges to the location of the maximum when the function is suitable and/or the initial approximation is good.

Mathematically, here is how the Newton Raphson method determines the value $\hat{\beta}$ at which a function $L(\beta)$ is maximized. Let $u = \left(\frac{\partial L(\beta)}{\partial \beta_1}, \ldots, \frac{\partial L(\beta)}{\partial \beta_q}\right)^\top$ and H denotes the matrix having entries $h_{jk} = \frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_k}$, called the Hessian matrix. Moreover, let $u^{(m)}$ and $\mathrm{H}^{(m)}$ be u and H evaluated at $\beta^{(m)}$, approximation m for $\hat{\beta}$. Step m in the iterative process ($m = 0, 1, 2, \ldots$) approximates $L(\beta)$ near $\beta^{(m)}$ by the terms up to second order in its Taylor series expansion,

$$L(\beta) \approx L(\beta^{(m)}) + u^{(m)^\top}(\beta - \beta^{(m)}) + \frac{1}{2}(\beta - \beta^{(m)})^\top \mathrm{H}^{(m)}(\beta - \beta^{(m)}).$$

Solving $\frac{\partial L(\beta)}{\partial \beta} \approx u^{(m)} + \mathrm{H}^{(m)}(\beta - \beta^{(m)}) = 0$ for β yields the next guess.

That guess can be expressed as

$$\beta^{(m+1)} = \beta^{(m)} - \left(\mathrm{H}^{(m)}\right)^{-1} u^{(m)}$$

assuming that $\mathrm{H}^{(m)}$ is nonsingular. (However, computing routines use standard methods for solving the linear equations rather than explicitly calculating the inverse.)

Iterations proceed until changes in $L(\beta^{(m)})$ between successive cycles are sufficiently small. The ML estimator is the limit of $\beta^{(m)}$ as $m \to \infty$; however, this need not happen if $L(\beta)$ has other local maxima at which $u(\beta) = 0$. In that case, a good initial approximation is crucial (Agresti, 2015). Figure 2.2 illustrates a cycle of the method, showing parabolic (second-order) approximation at a given step.

Fisher scoring method: Fisher scoring is an alternative iterative method for solving likelihood equations. The difference from Newton Raphson is in the way it uses the Hessian matrix. Fisher scoring uses the expected value of this matrix, called

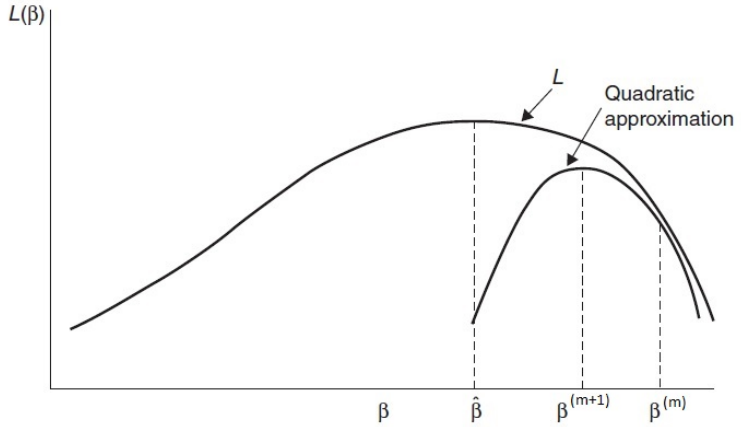

Figure 2.2. Illustration of a cycle of the Newton Raphson method (Agresti, 2015)

expected information, whereas Newton Raphson uses the Hessian matrix itself, called the observed information.

Let $I^{(m)}$ denotes approximation m for the ML estimate of the expected information matrix; that is $I^{(m)}$ has elements $-\mathrm{E}(\partial^2 L(\beta)/\partial\beta_j\partial\beta_k)$, evaluated at $\beta^{(m)}$. The formula for Fisher scoring is

$$\mathrm{I}^{(m)}\beta^{(m+1)} = \mathrm{I}^{(m)}\beta^{(m)} + u^{(m)}. \tag{2.6.}$$

Equation (2.6.) shows that $\mathrm{I}^{(m)} = X^\top W^{(m)} X$ where $W^{(m)}$ is weight matrix evaluated at $\beta^{(m)}$. The estimated asymptotic covariance matrix I^{-1} of $\hat{\beta}$ occurs as a by product of this algorithm as $(\mathrm{I}^{(m)})^{-1}$ for m at which convergence is adequate. For GLMs with a canonical link function, the observed and expected information are the same (Agresti, 2015).

It is used the algorithm as the Newton Raphson algorithm for solving the score equations for a canonical link function. When the algorithm is used for other link functions it is called the Fisher scoring algorithm. It corresponds to using the expected Fisher information instead of the observed information (Madsen and Thyre-

god, 2010).

Wedderburn (1976) discusses the existence and uniqueness of solutions to the ML equations in GLMs. When a solution vector can be found, it is consistent, asymptotically normal and asymptotically efficient, with distribution approximately $N_p(\beta, I^{-1})$. The information matrix $I = I(\beta)$ has elements

$$I(\beta)_{jk} = \sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{\phi \mathrm{Var}(\mu_i)\left[g'(\mu_i)\right]^2}$$

where ϕ is the dispersion parameter.

Let $\hat{\beta}^{(0)}$ be an initial estimate of β, and let $\hat{\beta}^{(m)}$ be the estimate of β at the mth iteration ($m \geq 1$). Then,

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} - \left\{ \mathrm{E}[\mathrm{H}_l(\beta)]_{\beta=\hat{\beta}^{(m)}} \right\}^{-1} \left[\frac{\partial l}{\partial \beta} \right]_{\beta=\hat{\beta}^{(m)}}, \quad m = 0, 1, \ldots \tag{2.7.}$$

with

$$\mathrm{E}[\mathrm{H}_l(\beta)] = -\frac{1}{a(\phi)} X^\top W X, \tag{2.8.}$$

$$\frac{\partial l}{\partial \beta} = \frac{1}{a(\phi)} X^\top W D (y - \mu) \tag{2.9.}$$

where X is an $n \times q$ model matrix whose ith row is $x_i^\top$ for the linear predictor, $W = diag(w_i)$, $D = diag[g'(\mu_i)]$, $y = (y_1, \ldots, y_n)^\top$, $\mu = (\mu_1, \ldots, \mu_n)^\top$ and "$diag$" superscript denotes the diagonal elements and $\mathrm{H}_l(\beta) = \partial^2 l / \partial \beta'$ is the Hessian matrix with the (j,k)th element

$$\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} = \frac{1}{a(\phi)} \sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial}{\partial \beta_k} \left[\frac{g'(\mu_i)}{w_i} x_{ij} \right] - \frac{1}{a(\phi)} \sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{w_i}$$

where x_{ik} is the kth element of $x_i^\top$.

In Equation (2.5.), the derivative of the log-likelihood can be denoted as

$$s(\beta, y) = \frac{\partial l(\beta, y)}{\partial \beta},$$

where $s(\beta, y)$ is commonly known as the score function when regarded as a function of β (for fixed y), or the score statistic when regarded as a function of y (for fixed β). The score statistic can be regarded as a random variable that is obtained from a transformation of y from R^n to R. The expected Fisher information is the second moment of this random variable (Millar, 2011).

Using Equations (2.8.) and (2.9.), Equation (2.7.) is then expressed as

$$\begin{aligned} \hat{\beta}^{(m+1)} &= \hat{\beta}^{(m)} + [(X^\top WX)^{-1}X^\top WD(y-\mu)]_{\beta=\hat{\beta}^{(m)}} \\ &= [(X^\top WX)^{-1}X^\top Wz]_{\beta=\hat{\beta}^{(m)}}, \quad m = 0,1,\ldots \end{aligned} \tag{2.10.}$$

where $z = X\beta + D(y-\mu)$ is the $n \times 1$ vector of working-response with elements

$$z_i = x_i^\top \beta + (y_i - \mu_i)g'(\mu_i), \quad i = 1,\ldots,n.$$

Starting values for IRLS in the case of a GLM are $\hat{\mu}_i^{(0)} = y_i$, $\hat{\eta}_i^{(0)} = g(\hat{\mu}_i^{(0)})$, although some adjustment may be necessary with particular link functions.

If the successive estimates $\hat{\beta}^{(m+1)}$ converge to, say $\hat{\beta}$, as m tends to infinity, we obtain $\hat{\beta} = (X^\top WX)^{-1}X^\top Wz$ where W and z are evaluated at the final iteration.

We conclude that $\hat{\beta}^{(m+1)} = (X^\top W^{(m)}X)^{-1}X^\top W^{(m)}z^{(m)}$ in Equation (2.10.) has the same form as a generalized (or weighted) least squares estimator of β in a linear model whose response vector is $z^{(m)}$ with given variance covariance matrix $W^{(m)}$. Several iterations of Equation (2.10.) can then be carried out until some convergence is achieved in the resulting values. We can therefore state that the solution of Equation (2.5.) is obtained by performing an IRLS procedure (Khuri, 2009).

Convergence criterion: There are different convergence criterion in the literature. Agresti (2002) suggested the convergence of $\beta^{(m)}$ to $\hat{\beta}$ for Newton Raphson method. For large m, the convergence satisfies, for each j,

$$\left|\beta_j^{(m+1)} - \hat{\beta}_j\right| \leq c\left|\beta_j^{(m)} - \hat{\beta}_j\right|^2$$

for some $c > 0$ and is referred to as second-order. This implies that the number of correct decimals in the approximation roughly doubles after sufficiently many iterations. In practice, it often takes relatively few iterations for satisfactory convergence.

The convergence criteria given by Myers et al (2010) is based on

$$\left| \frac{\beta_j^{(m+1)} - \hat{\beta}_j^{(m)}}{\beta_j^{(m)}} \right| < \delta, \quad j = 1, \ldots, q$$

where δ is some small number, say 10^{-6}.

Fahrmeir and Tutz (1994) presented the convergence criteria as

$$\frac{\left\| \hat{\beta}^{(m+1)} - \hat{\beta}^{(m)} \right\|}{\left\| \hat{\beta}^{(m)} \right\|} < \varepsilon$$

for some prechosen small number $\varepsilon > 0$.

However, a fairly natural iterative procedure is the following:

1. Start with initial estimates $\hat{\mu}_i^{(0)}$ and $\hat{\eta}_i^{(0)} = g(\hat{\mu}_i^{(0)})$ for the means and linear predictors.
2. The existence of an explicit solution in this special case suggests a particular method of solution more generally. If we consider the quantities

$$z_i = \eta_i + (y_i - \mu_i) g'(\mu_i)$$

then $\mathrm{E}(z_i) = \eta_i = \sum_{i=1}^{n} x_{ij}\beta_j$ is a linear model. Thus if z_i are known, $\beta_1, \ldots, \beta_q$ can be estimated by weighted least squares, with weights given by reciprocals of the variances $\mathrm{Var}(z_i) = [g'(\mu_i)]^2 \phi_i \mathrm{Var}(\mu_i)$. Of couse, $z_1, \ldots, z_n$ are in practice unknown since the η_i and μ_i are unknown. Given $\hat{\mu}_i^{(m)}$ and $\hat{\eta}_i^{(m)}$, form the adjusted dependent variate,

$$\hat{z}_i^{(m)} = \hat{\eta}_i^{(m)} + (y_i - \hat{\mu}_i^{(m)}) g'(\hat{\mu}_i^{(m)}), \quad i = 1, \ldots, n$$

and iterative weight

$$\hat{w}_i^{(m)} = \frac{1}{\phi_i \text{Var}(\hat{\mu}_i^{(m)})[g'(\hat{\mu}_i^{(m)})]^2}, \quad i = 1, \ldots, n.$$

3. Calculate $\hat{\beta}^{(m+1)}$ by weighted least squares regression of $\hat{z}^{(m)}$ on X:

$$\hat{\beta}^{(m+1)} = (X^\top \hat{W}^{(m)} X)^{-1} X^\top \hat{W}^{(m)} z^{(m)}$$

where $\hat{W}^{(m)} = \text{diag}(\hat{w}_i^{(m)})$. Define $\hat{\eta}^{(m+1)} = X\hat{\beta}^{(m+1)}$ and $\hat{\mu}^{(m+1)} = g^{-1}(\hat{\eta}^{(m+1)})$.

4. Repeat steps 2 and 3 until a suitable convergence criterion is satisfied.

This procedure is known as IRLS. It is simply the Newton Raphson method in models with canonical link, and more generally it coincides with the method known as Fisher scoring; see McCullagh and Nelder (1989). GLMs are fit by applying Fisher scoring within the Newton Raphson method applied to the entire single-parameter exponential family of distributions. The estimation algorithm is then referred to as IRLS.

2.3.2. Estimation of the Dispersion Parameter

Note that, thus far, we have assumed that the dispersion parameter ϕ was a known fixed parameter, and hence that the likelihood was a function only of the regression parameters in the linear predictor. As indicated earlier, this is entirely appropriate for binomial and Poisson responses when the dispersion parameter is defined to be one. For other distributions in the exponential family, however, this parameter will also require estimation (Cook, 2005).

In principle, ϕ could be estimated by ML. However, there are practical difficulties associated with this in some families; see, for instance, McCullagh and

Nelder (1989). In practice it is more common to use a "method of moments" estimator. If the regression parameters $\beta_1, \ldots, \beta_q$ were known, an unbiased estimator of $\phi = \text{Var}(Y_i)/\text{Var}(\mu_i)$ would be

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(y_i-\mu_i)^2}{\text{Var}(\mu_i)}.$$

Allowance for the fact that $\beta_1, \ldots, \beta_q$ must be estimated may be made by analogy with the case $\text{Var}(\mu) = 1$ and $g(\mu) = \mu$, resulting in the "degrees of freedom" corrected estimate for fitted value $\hat{\mu}_i$

$$\tilde{\phi} = \frac{1}{n-q}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu}_i)^2}{\text{Var}(\hat{\mu}_i)}.$$

This estimate is unbiased for ϕ in the case of the constant variance, linear model and consistent more generally (Firth, 1991).

3. BIASED ESTIMATION METHODS IN GLMS

3.1. Multicollinearity in GLMs

Collinearity has long been recognised as a potential source of problems in the estimation, computation and interpretation of linear model parameters (Belsley et al, 1980). However, its effects in GLMs are relatively unexplored. Collinearity causes difficulty in the interpretation of estimated regression coefficients as partial derivatives. The usual correlation matrix will not always be adequate for the detection of collinearity. Instead, collinearity diagnostics should be based on the estimated information matrix (Mackinnon and Puterman, 1989).

It has been well known that high linear dependency among covariates in any regression model, including not only traditional regression models but GLMs, can pose frustrating statistical problems, such as unsatisfactorily inflated variances, lowered power in prediction and even incorrect signs, to the estimation results (Belsley et al, 1980). An evident example resides in the OLS estimation in linear regression model. Strong linear dependecy among certain covariates renders the data matrix X rank deficient, and so is the $X^\top X$ matrix that plays an important role in the OLS estimation. Similar problems can occur in the GLMs. Although the estimation procedure is based on the ML estimation, the ML estimator is often obtained through the IRLS algorithm, involving again the computation of matrix inverse (McCullough and Nelder, 1989). When the $X^\top X$ matrix is ill conditioned, the weight matrix $X^\top WX$ can be ill conditioned as well, for some weight matrix W and then problems take place (Lee and Silvapulle, 1988; Schaefer et al, 1984).

In the context of linear regression, the detection of the presence of collinearity has been well studied. For instance, the correlation coefficients, the condition number and the variance inflation factors, all which are based upon the information

from correlation matrix, have been thoroughly discussed (Belsley et al, 1980). However, a few number of works have been done for GLMs. In the work of Schaefer et al (1984), authors discussed three aspects that serve as the criteria for collinearity diagnostic, inclusive of i) The R^2 value when regressing one variable on the others; ii) The sum of squared residuals in (i); iii) The minimal eigenvalue of the $X^\top \hat{W} X$ matrix. It is intuitive to see that when collinearity is high, the R^2 tends to be 1, the sum of squared residuals tends to be 0, and the minimal eigenvalue tends to be 0. A similar discussion can be found in Schaefer (1986). In conjunction with (iii) above, the criterion of looking at the minimal eigenvalue is equivalent to looking at the condition number of a matrix. The authors (Lee and Silvapulle, 1988) suggested that condition number can be a formal diagnostic tool for the detection of collinearity and a value exceeding 10 (Belsley et al, 1980) can be used as a threshold to tell if collinearity is present. Similarly, in the work of Lesaffre and Marx (1993), authors also treated condition number as the main tool for collinearity diagnostic. In other works such as Weissfeld and Sereika (1991), authors developed some collinearity diagnostics using the singular value decomposition technique.

Biased estimation has been a widely accepted techique in dealing with non-orthogonal problem in regression. The most famous techniques are the ridge regression (Hoerl and Kennard, 1970), restricted ML and Liu (Liu, 1993) estimators.

Schaefer (1986) and Schaefer et al (1984) proposed the ridge logistic estimator as a remedy for logistic regression having collinearity problem. Follow-up research, such as Le Cessie and Houwelingen (1992), Lee and Silvapulle (1988), all discussed the ridge logistic estimator in the same fashion. Besides, similar idea applied to the Poisson regression has been developed as well (Månsson and Shukur, 2011).

Collinearity in a GLM is more difficult to define than its counterpart in the

standard linear case. Intuitively it is again said to be present when (Huang et al, 2016):

- Two or more explanatory variables are highly correlated,
- Coefficient estimates are different in magnitude or sign from those hypothesised,
- Coefficient estimates are highly variable,
- Data collection is deficient,
- Matrices used in the computation of the coefficient estimates are ill conditioned (Mackinnon, 1986).

3.1.1. Diagnostics for Multicollinearity in GLMs

Several techniques have been proposed for detecting multicollinearity in a standard linear model. Collinearity in a GLM is more difficult to define than its counterpart in the standard linear case (Mackinnon, 1986). The diagnostics of multicollinearity are also explored by various authors in GLMs. Collinearity in GLMs can inflate variances of the estimated coefficients and cause poor prediction in certain regions of the regression space. In the past several decades, some literature discussed the collinearity problems in the logistic regression framework (see Schaefer et al, 1984; Schaefer, 1986; Marx and Smith, 1990b). Others explored this problem in the framework of GLMs (see Mackinnon and Puterman, 1989; Marx and Smith, 1990a; Weissfeld and Sereika, 1991; Lesaffre and Marx, 1993). Lesaffre and Marx (1993), following a suggestion of Mackinnon and Puterman (1989), investigated the dependence of the ill conditioning problems on the particular value of β. They concluded that in GLMs the response and the choice of the model also play a role in the degree of collinearity (ill conditioning) of the information matrix. The term, ML-collinearity, is given in their paper to describe the scenario when the explanatory variables are not

collinear, but at the ML estimate of the model parameter there is collinearity among the weighted explanatory variables.

In GLMs, the condition number for ill conditioned information matrix is defined by Mackinnon and Puterman (1989) as follows,

$$CN_{X^\top \hat{W} X} = (\lambda_{\max}/\lambda_{\min})^{1/2}.$$

Mackinnon and Puterman (1989) examined the relationship between the collinearity indices in GLMs and corresponding standard linear model as

$$\frac{w_{\min}}{w_{\max}} CN^2_{X^\top X} \leq CN^2_{X^\top W X} \leq \frac{w_{\max}}{w_{\min}} CN^2_{X^\top X}$$

where $CN_{X^\top X}$ and $CN_{X^\top W X}$ are the condition number of $X^\top X$ and $X^\top W X$, respectively, and W is a diagonal positive matrix with $w_{\min}$ and $w_{\max}$ its minimal and maximal components. The condition number of the GLMs system is bounded by the multiplies of that in the standard linear model where the multipliers are the ratios of the minimal and maximal scaling weights. Besides collinearity among the explanatory variables, Lesaffre and Marx (1993) defined the ML-collinearity, which has similar detrimental effects on the covariance matrix, e.g. inflation of some of the estimated standard errors of the regression coefficients.

Belsley et al (1980) defined an ill conditioned information matrix as one which has a small eigenvalue relative to the largest eigenvalue. Let $\lambda_0, \ldots, \lambda_q$ be the eigenvalues of information matrix in decreasing order. Marx and Smith (1990b) define a condition index as

$$CI_j = (\lambda_{\max}/\lambda_j)^{1/2}, \quad j = 0, \ldots, q.$$

Marx and Smith (1990b) emphasized that more detrimental to the GLM, than multicollinearity among the columns X, is an information matrix which is nearly deficient in rank.

3.2. Estimation in GLMs

Multicollinearity problem is very common in applied research and it may lead to some of regression coefficients being statistically insignificant and difficulties in interpreting the estimates of an individual coefficient. Several techniques in linear regression have been proposed for dealing with the problems caused by multicollinearity. The general approaches include collecting additional data, model respecification, and the use of estimation methods other than least squares that are specifically designed to combat the problems induced by multicollinearity (Montgomery et al, 2001). By analogy with the linear regression models, Mackinnon (1986), Mackinnon and Puterman (1989), Nyquist (1991) and Segerstedt (1992) stated that collinearity can be present in the GLMs and stated that collinearity has similar effects in GLMs as in linear regression models. In the literature, many studies have been conducted for linear regression models and some of these estimators are adapted to GLMs.

3.2.1. Restricted ML Estimation in GLMs

Suppose that in addition to the linear predictor, the set of q linearly independent restrictions on the parameter vector β exists: $R_t^\top \beta = r_t$, $t = 1,\ldots,q$ where R_t are $p \times 1$ vectors $R_t^\top = (R_{t1},\ldots,R_{tp})$ and r_t are scalars, both applying to known fixed numbers. We consider the estimation of β under a set of linear restrictions $R_t^\top \beta = r_t$. Nyquist (1991) considered ML estimation of the GLM under linear restrictions on the parameters. Let us denote the Lagrange multipliers $\lambda_1,\ldots,\lambda_q$ by the vector $\lambda = (\lambda_1,\ldots,\lambda_q)$. He proposed restricted ML estimation of GLMs using a quadratic penalty function

$$P_1(\beta,\lambda_1,\ldots,\lambda_q) = \sum_{i=1}^{n} \ln c(y_i,\phi_i) + \sum_{i=1}^{n} \{\theta_i y_i - b(\theta_i)\} - \frac{1}{2}\sum_{t=1}^{q} \lambda_t (r_t - R_t^\top \beta)^2$$

and find a solution to the unconstrained problem $P_1(\beta,\lambda)$ for fixed and positive values of λ_t, yielding the estimate $\hat{\beta}_r^{(m+1)}$. The restricted ML estimator is then found as

$$\hat{\beta}_r^{(m+1)} = \hat{\beta}^{(m+1)} + [(X^\top W^{(m)} X)^{-1} R^\top \left\{ R(X^\top W^{(m)} X)^{-1} R^\top \right\}^{-1}](r - R\hat{\beta}^{(m+1)})$$

where $\hat{\beta}^{(m+1)}$ is the ML estimator given by Equation (2.10.).

3.2.2. Ridge Estimation in GLMs

Considering the similarities between linear regression models, ridge estimator has been extended to GLMs by Segerstedt (1992). He considered the Lagrange function

$$P_2(\beta,k) = \beta^\top \beta - \frac{2}{k}\left[\sum_{i=1}^{n} \ln c(y_i,\phi_i) + \sum_{i=1}^{n}\{\theta_i y_i - b(\theta_i)\} + (\omega - l_{\max})\right]$$

where $2/k$ is the Lagrange multiplier, $\omega = diff(l) = l_{\max} - l(\beta)$ is the difference between the values of the log-likelihood function evaluated at the ML estimate and any estimator of the vector β. He introduces the ridge regression estimator in GLMs by solving

$$(X^\top W^{(m)} X + kI)\hat{\beta}_k^{(m+1)} = X^\top W^{(m)} z^{(m)}$$

where $k > 0$ is called the ridge parameter.

3.2.3. Restricted Ridge Estimation in GLMs

In this section, we are going to propose the restricted ridge estimator in GLMs. In this case, the log-likelihood function with respect to the penalty function $\sum_{t=1}^{q}(R_t^\top \beta - r_t)^2$ is

$$P(\beta,k,\lambda_1,\ldots,\lambda_q) = \beta^\top \beta - \frac{2}{k}\sum_{i=1}^{n} \ln c(y_i,\phi_i) + \sum_{i=1}^{n}\{\theta_i y_i - b(\theta_i)\} + \sum_{t=1}^{q}\lambda_t (R_t^\top \beta - r_t)^2$$

where $2/k$ and $\lambda_1,\ldots,\lambda_q$ are Lagrange multipliers. That is, we define an estimator for GLMs as the shortest estimator in an equivalence class of estimators with the same log-likelihood function with respect to the penalty function $\sum_{t=1}^{q}(R_t^\top\beta - r_t)^2$.

Differentiating $P(\beta,k,\lambda)$ with respect to β_j results in

$$\begin{aligned} q_j(\beta,k,\lambda) &= \frac{\partial P(\beta,k,\lambda)}{\partial\beta_j} \\ &= 2\beta_j - \frac{2}{k}\sum_{i=1}^{n}\frac{y_i-\mu_i}{\mathrm{var}(Y_i)}\frac{d\mu_i}{d\eta_i}x_{ij} + 2\sum_{t=1}^{q}R_{tj}\lambda_t(R_t^\top\beta - r_t),\ j=1,\ldots,p \end{aligned}$$

Thus, the expectations of minus times the second order derivatives are

$$\begin{aligned} s_{jv}(\beta,k,\lambda) &= \mathrm{E}\left(-\frac{\partial^2 P}{\partial\beta_j\partial\beta_v}\right) \\ &= -\left\{2\delta_{jv} + \frac{2}{k}\sum_{i=1}^{n}\frac{x_{ij}x_{iv}}{\mathrm{var}(Y_i)}\left(\frac{d\mu_i}{d\eta_i}\right)^2 + 2\sum_{t=1}^{q}\lambda_t R_{tj}R_{tv}\right\} \end{aligned}$$

where $\delta_{jv}=1$ if $j=v$ and zero otherwise.

Let $\hat{\beta}_r(k,\lambda)^{(0)}$ be the initial restricted ridge estimate of β, and let $\hat{\beta}_r(k,\lambda)^{(m)}$ denotes the estimator of β at the restricted ridge mth iteration ($m\geq 1$). Then, the Fisher scoring method for the new estimator can be written as

$$\hat{\beta}_r(k,\lambda)^{(m+1)} = \hat{\beta}_r(k,\lambda)^{(m)} + S(\beta_r^{(m)},k,\lambda)^{-1}Q(\beta_r^{(m)},k,\lambda) \tag{3.1.}$$

where $S(\beta_r^{(m)},k,\lambda)$ is the $p\times p$ matrix with elements $s_{jv}(\beta_r^{(m)},k,\lambda)$ and $Q(\beta_r^{(m)},k,\lambda)$ is the $p\times 1$ vector with elements $q_j(\beta_r^{(m)},k,\lambda)$, both treated at the estimate $\hat{\beta}_r(k,\lambda)^{(m)}$. Iteration is continued until the distance between successive estimates becomes less than some tolerance level. Multiplying both sides of Equation (3.1.) by $S(\beta_r^{(m)},k,\lambda)$ gives

$$S(\beta_r^{(m)},k,\lambda)\hat{\beta}_r(k,\lambda)^{(m+1)} = S(\beta_r^{(m)},k,\lambda)\hat{\beta}_r(k,\lambda)^{(m)} + Q(\beta_r^{(m)},k,\lambda) \tag{3.2.}$$

where

$$S(\beta_r^{(m)},k,\lambda) = -\frac{2}{k}\left\{(X^\top W^{(m)}X + kI) + R^\top\Lambda^* R\right\}, \tag{3.3.}$$

$$\begin{aligned} Q(\beta_r^{(m)},k,\lambda) &= 2\hat{\beta}_r(k,\lambda)^{(m)} - \frac{2}{k}[X^\top W^{(m)} z^{(m)} - X^\top W^{(m)} X\hat{\beta}_r(k,\lambda)^{(m)} \\ &\quad -R^\top \Lambda^* \{R\hat{\beta}_r(k,\lambda)^{(m)} - r\}], \end{aligned} \tag{3.4.}$$

$\Lambda^* = \text{diag}(k\lambda_1,\ldots,k\lambda_q)$, $R^\top = (R_1 \ldots R_q)$ and $r = (r_1 \ldots r_q)^\top$. In these equations, the weights are evaluated at $\hat{\beta}_r(k,\lambda)^{(m)}$. One important note is that $z^{(m)}$ in Equation (3.4.) is also computed by using $\hat{\beta}_r(k,\lambda)^{(m)}$.

From Equations (3.3.) and (3.4.), the right hand side of Equation (3.2.) can be written as

$$\begin{aligned} & S(\beta_r^{(m)},k,\lambda)\hat{\beta}_r(k,\lambda)^{(m)} + Q(\beta_r^{(m)},k,\lambda) \\ = & -\frac{2}{k}\left\{(X^\top W^{(m)} X + kI) + R^\top \Lambda^* R\right\}\hat{\beta}_r(k,\lambda)^{(m)} + 2\hat{\beta}_r(k,\lambda)^{(m)} \\ & -\frac{2}{k}[(X^\top W^{(m)} z^{(m)} - X^\top W^{(m)} X\hat{\beta}_r(k,\lambda)^{(m)} - R^\top \Lambda^* \left\{R\hat{\beta}_r(k,\lambda)^{(m)} - r\right\}] \\ = & -\frac{2}{k}\left(X^\top W^{(m)} z^{(m)} + R^\top \Lambda^* r\right). \end{aligned}$$

Then Equation (3.2.) simplifies to

$$\left\{(X^\top W^{(m)} X + kI) + R^\top \Lambda^* R\right\}\hat{\beta}_r(k,\lambda)^{(m+1)} = X^\top W^{(m)} z^{(m)} + R^\top \Lambda^* r.$$

Since $X^\top W^{(m)} X + kI$ is invertible and under the assumption that Λ^* is invertible, applying the inverse theorem for $\left\{(X^\top W^{(m)} X + kI) + R^\top \Lambda^* R\right\}$ yields

$$\begin{aligned} \left\{(X^\top W^{(m)} X + kI) + R^\top \Lambda^* R\right\}^{-1} &= (X^\top W^{(m)} X + kI)^{-1} - (X^\top W^{(m)} X + kI)^{-1} \\ &\quad \times R^\top \left\{(\Lambda^*)^{-1} + R(X^\top W^{(m)} X + kI)^{-1} R^\top\right\}^{-1} \\ &\quad \times R(X^\top W^{(m)} X + kI)^{-1}. \end{aligned}$$

$\hat{\beta}_r(k)^{(m+1)}$ can be written as

$$\begin{aligned} \hat{\beta}_r(k,\lambda)^{(m+1)} &= [(X^\top W^{(m)} X + kI)^{-1} - (X^\top W^{(m)} X + kI)^{-1} R^\top \\ &\quad \times \left\{(\Lambda^*)^{-1} + R(X^\top W^{(m)} X + kI)^{-1} R^\top\right\}^{-1} \\ &\quad \times R(X^\top W^{(m)} X + kI)^{-1}] \left(X^\top W^{(m)} z^{(m)} + R^\top \Lambda^* r\right). \end{aligned}$$

By using the result

$$
\begin{aligned}
&[(X^\top W^{(m)}X+kI)^{-1}-(X^\top W^{(m)}X+kI)^{-1}R^\top\{(\Lambda^*)^{-1}\\
&+R(X^\top W^{(m)}X+kI)^{-1}R^\top\}^{-1}R(X^\top W^{(m)}X+kI)^{-1}]R^\top\Lambda^* r\\
&=(X^\top W^{(m)}X+kI)^{-1}R^\top[I-\left\{(\Lambda^*)^{-1}+R(X^\top W^{(m)}X+kI)^{-1}R^\top\right\}^{-1}\\
&\times R(X^\top W^{(m)}X+kI)^{-1}R^\top]\Lambda^* r\\
&=(X^\top W^{(m)}X+kI)^{-1}R^\top\left\{(\Lambda^*)^{-1}+R(X^\top W^{(m)}X+kI)^{-1}R^\top\right\}^{-1}\{(\Lambda^*)^{-1}\\
&+R(X^\top W^{(m)}X+kI)^{-1}R^\top-R(X^\top W^{(m)}X+kI)^{-1}R^\top\}\Lambda^* r\\
&=(X^\top W^{(m)}X+kI)^{-1}R^\top\left\{(\Lambda^*)^{-1}+R(X^\top W^{(m)}X+kI)^{-1}R^\top\right\}^{-1}r,
\end{aligned}
$$

we get

$$
\begin{aligned}
\hat{\beta}_r(k,\lambda)^{(m+1)} &= (X^\top W^{(m)}X+kI)^{-1}X^\top W^{(m)}z^{(m)}-(X^\top W^{(m)}X+kI)^{-1}\\
&\times R^\top\left\{(\Lambda^*)^{-1}+R(X^\top W^{(m)}X+kI)^{-1}R^\top\right\}^{-1}\\
&\times R(X^\top W^{(m)}X+kI)^{-1}X^\top W^{(m)}z^{(m)}+(X^\top W^{(m)}X+kI)^{-1}\\
&\times R^\top\left\{(\Lambda^*)^{-1}+R(X^\top W^{(m)}X+kI)^{-1}R^\top\right\}^{-1}r.
\end{aligned}
$$

After necessary arrangements, we obtain $\hat{\beta}_r(k,\lambda)^{(m+1)}$ as

$$
\begin{aligned}
\hat{\beta}_r(k,\lambda)^{(m+1)} &= (X^\top W^{(m)}X+kI)^{-1}X^\top W^{(m)}z^{(m)}+[(X^\top W^{(m)}X+kI)^{-1}\\
&\times R^\top\left\{(\Lambda^*)^{-1}+R(X^\top W^{(m)}X+kI)^{-1}R^\top\right\}^{-1}]\\
&\times\left\{r-R(X^\top W^{(m)}X+kI)^{-1}X^\top W^{(m)}z^{(m)}\right\}. \qquad (3.5.)
\end{aligned}
$$

Since $(X^\top W^{(m)}X+kI)^{-1}X^\top W^{(m)}z^{(m)}$ is similar to the ridge estimator proposed by Segerstedt (1992) and $\hat{\beta}_r(k,\lambda)^{(m+1)}$ in form is similar to the restricted ridge estimator of Groß (2003) in linear models, we call the estimator $\hat{\beta}_r(k,\lambda)^{(m+1)}$ as the restricted ridge estimator in GLMs. The $(m+1)$th approximation $\hat{\beta}_r(k)^{(m+1)}$ of the restricted

ridge estimator is finally obtained as

$$\begin{aligned}\hat{\beta}_r(k)^{(m+1)} &= \lim_{\lambda_1,\ldots,\lambda_q\to\infty}\hat{\beta}_r(k,\lambda)^{(m+1)} \\ &= (X^\top W^{(m)}X+kI)^{-1}X^\top W^{(m)}z^{(m)}+[(X^\top W^{(m)}X+kI)^{-1}R^\top \\ &\quad\times\left\{R(X^\top W^{(m)}X+kI)^{-1}R^\top\right\}^{-1}] \\ &\quad\times\left\{r-R(X^\top W^{(m)}X+kI)^{-1}X^\top W^{(m)}z^{(m)}\right\}. \qquad (3.6.)\end{aligned}$$

As k approaches zero, $\hat{\beta}_r(k)^{(m+1)}$ approaches the restricted ML estimation in GLMs given by Nyquist (1991)

$$\begin{aligned}\hat{\beta}_r^{(m+1)} &= \lim_{k\to 0}\hat{\beta}_r(k)^{(m+1)} \\ &= \hat{\beta}^{(m+1)}+[(X^\top W^{(m)}X)^{-1}R^\top\left\{R(X^\top W^{(m)}X)^{-1}R^\top\right\}^{-1}](r-R\hat{\beta}^{(m+1)})\end{aligned}$$

where $\hat{\beta}^{(m+1)}=(X^\top W^{(m)}X)^{-1}X^\top W^{(m)}z^{(m)}$ is in the form of the IRLS estimator with weights evaluated at $\hat{\beta}_r^{(m)}$.

If the successive estimates $\hat{\beta}_r(k)^{(m+1)}$ and $\hat{\beta}_r^{(m+1)}$ converge to, say $\hat{\beta}_r(k)$ and $\hat{\beta}_r$ respectively, as m tends to infinity we obtain

$$\begin{aligned}\hat{\beta}_r(k) &= (X^\top\hat{W}_kX+kI)^{-1}X^\top\hat{W}_k\hat{z}_k+(X^\top\hat{W}_kX+kI)^{-1}R^\top\left\{R(X^\top\hat{W}_kX+kI)^{-1}R^\top\right\}^{-1} \\ &\quad\times\left\{r-R(X^\top\hat{W}_kX+kI)^{-1}X^\top\hat{W}_k\hat{z}_k\right\} \qquad (3.7.)\end{aligned}$$

and

$$\begin{aligned}\hat{\beta}_r &= (X^\top\hat{W}_rX)^{-1}X^\top\hat{W}_r\hat{z}_r+(X^\top\hat{W}_rX)^{-1}R^\top\left\{R(X^\top\hat{W}_rX)^{-1}R^\top\right\}^{-1} \\ &\quad\times\left\{r-R(X^\top\hat{W}_rX)^{-1}X^\top\hat{W}_r\hat{z}_r\right\}\end{aligned}$$

where $\hat{W}_k$ and $\hat{W}_r$ are evaluated at the final iteration corresponding to restricted ridge and restricted ML estimators and $\hat{z}_k$ and $\hat{z}_r$ are the estimated working-responses.

3.2.4. Liu Estimation in GLMs

Our aim in this section is to propose Liu estimator in GLMs. Similar to the linear regression case, we define the Liu estimator for GLMs. We consider the quadratic penalty function

$$\begin{aligned} F(\beta,d) &= l(\beta)-\frac{1}{2a(\phi)}(\beta-d\hat{\beta})^{\top}(\beta-d\hat{\beta}) \\ &= \left[\sum_{i=1}^{n}\frac{\theta_i y_i-b(\theta_i)}{a(\phi_i)}+c(y_i,\phi_i)\right]-\frac{1}{2a(\phi)}(\beta-d\hat{\beta})^{\top}(\beta-d\hat{\beta}) \end{aligned}$$

where $0<d<1$ is the Liu-biasing parameter and $\hat{\beta}$ is the IRLS estimator at the final iteration. The function $F(\beta,d)$ results in an estimator with an equivalence class of estimators of β which are equal distance from $d\hat{\beta}$. Differentiating $F(\beta,d)$ with respect to β_j gives

$$Q(\beta,d)=\frac{1}{a(\phi)}[X^{\top}W^{-1}D(y-\mu)+d\hat{\beta}-\beta]$$

where $Q(\beta,d)$ is the $q\times 1$ vector with elements

$$q_j(\beta,d)=\frac{\partial F(\beta,d)}{\partial\beta_j}=\frac{1}{a(\phi)}\left[\sum_{i=1}^{n}\frac{y_i-\mu_i}{w_i}g'(\mu_i)x_{ij}+d\hat{\beta}-\beta_j\right],\quad j=1,\ldots,q.$$

Then, we get

$$\frac{\partial^2 F(\beta,d)}{\partial\beta_j\partial\beta_k}=\frac{1}{a(\phi)}\left[\sum_{i=1}^{n}(y_i-\mu_i)\frac{\partial}{\partial\beta_k}\left(\frac{g'(\mu_i)}{w_i}x_{ij}\right)-\sum_{i=1}^{n}\frac{f_j(x_i)}{w_i}x_{ik}-\delta_{jk}\right] \qquad (3.8.)$$

where $\delta_{jk}=1$ if $j=k$ and zero otherwise.

Taking the expected value of both sides of Equation (3.8.),

$$s_{jk}(\beta,d)=\mathrm{E}\left[\frac{\partial^2 F(\beta,d)}{\partial\beta_j\partial\beta_k}\right]=-\frac{1}{a(\phi)}\left[\sum_{i=1}^{n}\frac{1}{w_i}x_{ij}x_{ik}+\delta_{jk}\right],$$

the expected value of $H_l(\beta,d)=\partial^2 F(\beta,d)/\partial\beta\partial\beta'$ is obtained as

$$\mathrm{E}[H_l(\beta,d)]=-\frac{1}{a(\phi)}(X^{\top}W^{-1}X+I).$$

Application of the Fisher scoring method in this case yields

$$\begin{aligned}\hat{\beta}_d^{(m+1)} &= \hat{\beta}_d^{(m)} - \left\{\mathrm{E}[H_l(\beta,d)]_{\beta=\hat{\beta}_d^{(m)}}\right\}^{-1}\left[\frac{\partial F(\beta,d)}{\partial\beta}\right]_{\beta=\hat{\beta}_d^{(m)}} \\ &= \hat{\beta}_d^{(m)} + \{(X^\top WX+I)^{-1}[X^\top WD(y-\mu)+d\hat{\beta}-\beta]\}_{\beta=\hat{\beta}_d^{(m)}}. \quad (3.9.)\end{aligned}$$

The estimator in Equation (3.9.) is calculated iteratively and it depends on the IRLS estimator at the final iteration $\hat{\beta}$. To obtain the same weight matrix W and working response z with $\hat{\beta}$, the estimator in the first step can be written as

$$\hat{\beta}_d^{(1)} = \beta^{(0)} + \{(X^\top WX+I)^{-1}[X^\top WD(y-\mu)+d\hat{\beta}^{(1)}-\beta]\}_{\beta=\beta^{(0)}} \quad (3.10.)$$

where $\hat{\beta}^{(1)}$ is the IRLS at the first step written as

$$\hat{\beta}^{(1)} = \beta^{(0)} + [(X^\top WX)^{-1}X^\top WD(y-\mu)]_{\beta=\beta^{(0)}}. \quad (3.11.)$$

Thus, the weight matrix at $\hat{\beta}_d^{(1)}$ and $\hat{\beta}^{(1)}$ are calculated in the same $\beta^{(0)}$ initial value or real parameter (if it is known).

After writing Equation (3.11.) in Equation (3.10.) and rearranging the expression Equation (3.12.), we get

$$\begin{aligned}\hat{\beta}_d^{(1)} &= \beta^{(0)} + \{(X^\top WX+I)^{-1}[X^\top WD(y-\mu) \\ &\quad +d[\beta+(X^\top WX)^{-1}[X^\top WD(y-\mu)]]-\beta]\}_{\beta=\beta^{(0)}} \\ &= \{(X^\top WX+I)^{-1}[(X^\top WX+I)\beta+X^\top WD(y-\mu)+d\beta \\ &\quad +d(X^\top WX)^{-1}[X^\top WD(y-\mu)]-\beta]\}_{\beta=\beta^{(0)}} \\ &= \{(X^\top WX+I)^{-1}[(X^\top WX+dI)\beta \\ &\quad +(X^\top WX+dI)(X^\top WX)^{-1}[X^\top WD(y-\mu)]]\}_{\beta=\beta^{(0)}} \\ &= \{(X^\top WX+I)^{-1}(X^\top WX+dI)[\beta+(X^\top WX)^{-1}[X^\top WD(y-\mu)]]\}_{\beta=\beta^{(0)}} \\ &= (X^\top\hat{W}^{(0)}X+I)^{-1}(X^\top\hat{W}^{(0)}X+dI)\hat{\beta}^{(1)} \quad (3.12.)\end{aligned}$$

where $\hat{W}^{(0)}$ is the weight matrix evaluated at $\beta^{(0)}$.

Carrying the idea of linear regression to GLMs, a first-order approximated general Liu estimator $\hat{\beta}_d^{(1)}$ can be defined as

$$\hat{\beta}_{\text{Đ}}^{(1)} = (X^\top \hat{W}^{(0)} X + I)^{-1}(X^\top \hat{W}^{(0)} X + \text{Đ})\hat{\beta}^{(1)} \tag{3.13.}$$

where Đ$= diag(d_i)$ and $0 < d_i < 1$, $i = 1, \ldots, p$ are the Liu-biasing parameters.

3.2.5. Restricted Liu Estimation in GLMs

It is the objective of this section to present a restricted Liu estimator in GLMs. Both iterative and first-order approximated restricted Liu estimators are developed using quadratic penalized likelihood. Because there are different estimators used in the case of multicollinearity in linear regression, these estimators can also be defined for GLMs.

We suppose that in addition to the linear predictor, the set of q linearly independent restrictions on the parameter vector β exist: $R_t^\top \beta = r_t$, $t = 1, \ldots, q$, where $R_t^\top = [R_{t_1}, \ldots, R_{t_p}]$ are $p \times 1$ vectors and r_t are scalars. We define the restricted Liu estimator for GLMs with the log-likelihood function in Equation (2.4.) with respect to quadratic penalty function $\sum_{t=1}^{q} (R_t^\top \beta - r_t)^2$. The objective function is

$$\begin{aligned} \psi(\beta, d, \lambda) &= \left[\sum_{i=1}^{n} \frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi_i)\right] \\ &\quad - \frac{1}{2a(\phi)}(\beta - d\hat{\beta})^\top (\beta - d\hat{\beta}) - \frac{1}{2a(\phi)} \sum_{t=1}^{q} \lambda_t (R_t^\top \beta - r_t)^2 \end{aligned}$$

where $0 < d < 1$ is Liu-biasing parameter, $\lambda = (\lambda_1, \ldots, \lambda_q)$ is the vector of Lagrange multipliers and $\hat{\beta}$ is the IRLS estimator evaluated at the final iteration.

Differentiating $\psi(\beta, d, \lambda)$ with respect to β and making use of Fisher's

method of scoring, we get

$$\begin{aligned} &Q(\beta,d,\lambda)|_{\beta=\hat{\beta}_r(d)^{(m)}} = \left[\frac{\partial\psi(\beta,d,\lambda)}{\partial\beta}\right]_{\beta=\hat{\beta}_r(d)^{(m)}} \\ &= \frac{1}{a(\phi)}[X^\top WD(y-\mu)-\beta+d\hat{\beta}-R^\top\Lambda^*(R\beta-r)]_{\beta=\hat{\beta}_r(d)^{(m)}} \end{aligned} \tag{3.14.}$$

where $\Lambda^* = diag(\lambda_1,\ldots,\lambda_q)$, $R^\top = (R_1 \ldots R_q)$, $r = (r_1 \ldots r_q)^\top$ and $W = diag\{w_{ii}\}$ is the $n \times n$ diagonal matrix with weights $w_{ii} = \{1/\mathrm{var}(Y_i)\}\,(d\mu_i/d\eta_i)^2$. $Q(\beta,d,\lambda)$ is the $q \times 1$ vector with elements

$$\begin{aligned} &q_j(\beta,d,\lambda) = \frac{\partial\psi(\beta,d,\lambda)}{\partial\beta_j} \\ &= \frac{1}{a(\phi)}[\sum_{i=1}^{n}\frac{y_i-\mu_i}{w_i}g'(\mu_i)x_{ij}-\beta_j+d\hat{\beta}_j-\sum_{t=1}^{q}R_{tj}\lambda_t(R_t^\top\beta-r_t)] \end{aligned} \tag{3.15.}$$

where $j = 1,\ldots,q$. The (j,v)th element of the Hessian matrix $\mathrm{H}_l(\beta,d,\lambda) = \partial^2\psi(\beta,d,\lambda)/\partial\beta\partial\beta'$ is

$$\begin{aligned} \frac{\partial^2\psi(\beta,d,\lambda)}{\partial\beta_j\partial\beta_v} &= \frac{1}{a(\phi)}[\sum_{i=1}^{n}(y_i-\mu_i)\frac{\partial}{\partial\beta_v}\left(\frac{g'(\mu_i)}{w_i}x_{ij}\right) \\ &\quad -\sum_{i=1}^{n}\frac{x_{ij}}{w_i}x_{iv}-\delta_{jv}-\sum_{t=1}^{q}\lambda_t R_{tj}R_{tv}] \end{aligned} \tag{3.16.}$$

where $\delta_{jv} = 1$ if $j = v$ and zero otherwise.

Taking the expected value of both sides of Equation (3.16.), we get

$$\begin{aligned} s_{jv}(\beta,d,\lambda) &= -E\left[\frac{\partial^2\psi(\beta,d,\lambda)}{\partial\beta_j\partial\beta_v}\right] \\ &= \frac{1}{a(\phi)}\left[\sum_{i=1}^{n}\frac{1}{w_i}x_{ij}x_{iv}+\delta_{jv}+\sum_{t=1}^{q}\lambda_t R_{tj}R_{tv}\right]. \end{aligned}$$

Consequently, the expected value of $\mathrm{H}_l(\beta,d,\lambda)$ is of the form

$$\mathrm{E}[\mathrm{H}_l(\beta,d,\lambda)] = \frac{1}{a(\phi)}[(X^\top WX+I)+R^\top\Lambda^*R]. \tag{3.17.}$$

Assuming that Λ^* and $X^\top WX+I$ are invertible, application of inverse theorem (see Harville, 2008, p.428), we obtain

$$\left\{\mathrm{E}[\mathrm{H}_l(\beta,d,\lambda)]_{\beta=\hat{\beta}_r(d)^{(m)}}\right\}^{-1} = a(\phi)\{(X^\top WX+I)^{-1} - (X^\top WX+I)^{-1}R^\top \times[(\Lambda^*)^{-1}+R(X^\top WX+I)^{-1}R^\top]^{-1}R(X^\top WX+I)^{-1}\}. \tag{3.18.}$$

By using Equations (3.14.), (3.17.) and (3.18.), we get

$$\begin{aligned}&\mathrm{E}[\mathrm{H}_l(\beta,d,\lambda)]_{\beta=\hat{\beta}_r(d)^{(m)}}\hat{\beta}_r(d)^{(m)} + \left[\frac{\partial\psi(\beta,d,\lambda)}{\partial\beta}\right]_{\beta=\hat{\beta}_r(d)^{(m)}}\\ &= \frac{1}{a(\phi)}[(X^\top WX+I)\hat{\beta}_r(d)^{(m)}\\ &+X^\top WD(y-\mu)-\hat{\beta}_r(d)^{(m)}+d\hat{\beta}+R^\top\Lambda^* r].\end{aligned} \tag{3.19.}$$

Following Equations (3.18.) and (3.19.), $\hat{\beta}_r(d)^{(m+1)}$ simplifies to

$$\begin{aligned}\hat{\beta}_r(d)^{(m+1)} &= \{(X^\top WX+I)^{-1}-(X^\top WX+I)^{-1}R^\top\\ &\times[(\Lambda^*)^{-1}+R(X^\top WX+I)^{-1}R^\top]^{-1}R(X^\top WX+I)^{-1}\}\\ &\times[(X^\top WX+I)\hat{\beta}_r(d)^{(m)}+X^\top WD(y-\mu)-\hat{\beta}_r(d)^{(m)}\\ &+d\hat{\beta}+R^\top\Lambda^* r][(X^\top WX+I)^{-1}R^\top\Lambda^* r].\end{aligned}$$

After multiplying the terms and by using the following equation $\hat{\beta}(d)^{(m+1)} = (X^\top WX+I)^{-1}[(X^\top WX+I)\hat{\beta}_r(d)^{(m)}+X^\top WD(y-\mu)-\hat{\beta}_r(d)^{(m)}+d\hat{\beta}]$, we get, after algebric simplifications,

$$\begin{aligned}\hat{\beta}_r(d)^{(m+1)} &= \hat{\beta}(d)^{(m+1)}-(X^\top WX+I)^{-1}R^\top[(\Lambda^*)^{-1}+R(X^\top WX+I)^{-1}\\ &\times R^\top]^{-1}R\hat{\beta}(d)^{(m+1)}+[I-(X^\top WX+I)^{-1}R[(\Lambda^*)^{-1}\\ &+R(X^\top WX+I)^{-1}R^\top]^{-1}R]\\ &= \hat{\beta}(d)^{(m+1)}-(X^\top WX+I)^{-1}R^\top[(\Lambda^*)^{-1}+R(X^\top WX+I)^{-1}\\ &\times R^\top]^{-1}R\hat{\beta}(d)^{(m+1)}+[(X^\top WX+I)+R^\top\Lambda^* r]^{-1}R^\top\Lambda^* r.\end{aligned}$$

For further simplifications and from Equation (3.18.), we obtain

$$
\begin{aligned}
&[(X^\top WX+I)+R^\top\Lambda^* r]^{-1}R^\top\Lambda^* = (X^\top WX+I)^{-1}R^\top\Lambda^* - (X^\top WX+I)^{-1}R^\top \\
&\times[(\Lambda^*)^{-1}+R(X^\top WX+I)^{-1}R^\top]^{-1}R(X^\top WX+I)^{-1}R^\top\Lambda^* \\
&= (X^\top WX+I)^{-1}R^\top[(\Lambda^*)^{-1}+R(X^\top WX+I)^{-1}R^\top]^{-1} \\
&\times\{[(\Lambda^*)^{-1}+R(X^\top WX+I)^{-1}R^\top]-R(X^\top WX+I)^{-1}R^\top\}\Lambda^* \\
&= (X^\top WX+I)^{-1}R^\top[(\Lambda^*)^{-1}+R(X^\top WX+I)^{-1}R^\top]^{-1}(\Lambda^*)^{-1}\Lambda^*.
\end{aligned}
$$

Finally, $\hat{\beta}_r(d)^{(m+1)}$ results in

$$
\begin{aligned}
\hat{\beta}_r(d)^{(m+1)} &= \hat{\beta}(d)^{(m+1)} - (X^\top WX+I)^{-1}R^\top[(\Lambda^*)^{-1}+R(X^\top WX+I)^{-1} \\
&\quad \times R^\top]^{-1}R\hat{\beta}(d)^{(m+1)} + (X^\top WX+I)^{-1}R^\top[(\Lambda^*)^{-1} \\
&\quad +R(X^\top WX+I)^{-1}R^\top]^{-1}r \\
&= \hat{\beta}(d)^{(m+1)} - (X^\top WX+I)^{-1}R^\top[(\Lambda^*)^{-1}+R(X^\top WX+I)^{-1}R^\top]^{-1} \\
&\quad \times[R\hat{\beta}(d)^{(m+1)}-r].
\end{aligned}
$$

Application of the Fisher scoring method yields

$$
\begin{aligned}
\hat{\beta}_r(d)^{(m+1)} &= \hat{\beta}_r(d)^{(m)} + \left\{\mathrm{E}[\mathrm{H}_l(\beta,d,\lambda)]_{\beta=\hat{\beta}_r(d)^{(m)}}\right\}^{-1}\left[\frac{\partial\psi(\beta,d,\lambda)}{\partial\beta}\right]_{\beta=\hat{\beta}_r(d)^{(m)}} \\
&= \hat{\beta}_d^{(m+1)} - (X^\top W^{(m)}X+I)^{-1}R^\top[(\Lambda^*)^{-1} \\
&\quad +R(X^\top W^{(m)}X+I)^{-1}R^\top]^{-1}(R\hat{\beta}_d^{(m+1)}-r) \qquad (3.20.)
\end{aligned}
$$

where $\hat{\beta}_d^{(m+1)} = (X^\top W^{(m)}X+I)^{-1}(X^\top W^{(m)}X+dI)\hat{\beta}^{(m+1)}$ is the Liu estimator and $\hat{\beta}^{(m+1)}$ is in the form of the ML estimator given by Equation (2.10.) in GLMs and both $\hat{\beta}_d^{(m+1)}$ and $\hat{\beta}^{(m+1)}$ are evaluated at the iteration algorithm.

The $(m+1)$th approximation of the restricted Liu estimator is finally obtained as

$$\begin{aligned}\hat{\beta}_r(d)^{(m+1)} &= \lim_{\lambda_1,\ldots,\lambda_q\to\infty}\hat{\beta}_r(d)^{(m+1)} \\ &= \hat{\beta}_d^{(m+1)} - (X^\top W^{(m)}X+I)^{-1}R^\top[R(X^\top W^{(m)}X+I)^{-1}R^\top]^{-1} \\ &\quad \times(R\hat{\beta}_d^{(m+1)}-r). \end{aligned} \tag{3.21.}$$

The estimator in Equation (3.21.) is calculated iteratively. The estimator in the first step can be written as

$$\begin{aligned}\hat{\beta}_r(d)^{(1)} &= \hat{\beta}_d^{(1)} - (X^\top W^{(0)}X+I)^{-1}R^\top \\ &\quad \times[R(X^\top W^{(0)}X+I)^{-1}R^\top]^{-1}(R\hat{\beta}_d^{(1)}-r) \end{aligned} \tag{3.22.}$$

where $\hat{\beta}_d^{(1)} = (X^\top W^{(0)}X+I)^{-1}(X^\top W^{(0)}X+dI)\hat{\beta}^{(1)}$ is the Liu estimator at the first step and $\hat{\beta}^{(1)} = (X^\top W^{(0)}X)^{-1}X^\top W^{(0)}z^{(0)}$ is the IRLS estimator at the first step written where $z^{(0)} = X\beta^{(0)} + [D(y-\mu)]_{\beta=\beta^{(0)}}$ is the working response.

As special case,

$$\begin{aligned}\hat{\beta}_r^{(1)} &= \lim_{d\to 0}\hat{\beta}_r(d)^{(m+1)} \\ &= \hat{\beta}^{(1)} + (X^\top W^{(0)}X)^{-1}R^\top[R(X^\top W^{(0)}X)^{-1}R^\top]^{-1}(R\hat{\beta}^{(1)}-r)\end{aligned}$$

is the restricted ML estimator which is introduced by Nyquist (1991). Thus, the weight matrix and the working response of $\hat{\beta}_r(d)^{(1)}$, $\hat{\beta}_d^{(1)}$, $\hat{\beta}^{(1)}$ and $\hat{\beta}_r^{(1)}$ are calculated in the same $\beta^{(0)}$ initial value or real parameter and all these four estimators have the same working response. $\hat{\beta}^{(1)}$, $\hat{\beta}_r^{(1)}$, $\hat{\beta}_d^{(1)}$, and $\hat{\beta}_r(d)^{(1)}$ can be called respectively as the first-order approximated ML, first-order approximated restricted ML, first-order approximated Liu, and first-order approximated restricted Liu estimators.

3.3. Comparisons of the Estimators

To compare the estimators of the unknown coefficient vector β, a criterion for measuring the performance of an estimator is needed. Widely used method for

measuring the performance of an estimator of β is the matrix MSE because it includes all pertinent information about the feature of an estimator. Let $\tilde{\beta}$ be an estimator of β, then the MSE of $\tilde{\beta}$ is defined as

$$\text{MSE}(\tilde{\beta}) = \text{Var}(\tilde{\beta}) + [\text{Bias}(\tilde{\beta})][\text{Bias}(\tilde{\beta})]^{\top} \tag{3.23.}$$

where $\text{Var}(\tilde{\beta})$ and $\text{Bias}(\tilde{\beta})$ are the variance-covariance matrix and bias vector of $\tilde{\beta}$, respectively. Applying the trace operator, we have the scalar mean square error (sMSE): $\text{sMSE}(\tilde{\beta}) = tr[\text{MSE}(\tilde{\beta})]$.

Let two estimators $\tilde{\beta}_1$ and $\tilde{\beta}_2$ of β are given, the estimator $\tilde{\beta}_2$ is said to be superior to $\tilde{\beta}_1$ in the sense of MSE criterion if and only if the difference $\text{MSE}(\tilde{\beta}_1) - \text{MSE}(\tilde{\beta}_2)$ is a nonnegative definite (nnd) matrix. The estimator $\tilde{\beta}_2$ dominates the estimator $\tilde{\beta}_1$ in the sense of sMSE criterion, if and only if $\text{sMSE}(\tilde{\beta}_1) \geq \text{sMSE}(\tilde{\beta}_2)$.

3.3.1. MSE of the Restricted Ridge Estimator

Mean squared error properties of restricted ridge estimator at convergence in GLMs is going to be investigated. Equation (3.7.) can also be written as

$$\hat{\beta}_r(k) = S_k^{-1}R^{\top}(RS_k^{-1}R^{\top})^{-1}r + M_k X^{\top}\hat{W}\hat{z} \tag{3.24.}$$

where $M_k = S_k^{-1} - S_k^{-1}R^{\top}(RS_k^{-1}R^{\top})^{-1}RS_k^{-1}$ and $S_k = X^{\top}\hat{W}X + kI$.

By applying the idea of Groß (2003) to M_k, the matrix M_k in GLMs can be written as

$$M_k = (QS_kQ)^{+} = V \begin{pmatrix} (\Lambda_{p-q} + kI_{p-q})^{-1} & 0 \\ 0 & 0 \end{pmatrix} V^{\top}$$

where $Q = I_p - R^{\top}(RR^{\top})^{-1}R$ and "+" superscript denotes the Moore-Penrose inverse of QS_kQ. $\Lambda_{p-q} = \text{diag}(t_1, \ldots, t_{p-q})$ is a diagonal matrix including the $p-q$ nonzero eigenvalues of $QX^{\top}\hat{W}XQ$ on its diagonal. V is an orthogonal matrix with columns that constitute the eigenvectors of $QX^{\top}\hat{W}XQ$.

Using Equation (3.23.), the variance-covariance matrix and the bias vector are obtained asymptotically as $\text{Var}\left\{\hat{\beta}_r(k)\right\} = M_k X^\top \hat{W} X M_k$ and

$$\text{Bias}\left\{\hat{\beta}_r(k)\right\} = -kM_k\beta - S_k^{-1}R^\top(RS_k^{-1}R^\top)^{-1}(R\beta - r).$$

Then, the matrix MSE of $\hat{\beta}_r(k)$ is obtained asymptotically as

$$\begin{aligned}
\text{MSE}\left\{\hat{\beta}_r(k)\right\} &= \text{Var}\left\{\hat{\beta}_r(k)\right\} + \text{Bias}\left\{\hat{\beta}_r(k)\right\} Bias\left\{\hat{\beta}_r(k)\right\}^\top \\
&= M_k X^\top \hat{W} X M_k + \left\{-kM_k\beta - S_k^{-1}R^\top(RS_k^{-1}R^\top)^{-1}(R\beta - r)\right\} \\
&\quad \times \left\{-kM_k\beta - S_k^{-1}R^\top(RS_k^{-1}R^\top)^{-1}(R\beta - r)\right\}^\top.
\end{aligned}$$

We examine the MSE properties of $\hat{\beta}_r(k)$ when the restrictions hold true. In this case, the matrix MSE of $\hat{\beta}_r(k)$ reduces to

$$\text{MSE}\left\{\hat{\beta}_r(k)\right\} = M_k X^\top \hat{W} X M_k + (-kM_k\beta)(-kM_k\beta)^\top.$$

Since $M_k X^\top \hat{W} X M_k = M_k - kM_k^2$, the asymptotic variance-covariance matrix and the squared bias of $\hat{\beta}_r(k)$ can equivalently be written as

$$\begin{aligned}
\text{Var}\left\{\hat{\beta}_r(k)\right\} &= M_k X^\top \hat{W} X M_k \\
&= V \begin{pmatrix} (\Lambda_{p-q} + kI_{p-q})^{-1}\Lambda_{p-q}(\Lambda_{p-q} + kI_{p-q})^{-1} & 0 \\ 0 & 0 \end{pmatrix} V^\top \\
&= \sum_{j=1}^{p-q} \frac{t_j}{(t_j + k)^2},
\end{aligned}$$

$$\begin{aligned}
\text{Bias}\left\{\hat{\beta}_r(k)\right\}^\top \text{Bias}\left\{\hat{\beta}_r(k)\right\} &= (-kM_k\beta)^\top(-kM_k\beta) \\
&= \alpha^\top \begin{pmatrix} k^2(\Lambda_{p-q} + kI_{p-q})^{-1}(\Lambda_{p-q} + kI_{p-q})^{-1} & 0 \\ 0 & 0 \end{pmatrix} \alpha \\
&= \sum_{j=1}^{p-q} \frac{k^2\alpha_j^2}{(t_j + k)^2}
\end{aligned}$$

where $\alpha = V^\top \beta$. So, the asymptotic sMSE value of $\hat{\beta}_r(k)$ is written as

$$\text{sMSE}\left\{\hat{\beta}_r(k)\right\} = \sum_{j=1}^{p-q} \frac{t_j}{(t_j+k)^2} + \sum_{j=1}^{p-q} \frac{k^2\alpha_j^2}{(t_j+k)^2} = \gamma_1(k) + \gamma_2(k)$$

where the first term $\gamma_1(k)$ is the sum of the variance and second term $\gamma_2(k)$ is the square of bias of the parameter estimates.

In the following theorems, the MSE properties of the restricted ridge estimator in GLMs are derived. When the restrictions hold true, it is obvious that these theorems also hold for the ridge estimator in linear regression given by Hoerl and Kennard (1970). However, due to the restrictions, j goes from one to $p-q$. In this thesis, we give following Theorems.

Theorem 3.1 For all $k > 0$, the total variance $\gamma_1(k)$ is a continuous and monotonic decreasing function of k.

Proof: The first derivative with respect to k equals

$$\frac{d\gamma_1(k)}{dk} = -2\sum_{j=1}^{p-q} \frac{t_j}{(t_j+k)^3}. \tag{3.25.}$$

According to Equation (3.25.), one can conclude that $\gamma_1(k)$ is a continuous decreasing function of k since $t_j > 0$ for $j = 1,\ldots,p-q$ and $k > 0$.

Theorem 3.2 For all $k > 0$, the squared bias $\gamma_2(k)$ is a continuous and monotonic increasing function of k.

Proof: The first derivative with respect to k equals

$$\frac{d\gamma_2(k)}{dk} = 2k\sum_{j=1}^{p-q} \frac{\alpha_j^2 t_j}{(t_j+k)^3}. \tag{3.26.}$$

Since $t_j > 0$ for $j = 1,\ldots,p-q$ and $k > 0$, Equation (3.26.) shows that $\gamma_2(k)$ is a continuous, monotonically increasing function of k.

To compare the estimators, the weight matrix and the working response for each estimator have to be same. Thus, we can compare the first-order approximated restricted ridge estimator to the first-order approximated ML, restricted ML and ridge estimators according to the MSE criterion since the weight matrices and the working responses are evaluated at the initial value $\hat{\beta}^{(0)}$.

We compute the approximated biasing vector and variance-covariance matrix of the estimator $\hat{\beta}_r(k)^{(1)}$ as

$$\text{Bias}[\hat{\beta}_r(k)^{(1)}] = -kM_k^{(0)}\beta - (S_k^{(0)})^{-1}R^{\top}(R(S_k^{(0)})^{-1}R^{\top})^{-1}(R\beta - r) \tag{3.27.}$$

and

$$\text{Var}[\hat{\beta}_r(k)^{(1)}] = M_k^{(0)}S^{(0)}M_k^{(0)} \tag{3.28.}$$

where $M_k^{(0)} = (S_k^{(0)})^{-1} - (S_k^{(0)})^{-1}R^{\top}(R(S_k^{(0)})^{-1}R^{\top})^{-1}R(S_k^{(0)})^{-1}$, $S_k^{(0)} = X^{\top}\hat{W}^{(0)}X + kI$ and $S^{(0)} = X^{\top}\hat{W}^{(0)}X$.

Combining Equations (3.27.)-(3.28.), the matrix MSE of $\hat{\beta}_r(k)^{(1)}$ in GLMs is found as

$$\text{MSE}[\hat{\beta}_r(k)^{(1)}] = M_k^{(0)}S^{(0)}M_k^{(0)} + (-kM_k^{(0)}\beta - P_k(R\beta - r))(-kM_k^{(0)}\beta - P_k(R\beta - r))^{\top}$$

where $P_k = (S_k^{(0)})^{-1}R^{\top}(R(S_k^{(0)})^{-1}R^{\top})^{-1}$.
MSE matrices of $\hat{\beta}^{(1)}$, $\hat{\beta}_r^{(1)}$ and $\hat{\beta}_k^{(1)}$ can be written respectively as

$$\text{MSE}[\hat{\beta}^{(1)}] = S^{(0)},$$

$$\text{MSE}[\hat{\beta}_r^{(1)}] = M_0^{(0)}S^{(0)}M_0^{(0)} + P_0(R\beta - r)(R\beta - r)^{\top}P_0{}^{\top},$$

$$\text{MSE}[\hat{\beta}_k^{(1)}] = (S_k^{(0)})^{-1}S_k^{(0)}(S^{(0)})^{-1}S_k^{(0)}(S_k^{(0)})^{-1} + k^2(S_k^{(0)})^{-1}\beta\beta^{\top}(S_k^{(0)})^{-1}$$

where $M_0^{(0)} = (S^{(0)})^{-1} - (S^{(0)})^{-1}R^{\top}(R(S^{(0)})^{-1}R^{\top})^{-1}R(S^{(0)})^{-1}$ and
$P_0 = (S^{(0)})^{-1}R^{\top}(R(S^{(0)})^{-1}R^{\top})^{-1}$.

3.3.1.1. The Comparison Between the First-Order Approximated Restricted Ridge Estimator and the First-Order Approximated ML Estimator

For the superiority of $\hat{\beta}_r(k)^{(1)}$ over $\hat{\beta}^{(1)}$, we consider the matrix MSE difference

$$\begin{aligned}\Delta_1 &= \text{MSE}[\hat{\beta}^{(1)}] - \text{MSE}[\hat{\beta}_r(k)^{(1)}] \\ &= (S^{(0)})^{-1} - M_k^{(0)} S^{(0)} M_k^{(0)} - (M_k^{(0)} S^{(0)} - I)(\beta - R^\top (RR^\top)^{-1} r) \\ &\quad \times (\beta - R^\top (RR^\top)^{-1} r)^\top (M_k^{(0)} S^{(0)} - I)^\top.\end{aligned}$$

The comparison is considered in two cases where $R\beta \neq r$ and $R\beta = r$. In this thesis, we propose following Theorems.

Case 1. The restrictions are not true, i.e. $R\beta \neq r$

Theorem 3.3 Under the condition that $(S^{(0)})^{-1} - M_k^{(0)} S^{(0)} M_k^{(0)}$ is nnd, the estimator $\hat{\beta}_r(k)^{(1)}$ is better than the estimator $\hat{\beta}^{(1)}$ in the sense of matrix MSE criterion if and only if

$$\begin{aligned}(\beta - R^\top (RR^\top)^{-1} r)^\top (M_k^{(0)} S^{(0)} - I)^\top ((S^{(0)})^{-1} - M_k^{(0)} S^{(0)} M_k^{(0)})^- \\ \times (M_k^{(0)} S^{(0)} - I)(\beta - R^\top (RR^\top)^{-1} r) \leq 1\end{aligned}$$

when the restrictions are not true.

Proof: Let $\mathbb{C}(A)$ denotes the column space of A where $A = (S^{(0)})^{-1} - M_k^{(0)} S^{(0)} M_k^{(0)}$. Then if $(M_k^{(0)} S^{(0)} - I)[\beta - R^\top (RR^\top)^{-1} r] \in \mathbb{C}(A)$, Δ_1 is nnd if and only if $AA^-(M_k^{(0)} S^{(0)} - I) = M_k^{(0)} S^{(0)} - I$. The proof of the equality $AA^-(M_k^{(0)} S^{(0)} - I) = M_k^{(0)} S^{(0)} - I$ can be followed from Özkale (2014).

Case 2. The restrictions are true, i.e. $R\beta = r$

Theorem 3.4 Under the conditions that $(S^{(0)})^{-1} - M_k^{(0)} S^{(0)} M_k^{(0)}$ is nnd, $\hat{\beta}_r(k)^{(1)}$ is superior to $\hat{\beta}^{(1)}$ in the sense of matrix MSE criterion if and only if

$$k^2 \beta^\top M_k^{(0)} ((S^{(0)})^{-1} - M_k^{(0)} S^{(0)} M_k^{(0)})^- M_k^{(0)} \beta \leq 1$$

when the restrictions are correct.

Proof: If $kM_k^{(0)}\beta \in \mathbb{C}(A)$, then Δ_1 is nnd if and only if $k^2\beta^\top M_k^{(0)} A^- M_k^{(0)} \beta \leq 1$ where A^- is a g-inverse of A under the condition that A is nnd (see for example, Trenkler and Toutenburg, 1990).

Since $\mathbb{C}(M_k^{(0)} S^{(0)} M_k^{(0)}) \subset \mathbb{C}((S^{(0)})^{-1})$ and $\mathfrak{R}(M_k^{(0)} S^{(0)} M_k^{(0)}) \subset \mathfrak{R}((S^{(0)})^{-1})$ where $\mathfrak{R}(\cdot)$ denotes the row space of a matrix, by applying Theorem A1 in Appendix-A from Harville (2008), and by following Özkale (2014), the g-inverse of A can be obtained as

$$A^- = S^{(0)} + S^{(0)} M_k^{(0)} ((S^{(0)})^{-1} - M_k^{(0)} S^{(0)} M_k^{(0)})^- M_k^{(0)} S^{(0)}.$$

Furthermore, $-kM_k^{(0)}\beta \in \mathbb{C}(A)$ if and only if $-kM_k^{(0)}\beta = -kAA^- M_k^{(0)}\beta$. The proof of the equality $-kM_k^{(0)}\beta = -kAA^- M_k^{(0)}\beta$ can be followed from Özkale (2014).

3.3.1.2. The Comparison Between the First-Order Approximated Restricted Ridge Estimator and the First-Order Approximated Restricted ML Estimator

For the superiority of $\hat{\beta}_r(k)^{(1)}$ over $\hat{\beta}_r^{(1)}$, we consider the matrix MSE difference as

$$\begin{aligned} \Delta_2 &= \text{MSE}[\hat{\beta}_r^{(1)}] - \text{MSE}[\hat{\beta}_r(k)^{(1)}] \\ &= M_0^{(0)} S^{(0)} M_0^{(0)} - M_k^{(0)} S^{(0)} M_k^{(0)} - k^2 M_k^{(0)} \beta\beta^\top M_k^{(0)}. \end{aligned}$$

The comparison is considered in two cases where $R\beta \neq r$ and $R\beta = r$. The results are presented by Theorems 3.5 and 3.6.

Case 1. The restrictions are not true, i.e. $R\beta \neq r$

Theorem 3.5 Under the condition that $M_0^{(0)} - M_k^{(0)} S^{(0)} M_k^{(0)}$ is a nnd matrix, the estimator $\hat{\beta}_r(k)^{(1)}$ dominates the estimator $\hat{\beta}_r^{(1)}$ in terms of the matrix MSE criterion if and only if

$$(\beta - R^\top (RR^\top)^{-1} r)^\top (M_k^{(0)} S^{(0)} - M_0^{(0)} S^{(0)})^\top (M_0^{(0)} - M_k^{(0)} S^{(0)} M_k^{(0)})^- \times (M_k^{(0)} S^{(0)} - M_0^{(0)} S^{(0)})(\beta - R^\top (RR^\top)^{-1} r) = 0$$

when the restrictions are incorrect.

Proof: Let $\mathbb{C}(D)$ denotes the column space of D where $D = M_0^{(0)} - M_k^{(0)} S^{(0)} M_k^{(0)}$. Δ_2 can be written as $\Delta_2 = D + d_1 d_1^\top - d_2 d_2^\top$ where d_1 and d_2 are the bias vectors of $\hat{\beta}_r^{(1)}$ and $\hat{\beta}_r(k)^{(1)}$, respectively. To the end of the comparison, it is important whether $d_i \in \mathbb{C}(D)$, $i = 1,2$ or not. From Theorem A3 in Appendix-A $d_1 \notin \mathbb{C}(D)$, $d_2 \notin \mathbb{C}(D)$ and $d_2 \in \mathbb{C}(D : d_1)$. d_1 and d_2 are linearly independent. Thus when D is nnd, Δ_2 is nnd. Because of similarity between the $\hat{\beta}_r(k)^{(1)}$ and $\hat{\beta}_r^{(1)}$ estimators in GLMs and restricted ridge and restricted ML estimators in linear model, the proof can be followed from Özkale (2014).

Case 2. The restrictions are true, i.e. $R\beta = r$

Theorem 3.6 Under the condition that $M_0^{(0)} - M_k^{(0)} S^{(0)} M_k^{(0)}$ is nnd, $\hat{\beta}_r(k)^{(1)}$ is superior to $\hat{\beta}_r^{(1)}$ in the sense of matrix MSE criterion if and only if

$$k^2 \beta^\top M_k^{(0)} ((S^{(0)})^{-1} - M_k^{(0)} S^{(0)} M_k^{(0)})^- M_k^{(0)} \beta \leq 1$$

when the restrictions are correct.

Proof: If $-kM_k^{(0)}\beta \in \mathbb{C}(D)$, then Δ_2 is nnd if and only if $k^2\beta^\top M_k^{(0)}D^- M_k^{(0)}\beta \leq 1$ where D^- is a g-inverse of D under the condition that D is nnd.

Since $\mathbb{C}(M_k^{(0)}S^{(0)}M_k^{(0)}) \subset \mathbb{C}(M_0^{(0)})$ and $\Re(M_k^{(0)}S^{(0)}M_k^{(0)}) \subset \Re(M_0^{(0)})$ where $\Re(\cdot)$ denotes the row space of a matrix, applying Theorem A1 in Appendix-A from Harville (2008), and by following Özkale (2014), the g-inverse of D can be written as

$$D^- = (M_0^{(0)})^- + (M_0^{(0)})^- M_k^{(0)}((S^{(0)})^{-1} - M_k^{(0)}(M_0^{(0)})^- M_k^{(0)})^- M_k^{(0)}(M_0^{(0)})^-.$$

Furthermore, $-kM_k^{(0)}\beta \in \mathbb{C}(D)$ if and only if $-kM_k^{(0)}\beta = -kDD^- M_k^{(0)}\beta$. The proof of the equality $-kM_k^{(0)}\beta = -kDD^- M_k^{(0)}\beta$ can be followed from Özkale (2014).

3.3.1.3. The Comparison Between the First-Order Approximated Restricted Ridge Estimator and the First-Order Approximated Ridge Estimator

For the superiority of $\hat{\beta}_r(k)^{(1)}$ over $\hat{\beta}_k^{(1)}$, we consider the matrix MSE difference $\Delta_3 = \text{MSE}[\hat{\beta}_k^{(1)}] - \text{MSE}[\hat{\beta}_r(k)^{(1)}] = B - M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}$ where $B = \text{MSE}[\hat{\beta}_k^{(1)}]$. The comparison is considered in $R\beta = r$ case. In this thesis, we propose the following Theorems.

Theorem 3.7 The $\hat{\beta}_r(k)^{(1)}$ estimator of β is superior to the $\hat{\beta}_k^{(1)}$ estimator by the criterion of matrix MSE if and only if $\lambda_{\max}(B^{-1}M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}) \leq 1$ when $R\beta = r$ is assumed to be true.

Proof: By applying Theorem A2 in Appendix-A from Graybill (1976) we can derive the necessary and sufficient condition for Δ_3 to be nnd. Since B is positive definite (pd) and $M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}$ is a symmetric matrix, there exists a nonsingular matrix Q such that $Q^\top BQ = I$ and $Q^\top M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}Q = \Lambda$, where Λ is a diagonal matrix whose elements which are the roots of the polynomial equation

$\left|M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}-\lambda B\right|=0$. Since

$$\left|M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}-\lambda B\right|=\left|M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}B^{-1}-\lambda I_p\right||B|=0,$$

λ_i is an eigenvalue of $M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}B^{-1}$. $M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}B^{-1}$ has the same eigenvalues with $B^{-1}M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}$. If $B-M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}$ is a nnd matrix and $Q^\top BQ-Q^\top M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}Q=I_p-\Lambda$ is a nnd matrix, then $1-\lambda_i\geq 0$, so we get $\lambda_{\max}(B^{-1}M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)})\leq 1$.

Let $\lambda_{\max}(B^{-1}M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)})\leq 1$. Since B is pd and $M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}$ is a symmetric matrix, we get

$$\lambda_p\leq\frac{x^\top M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}x}{x^\top Bx}\leq\lambda_1,$$

where $\lambda_1(B^{-1}M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)})\geq\ldots\geq\lambda_p(B^{-1}M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)})$ are the roots of $\left|M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}-\lambda B\right|=0$. Then we get $x^\top M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}x\leq x^\top Bx$, so $B-M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)}$ is a nnd matrix. It is obvious that Δ_3 is nnd if and only if $\lambda_{\max}(B^{-1}M_k^{(0)}S_k^{(0)}BS_k^{(0)}M_k^{(0)})\leq 1$.

3.3.2. The MSE of the First-Order Approximated Liu Estimator

We get the asymptotic bias vector and asymptotic variance-covariance matrix of the first-order approximated Liu estimator:

$$\text{Bias}[\hat{\beta}_d^{(1)}]=\text{E}[\hat{\beta}_d^{(1)}]-\beta=(d-1)(X^\top\hat{W}^{(0)}X+I)^{-1}\beta,$$

$$\begin{aligned}\text{Var}[\hat{\beta}_d^{(1)}] &= a(\phi)[(X^\top\hat{W}^{(0)}X+I)^{-1}(X^\top\hat{W}^{(0)}X+dI)(X^\top\hat{W}^{(0)}X)^{-1}\\ &\quad\times(X^\top\hat{W}^{(0)}X+dI)(X^\top\hat{W}^{(0)}X+I)^{-1}].\end{aligned}$$

Hence, by using Equation (3.23.), we get

$$\begin{aligned} \mathrm{MSE}[\hat{\beta}_d^{(1)}] &= a(\phi)[(X^\top \hat{W}^{(0)} X + I)^{-1}(X^\top \hat{W}^{(0)} X + dI)(X^\top \hat{W}^{(0)} X)^{-1} \\ &\quad \times (X^\top \hat{W}^{(0)} X + dI)(X^\top \hat{W}^{(0)} X + I)^{-1}] \\ &\quad + (d-1)^2 (X^\top \hat{W}^{(0)} X + I)^{-1} \beta \beta^\top (X^\top \hat{W}^{(0)} X + I)^{-1}. \end{aligned} \quad (3.29.)$$

3.3.2.1. Comparison of the First-Order Approximated Liu Estimator to the ML Estimator by the Matrix MSE Criterion in GLMs

Since $\hat{\beta}_d^{(1)}$ and $\hat{\beta}^{(1)}$ have the same working response and weight matrix, we can compare these two estimators in the sense of matrix MSE. $\hat{\beta}^{(1)}$ is asymptotically unbiased and has asymptotic variance-covariance matrix

$$\mathrm{Var}(\hat{\beta}^{(1)}) = a(\phi)(X^\top \hat{W}^{(0)} X)^{-1}. \quad (3.30.)$$

We are interested in investigating under which conditions $\hat{\beta}_d^{(1)}$ is better than $\hat{\beta}^{(1)}$. For this, we consider the difference $\Delta_4 = \mathrm{MSE}(\hat{\beta}^{(1)}) - \mathrm{MSE}(\hat{\beta}_d^{(1)})$. In this thesis, we propose following Theorems.

Theorem 3.8 $\mathrm{MSE}(\hat{\beta}^{(1)}) - \mathrm{MSE}(\hat{\beta}_d^{(1)})$ is pd if and only if

$$\beta^\top \left[I + \frac{(1+d)}{2} (X^\top \hat{W}^{(0)} X)^{-1} \right]^{-1} \beta < \frac{2a(\phi)}{1-d}.$$

Proof: From Equations (3.29.) and (3.30.), we find the matrix difference

$$\begin{aligned} \Delta_4 &= a(\phi)[(X^\top \hat{W}^{(0)} X)^{-1} - (X^\top \hat{W}^{(0)} X + I)^{-1}(X^\top \hat{W}^{(0)} X + dI) \\ &\quad \times (X^\top \hat{W}^{(0)} X)^{-1}(X^\top \hat{W}^{(0)} X + dI)(X^\top \hat{W}^{(0)} X + I)^{-1}] \\ &\quad - (d-1)^2 (X^\top \hat{W}^{(0)} X + I)^{-1} \beta \beta^\top (X^\top \hat{W}^{(0)} X + I)^{-1}. \end{aligned}$$

After some algebra, we get

$$\begin{aligned}\Delta_4 &= (1-d)(X^\top\hat{W}^{(0)}X+I)^{-1}\{[2a(\phi)I+(1+d)a(\phi)(X^\top\hat{W}^{(0)}X)^{-1}] \\ &\quad -(1-d)\beta\beta^\top\}(X^\top\hat{W}^{(0)}X+I)^{-1} \\ &= (X^\top\hat{W}^{(0)}X+I)^{-1}M(X^\top\hat{W}^{(0)}X+I)^{-1}\end{aligned}$$

where $M=(1-d)\{[2a(\phi)I+(1+d)a(\phi)(X^\top\hat{W}^{(0)}X)^{-1}]-(1-d)\beta\beta^\top\}$.
We may write M as follows:

$$\begin{aligned}M &= (1-d)\left[(1-d)\left\{\frac{a(\phi)(2I+(1+d)(X^\top\hat{W}^{(0)}X)^{-1})}{1-d}-\beta\beta^\top\right\}\right] \\ &= (1-d)^2\left[\frac{2a(\phi)}{1-d}\left(I+\frac{(1+d)}{2}(X^\top\hat{W}^{(0)}X)^{-1}\right)-\beta\beta^\top\right] \\ &= (1-d)^2[cA-\beta\beta^\top]\end{aligned}$$

where $c=\frac{2a(\phi)}{1-d}$ is a positive scalar and $A_1=I+\frac{(1+d)}{2}(X^\top\hat{W}^{(0)}X)^{-1}$ is a pd matrix. Thus, from Farebrother (1976), $cA_1-\beta\beta^\top$ is pd if and only if $\beta^\top A_1^{-1}\beta<c$. Writing A_1 and c, we complete the proof of the theorem.

3.3.2.2. Comparison of the First-Order Approximated Liu Estimator to the ML Estimator by sMSE Criterion in GLMs

It is well known that a linear regression model can be transformed to a canonical form by orthogonal transformation. By using spectral decomposition, the information matrix at the first order iteration can be expressed as $X^\top\hat{W}^{(0)}X=T\Lambda T^\top$ where $T=[T_1,\ldots,T_q]$ is a $q\times q$ orthogonal matrix with columns that constitute the eigenvectors of $X^\top\hat{W}^{(0)}X$ and $T^\top X^\top\hat{W}^{(0)}XT=Z^\top\hat{W}^{(0)}Z=\Lambda=diag(\lambda_j)$ is a $q\times q$ diagonal matrix including the n nonzero eigenvalues of $X^\top\hat{W}^{(0)}X$ ($\lambda_1=\lambda_{\max}\geq\lambda_2\geq\ldots\geq\lambda_q=\lambda_{\min}$) on its diagonal where $Z=XT$. Then, the MSE matrices of the first-order approximated ML and first-order approximated Liu estimators in GLMs given

by Equations (3.31.)-(3.32.) can be written in terms of canonical form as follows:

$$\mathrm{MSE}(\hat{\alpha}^{(1)}) = a(\phi)\Lambda^{-1}, \tag{3.31.}$$

$$\begin{aligned}\mathrm{MSE}(\hat{\alpha}_d^{(1)}) &= a(\phi)[(\Lambda+I)^{-1}(\Lambda+dI)\Lambda^{-1}(\Lambda+dI)(\Lambda+I)^{-1}] \\ &\quad +(d-1)^2(\Lambda+I)^{-1}\alpha\alpha^{\top}(\Lambda+I)^{-1},\end{aligned} \tag{3.32.}$$

where $\alpha = T^{\top}\beta$, $\hat{\alpha}^{(1)} = T^{\top}\hat{\beta}^{(1)}$ and $\hat{\alpha}_d^{(1)} = T^{\top}\hat{\beta}_d^{(1)}$. The sMSE of the first-order approximated ML estimator and the first-order approximated Liu estimator in GLMs are

$$\mathrm{sMSE}[\hat{\alpha}^{(1)}] = a(\phi)\sum_{j=1}^{q}\frac{1}{\lambda_j}, \tag{3.33.}$$

and

$$\mathrm{sMSE}[\hat{\alpha}_d^{(1)}] = a(\phi)\sum_{j=1}^{q}\frac{(\lambda_j+d)^2}{\lambda_j(\lambda_j+1)^2} + (d-1)^2\sum_{j=1}^{q}\frac{\alpha_j^2}{(\lambda_j+1)^2} \tag{3.34.}$$

respectively, where α_j is defined as the jth element of α.

For the superiorities of $\hat{\alpha}_d^{(1)}$ and $\hat{\alpha}_{\text{Đ}}^{(1)} = T^{\top}\hat{\beta}_{\text{Đ}}^{(1)}$ over $\hat{\alpha}^{(1)}$, we consider the differences $\mathrm{sMSE}[\hat{\alpha}^{(1)}] - \mathrm{sMSE}[\hat{\alpha}_d^{(1)}]$ and $\mathrm{sMSE}[\hat{\alpha}^{(1)}] - \mathrm{sMSE}[\hat{\alpha}_{\text{Đ}}^{(1)}]$ and obtain the results given by Theorems 3.9 and 3.10.

Theorem 3.9 If $d > 1 - 2\sum_{j=1}^{q}\frac{a(\phi)}{\lambda_j(\lambda_j+1)} / \sum_{j=1}^{q}\frac{\lambda_j\alpha_j^2+a(\phi)}{\lambda_j(\lambda_j+1)^2}$, then $\hat{\alpha}_d^{(1)}$ is superior to $\hat{\alpha}^{(1)}$ in terms of sMSE.

Proof: From Equations (3.33.) and (3.34.), we have

$$\begin{aligned}\mathrm{sMSE}[\hat{\alpha}^{(1)}] - \mathrm{sMSE}[\hat{\alpha}_d^{(1)}] &= \sum_{j=1}^{q}\frac{a(\phi)[(1-d)(2\lambda_j+d+1)] - (d-1)^2\alpha_j^2\lambda_j}{\lambda_j(\lambda_j+1)^2} \\ &= (d-1)\sum_{j=1}^{q}\frac{\alpha_j^2\lambda_j - a(\phi) - d(\alpha_j^2\lambda_j + a(\phi)) - 2a(\phi)\lambda_j}{\lambda_j(\lambda_j+1)^2}.\end{aligned} \tag{3.35.}$$

Since $0 < d < 1$, $d-1$ is negative. Then for the difference $\text{sMSE}[\hat{\alpha}^{(1)}] - \text{sMSE}[\hat{\alpha}_d^{(1)}]$ is positive, the inequality $\sum_{j=1}^{q} \frac{\alpha_j^2\lambda_j - a(\phi) - d(\alpha_j^2\lambda_j + a(\phi)) - 2a(\phi)\lambda_j}{\lambda_j(\lambda_j+1)^2} < 0$ must be satisfied. Then the difference in Equation (3.35.) is positive if

$$d > 1 - 2\frac{\sum_{j=1}^{q} \frac{a(\phi)}{\lambda_j(\lambda_j+1)}}{\sum_{j=1}^{q} \frac{\lambda_j\alpha_j^2 + a(\phi)}{\lambda_j(\lambda_j+1)^2}}.$$

Theorem 3.10 If $d_j > 1 - 2a(\phi)\frac{1}{\lambda_j(\lambda_j+1)} / \frac{\lambda_j\alpha_j^2 + a(\phi)}{\lambda_j(\lambda_j+1)^2}$ for $j = 1, \ldots, q$, then $\hat{\alpha}_{\text{D}}^{(1)}$ is superior to $\hat{\alpha}^{(1)}$ in terms of sMSE.

Proof: For $\hat{\alpha}_{\text{D}}^{(1)}$, we have $\text{sMSE}[\hat{\alpha}_{\text{D}}^{(1)}] = a(\phi) \sum_{j=1}^{q} \frac{(\lambda_j + d_j)^2}{\lambda_j(\lambda_j+1)^2} + \sum_{j=1}^{q} \frac{\alpha_j^2(d_j-1)^2}{(\lambda_j+1)^2}$. So the difference $\text{sMSE}[\hat{\alpha}^{(1)}] - \text{sMSE}[\hat{\alpha}_{\text{D}}^{(1)}]$ to be positive if $d_j > 1 - 2a(\phi)\frac{1}{\lambda_j(\lambda_j+1)} / \frac{\lambda_j\alpha_j^2 + a(\phi)}{\lambda_j(\lambda_j+1)^2}$.

Remark 3.11 The conditions given by Theorems 3.8, 3.9 and 3.10 depend on the unknown parameters $a(\phi)$, α and α_j^2. Therefore, one estimator can perform better than the other for certain values of $a(\phi)$, α and α_j^2 but can be worse for others. We here prefer to use the estimators of $a(\phi)$ and α_j^2 given by $\hat{\varphi}^2$ and $(\hat{\alpha}_j^{(1)})^2 - \frac{\hat{\varphi}^2}{\lambda_j}$ where $\hat{\alpha}_j^{(1)}$ is the jth element of $\hat{\alpha}^{(1)}$.

3.3.2.3. Selection of the Biasing Parameter d

Differentiating $\text{sMSE}[\hat{\alpha}_d^{(1)}]$ with respect to d gives

$$\frac{\partial \text{sMSE}[\hat{\alpha}_d^{(1)}]}{\partial d} = a(\phi) \sum_{j=1}^{q} \frac{2(\lambda_j + d)}{\lambda_j(\lambda_j+1)^2} + 2(d-1) \sum_{j=1}^{q} \frac{\alpha_j^2}{(\lambda_j+1)^2}$$

and equating the result to zero, we obtain

$$d_{opt} = \frac{\sum_{j=1}^{q} \frac{\alpha_j^2 - a(\phi)}{(\lambda_j+1)^2}}{\sum_{j=1}^{q} \frac{a(\phi) + \lambda_j\alpha_j^2}{\lambda_j(\lambda_j+1)^2}} = 1 - \frac{\sum_{j=1}^{q} \frac{a(\phi)}{\lambda_j(\lambda_j+1)}}{\sum_{j=1}^{q} \frac{a(\phi) + \lambda_j\alpha_j^2}{\lambda_j(\lambda_j+1)^2}}.$$

Substituting α_j^2 and $a(\phi)$ by their asymptotically unbiased estimates $(\hat{\alpha}_j^{(1)})^2 - \frac{\hat{\varphi}^2}{\lambda_j}$ and $\hat{\varphi}^2$, we get the estimate of d_{opt} as

$$\hat{d}_{opt} = 1 - \left[\sum_{j=1}^{q} \frac{\hat{\varphi}^2}{\lambda_j(\lambda_j+1)} / \sum_{j=1}^{q} \frac{(\hat{\alpha}_j^{(1)})^2}{(\lambda_j+1)^2} \right]. \tag{3.36.}$$

In Equation (3.36.), $\hat{d}_{opt}$ is always smaller than one. However, it is positive for $\hat{\varphi}^2 < \left[\sum_{j=1}^{q} \frac{(\hat{\alpha}_j^{(1)})^2}{(\lambda_j+1)^2} / \sum_{j=1}^{q} \frac{1}{\lambda_j(\lambda_j+1)} \right]$. To get an estimate of d in interval $(0,1)$, we define

$$\hat{d}_h = 1 - h \left[\sum_{j=1}^{q} \frac{\hat{\varphi}^2}{\lambda_j(\lambda_j+1)} / \sum_{j=1}^{q} \frac{(\hat{\alpha}_j^{(1)})^2}{(\lambda_j+1)^2} \right] \tag{3.37.}$$

where h is an arbitrary constant which satisfies the inequality $0 < h < \frac{\sum_{j=1}^{q} \frac{(\hat{\alpha}_j^{(1)})^2}{(\lambda_j+1)^2}}{\sum_{j=1}^{q} \frac{\hat{\varphi}^2}{\lambda_j(\lambda_j+1)}}$. However, depending on the value of h, $\hat{d}_h$ may not minimize $\mathrm{sMSE}[\hat{\alpha}_d^{(1)}]$.

Furthermore, $\hat{d}_h$ in interval $(0,1)$ can be found such that $\mathrm{sMSE}[\hat{\alpha}_{\hat{d}_h}^{(1)}]$ is smaller than $\mathrm{sMSE}[\hat{\alpha}^{(1)}]$. In this case h should satisfy the condition in Theorem 3.13.

Lemma 3.12 (Stein, 1981). Assume $y \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 is known. Let $f(y)$ be a real function with $\mathrm{E}|f'(y)| < \infty$, then $\mathrm{E}(f'(y)) = \mathrm{E}((y-\mu)f(y))/\sigma^2$.

Theorem 3.13 Let $\lambda_1 = \lambda_{\max} \geq \lambda_2 \geq \ldots \geq \lambda_q = \lambda_{\min}$ be the eigenvalues of $X^\top \hat{W}^{(0)} X$. Assume that h satisfies

$$0 < h < \min \left\{ \frac{2(n-q)}{n-q+2} \left[1 - \frac{2}{\sum_{j=1}^{q} \frac{\lambda_{\min}(\lambda_{\min}+1)}{\lambda_j(\lambda_j+1)}} \right], \frac{\sum_{j=1}^{q} \frac{(\hat{\alpha}_j^{(1)})^2}{(\lambda_j+1)^2}}{\sum_{j=1}^{q} \frac{\hat{\varphi}^2}{\lambda_j(\lambda_j+1)}} \right\}$$

if $\sum_{j=1}^{q} \frac{\lambda_{\min}(\lambda_{\min}+1)}{\lambda_j(\lambda_j+1)} > 2$. Then $\mathrm{sMSE}[\hat{\alpha}_{\hat{d}_h}^{(1)}] < \mathrm{sMSE}[\hat{\alpha}^{(1)}]$ for all α and $a(\phi)$.

Proof: The ith element of $\hat{\alpha}^{(1)}_{\hat{d}_h}$ can be written as $\hat{\alpha}^{(1)}_{\hat{d}_{h,i}} = \frac{\lambda_i+d}{\lambda_i+1}\hat{\alpha}^{(1)}_i$. Then the asymptotic sMSE of $\hat{\alpha}^{(1)}_{\hat{d}_{h,i}}$ equals

$$\begin{aligned}
\mathrm{E}[(\hat{\alpha}^{(1)}_{\hat{d}_{h,i}} - \alpha_i)^2] &= \mathrm{E}[(\frac{\lambda_i + \hat{d}_h}{\lambda_i+1}\hat{\alpha}^{(1)}_i - \alpha_i)^2] \\
&= \mathrm{E}[(\hat{\alpha}^{(1)}_i - \alpha_i)^2] + \frac{2}{\lambda_i+1}\mathrm{E}[(\hat{\alpha}^{(1)}_i - \alpha_i)(\hat{d}_h - 1)\hat{\alpha}^{(1)}_i] \\
&\quad + \left(\frac{1}{\lambda_i+1}\right)^2 \mathrm{E}[(\hat{d}_h - 1)^2(\hat{\alpha}^{(1)}_i)^2]. \qquad (3.38.)
\end{aligned}$$

Since $A_2 = \mathrm{E}[(\hat{\alpha}^{(1)}_i - \alpha_i)(\hat{d}_h - 1)\hat{\alpha}^{(1)}_i] = -h\mathrm{E}\left[(\hat{\alpha}^{(1)}_i - \alpha_i)\hat{\varphi}^2 \frac{\sum_{j=1}^q \frac{1}{\lambda_j(\lambda_j+1)}}{\sum_{j=1}^q \frac{(\hat{\alpha}^{(1)}_j)^2}{(\lambda_j+1)^2}}\hat{\alpha}^{(1)}_i\right]$, let $f(\hat{\alpha}^{(1)}_i) = \frac{\sum_{j=1}^q \frac{1}{\lambda_j(\lambda_j+1)}}{\sum_{j=1}^q \frac{(\hat{\alpha}^{(1)}_j)^2}{(\lambda_j+1)^2}}\hat{\alpha}^{(1)}_i$. Then by using Lemma, we get

$$A_2 = -h\frac{a(\phi)}{\lambda_i}\left[\frac{\sum_{j=1}^q \frac{\hat{\varphi}^2}{\lambda_j(\lambda_j+1)}}{\sum_{j=1}^q \frac{(\hat{\alpha}^{(1)}_j)^2}{(\lambda_j+1)^2}} - \frac{2(\hat{\alpha}^{(1)}_i)^2}{(\lambda_i+1)^2}\frac{\sum_{j=1}^q \frac{\hat{\varphi}^2}{\lambda_j(\lambda_j+1)}}{\left[\sum_{j=1}^q \frac{(\hat{\alpha}^{(1)}_j)^2}{(\lambda_j+1)^2}\right]^2}\right]$$

where $\hat{\alpha}^{(1)}_i \sim \mathcal{N}(\alpha_i, \frac{a(\phi)}{\lambda_i})$.

Using A_2, the Equation (3.38.) can be presented as

$$\begin{aligned}
\mathrm{E}[(\hat{\alpha}^{(1)}_{\hat{d}_{h,i}} - \alpha_i)^2] &= \mathrm{E}[(\hat{\alpha}^{(1)}_i - \alpha_i)^2] + \mathrm{E}\{\frac{\sum_{j=1}^q \frac{1}{\lambda_j(\lambda_j+1)}}{\sum_{j=1}^q \frac{(\hat{\alpha}^{(1)}_j)^2}{(\lambda_j+1)^2}} \\
&\quad \times[\frac{2(-h)}{\lambda_i(\lambda_i+1)}a(\phi)\hat{\varphi}^2 + \frac{4h(\hat{\alpha}^{(1)}_i)^2}{\lambda_i(\lambda_i+1)^3}a(\phi)\hat{\varphi}^2\frac{1}{\sum_{j=1}^q \frac{(\hat{\alpha}^{(1)}_j)^2}{(\lambda_j+1)^2}} \\
&\quad + \frac{h^2(\hat{\alpha}^{(1)}_i)^2}{(\lambda_i+1)^2}(\hat{\varphi}^2)^2\frac{\sum_{j=1}^q \frac{1}{\lambda_j(\lambda_j+1)}}{\sum_{j=1}^q \frac{(\hat{\alpha}^{(1)}_j)^2}{(\lambda_j+1)^2}}]\}. \qquad (3.39.)
\end{aligned}$$

From Equation (3.39.), the difference $\Delta = \text{sMSE}(\hat{\alpha}^{(1)}_{\hat{d}_h}) - \text{sMSE}(\hat{\alpha}^{(1)})$ equals

$$\begin{aligned}\Delta &= \text{E}\{\frac{\sum_{j=1}^{q}\frac{\hat{\varphi}^2}{\lambda_j(\lambda_j+1)}}{\sum_{j=1}^{q}\frac{(\hat{\alpha}_j^{(1)})^2}{(\lambda_j+1)^2}}[\sum_{j=1}^{q}\frac{(-2h)}{\lambda_i(\lambda_i+1)}a(\phi) \\ &\quad +4ha(\phi)\sum_{j=1}^{q}\frac{(\hat{\alpha}_i^{(1)})^2}{\lambda_i(\lambda_i+1)^3}\frac{1}{\sum_{j=1}^{q}\frac{(\hat{\alpha}_j^{(1)})^2}{(\lambda_j+1)^2}}+h^2\sum_{j=1}^{q}\frac{\hat{\varphi}^2}{\lambda_j(\lambda_j+1)}]\}.\end{aligned}$$

By the asymptotic independence of $\hat{\alpha}_j^{(1)}$ and $\hat{\varphi}^2$ (see Agresti, 2015) and that $\hat{\varphi}_p^2$ is approximately unbiased, $\text{E}(\hat{\varphi}_p^2) = a(\phi)$ and $\text{E}[(\hat{\varphi}_p^2)^2] = \frac{n-q+2}{n-q}[a(\phi)]^2$, we get

$$\begin{aligned}\Delta &\leq \text{E}\left\{\frac{\sum_{j=1}^{q}\frac{\hat{\varphi}^2}{\lambda_j(\lambda_j+1)}}{\sum_{j=1}^{q}\frac{(\hat{\alpha}_j^{(1)})^2}{(\lambda_j+1)^2}}h\left[\sum_{j=1}^{q}\frac{-2a(\phi)}{\lambda_i(\lambda_i+1)}+\frac{4a(\phi)}{\lambda_{\min}(\lambda_{\min}+1)}+h\sum_{j=1}^{q}\frac{\hat{\varphi}^2}{\lambda_j(\lambda_j+1)}\right]\right\} \\ &= \text{E}\left\{\left[\frac{\sum_{j=1}^{q}\frac{\hat{\varphi}^2}{\lambda_j(\lambda_j+1)}}{\sum_{j=1}^{q}\frac{(\hat{\alpha}_j^{(1)})^2}{(\lambda_j+1)^2}}\right]\right\}h[\sum_{j=1}^{q}\frac{-2a(\phi)}{\lambda_i(\lambda_i+1)} \\ &\quad +\frac{4a(\phi)}{\lambda_{\min}(\lambda_{\min}+1)}+h\sum_{j=1}^{q}\frac{n-q+2}{(n-q)\lambda_j(\lambda_j+1)}].\end{aligned}$$

For Δ to be negative, $H(h) = h\left[\sum_{j=1}^{q}\frac{-2a(\phi)}{\lambda_i(\lambda_i+1)}+\frac{4a(\phi)}{\lambda_{\min}(\lambda_{\min}+1)}+h\sum_{j=1}^{q}\frac{n-q+2}{(n-q)\lambda_j(\lambda_j+1)}\right]$ should be negative. $H(h)$ is negative for $0 < h < \frac{2(n-q)}{n-q+2}\left[1-\frac{2}{\sum_{j=1}^{q}\frac{\lambda_{\min}(\lambda_{\min}+1)}{\lambda_j(\lambda_j+1)}}\right]$ if $\sum_{j=1}^{q}\frac{\lambda_{\min}(\lambda_{\min}+1)}{\lambda_j(\lambda_j+1)} > 2$ and

$$\frac{2(n-q)}{n-q+2}\left[1-\frac{2}{\sum_{j=1}^{q}\frac{\lambda_{\min}(\lambda_{\min}+1)}{\lambda_j(\lambda_j+1)}}\right] < h < 0 \text{ if } \sum_{j=1}^{q}\frac{\lambda_{\min}(\lambda_{\min}+1)}{\lambda_j(\lambda_j+1)} < 2. \qquad (3.40.)$$

Since $\hat{d}_h$ should be in interval $(0,1)$, h defined by Equation (3.40.) does not result in $\hat{d}_h \in (0,1)$. The proof is completed.

Remark 3.14 If the data set satisfies the condition $\sum_{j=1}^{q}\frac{\lambda_{\min}(\lambda_{\min}+1)}{\lambda_j(\lambda_j+1)} < 2$, then a $d \in (0,1)$ can be chosen as $\hat{d}_h$.

3.3.3. MSE of the First-Order Approximated Restricted Liu Estimator

In this section, we compare the first-order approximated restricted Liu estimator to the first-order approximated ML, restricted ML and Liu estimators according to the MSE criterion.

We compute approximated biasing vector and variance-covariance matrix of the estimator $\hat{\beta}_r(d)^{(1)}$ as (see also Özkale, 2014)

$$\text{Bias}[\hat{\beta}_r(d)^{(1)}] = (d-1)M_1^{(0)}\beta - (S_1^{(0)})^{-1}R^\top(R(S_1^{(0)})^{-1}R^\top)^{-1}(R\beta - r)$$

$$\text{Var}[\hat{\beta}_r(d)^{(1)}] = a(\phi)M_1^{(0)}S_d^{(0)}(S_1^{(0)})^{-1}S_d^{(0)}M_1^{(0)}$$

where $M_1^{(0)} = (S_1^{(0)})^{-1} - (S_1^{(0)})^{-1}R^\top(R(S_1^{(0)})^{-1}R^\top)^{-1}R(S_1^{(0)})^{-1}$, $S_d^{(0)} = X^\top\hat{W}^{(0)}X + dI$, and $S_1^{(0)} = X^\top\hat{W}^{(0)}X + I$. Then, the matrix MSE of the estimator $\hat{\beta}_r(d)^{(1)}$ is

$$\begin{aligned}\text{MSE}[\hat{\beta}_r(d)^{(1)}] &= a(\phi)M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)} \\ &\quad + [(d-1)M_1^{(0)}\beta - (S_1^{(0)})^{-1}R^\top(R(S_1^{(0)})^{-1}R^\top)^{-1}(R\beta - r)] \\ &\quad \times [(d-1)M_1^{(0)}\beta - (S_1^{(0)})^{-1}R^\top(R(S_1^{(0)})^{-1}R^\top)^{-1}(R\beta - r)]^\top. \quad (3.41.)\end{aligned}$$

Similarly, assuming that the linear restrictions $R\beta = r$ hold true for $\hat{\beta}_r^{(1)}$, the MSE matrices of the estimators $\hat{\beta}^{(1)}$, $\hat{\beta}_r^{(1)}$ and $\hat{\beta}_d^{(1)}$ are respectively as

$$\text{MSE}[\hat{\beta}^{(1)}] = a(\phi)(S^{(0)})^{-1}, \quad (3.42.)$$

$$\text{MSE}[\hat{\beta}_r^{(1)}] = a(\phi)M_0^{(0)}, \quad (3.43.)$$

and

$$\text{MSE}[\hat{\beta}_d^{(1)}] = a(\phi)(S_1^{(0)})^{-1}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}(S_1^{(0)})^{-1} + (d-1)^2(S_1^{(0)})^{-1}\beta\beta^\top(S_1^{(0)})^{-1}. \quad (3.44.)$$

3.3.3.1. The Comparison Between the First-Order Approximated Restricted Liu Estimator and the First-Order Approximated Liu Estimator

For the superiority of $\hat{\beta}_r(d)^{(1)}$ over $\hat{\beta}_d^{(1)}$ when $R\beta = r$ holds true, Theorem 3.15 can be given. The difference $\Delta_5 = \text{MSE}[\hat{\beta}_d^{(1)}] - \text{MSE}[\hat{\beta}_r(d)^{(1)}]$ is given by $\Delta_5 = B - M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)}$ where $B = \text{MSE}[\hat{\beta}_d^{(1)}]$.

Theorem 3.15 The estimator $\hat{\beta}_r(d)^{(1)}$ of β is superior to the estimator $\hat{\beta}_d^{(1)}$ by the criterion of matrix MSE if and only if $\lambda_{\max}(B^{-1} M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)}) \leq 1$ where $\lambda_{\max}$ is the maximum eigenvalue of $M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)} B^{-1}$.

Proof: For the superiority of $\hat{\beta}_r(d)^{(1)}$ over $\hat{\beta}_d^{(1)}$, we examine the difference $\Delta_5 = \text{MSE}[\hat{\beta}_d^{(1)}] - \text{MSE}[\hat{\beta}_r(d)^{(1)}]$ is given by $\Delta_5 = B - M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)}$. By applying Theorem A2 in Appendix-A from Graybill (1976) we can derive the necessary and sufficient condition for Δ_5 to be nnd. Since B is pd and $M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)}$ is a symmetric matrix, there exists a nonsingular matrix Q such that $Q^\top B Q = I$ and $Q^\top M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)} Q = \Lambda$, where Λ is a diagonal elements are the roots of the polynomial equation $\left| M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)} - \lambda B \right| = 0$. Since

$$\left| M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)} - \lambda B \right| = \left| M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)} B^{-1} - \lambda I_q \right| |B| = 0,$$

λ_i is an eigenvalue of $M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)} B^{-1}$. $M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)} B^{-1}$ has the same eigenvalues with $B^{-1} M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)}$. If $B - M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)}$ is a nnd matrix, $Q^\top B Q - Q^\top M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)} Q = I_q - \Lambda$ is a nnd matrix. Then $1 - \lambda_i \geq 0$, so we get $\lambda_{\max}(B^{-1} M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)}) \leq 1$.

Let $\lambda_{\max}(B^{-1} M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)}) \leq 1$. Since B is pd and $M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)}$ is a symmetric matrix, we have get

$$\lambda_q \leq \frac{x^\top M_1^{(0)} S_1^{(0)} B S_1^{(0)} M_1^{(0)} x}{x^\top B x} \leq \lambda_1,$$

where $\lambda_1(B^{-1}M_1^{(0)}S_1^{(0)}BS_1^{(0)}M_1^{(0)}) \geq \ldots \geq \lambda_q(B^{-1}M_1^{(0)}S_1^{(0)}BS_1^{(0)}M_1^{(0)})$ are the roots of $\left|M_1^{(0)}S_1^{(0)}BS_1^{(0)}M_1^{(0)} - \lambda B\right| = 0$. Then we get $x^\top M_1^{(0)}S_1^{(0)}BS_1^{(0)}M_1^{(0)}x \leq x^\top Bx$, so $B - M_1^{(0)}S_1^{(0)}BS_1^{(0)}M_1^{(0)}$ is a nnd matrix. It is obvious that Δ_5 is nnd if and only if $\lambda_{\max}(B^{-1}M_1^{(0)}S_1^{(0)}BS_1^{(0)}M_1^{(0)}) \leq 1$.

3.3.3.2. The Comparison Between the First-Order Approximated Restricted Liu Estimator and the First-Order Approximated Restricted ML Estimator

Case 1. The restrictions are true, i.e. $R\beta = r$

For the superiority of $\hat{\beta}_r(d)^{(1)}$ over $\hat{\beta}_r^{(1)}$, we consider the matrix MSE difference $\Delta_6 = \mathrm{MSE}[\hat{\beta}_r^{(1)}] - \mathrm{MSE}[\hat{\beta}_r(d)^{(1)}]$. In the case of $R\beta = r$, the matrix MSE difference Δ_6 equals to

$$\Delta_6 = a(\phi)M_0 - a(\phi)M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)} - (d-1)^2M_1^{(0)}\beta\beta^\top M_1^{(0)}.$$

The result can be presented by Theorem 3.16.

Theorem 3.16 Under the conditions that $a(\phi)[M_0^{(0)} - M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)}]$ is nnd, the estimator $\hat{\beta}_r(d)^{(1)}$ is superior to the estimator $\hat{\beta}_r^{(1)}$ by the matrix MSE criterion if and only if

$$(d-1)^2\beta^\top M_1^{(0)}(M_0^{(0)} - M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)})^- M_1^{(0)}\beta \leq 1$$

when the restrictions are true.

Proof: Let $\mathbb{C}(D)$ denotes the column space of D where $D = a(\phi)[M_0^{(0)} - M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)}]$. Then if $(1-d)M_1^{(0)}\beta \in \mathbb{C}(D)$, Δ_6 is nnd if and only if $(d-1)^2\beta^\top M_1^{(0)}D^-M_1^{(0)}\beta \leq 1$ where D^- is a g-inverse of D under the condition that D is nnd.

Since $\mathbb{C}(M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)}) \subset \mathbb{C}(M_0^{(0)})$ and $\Re(M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)}) \subset \Re(M_0^{(0)})$ where $\Re(\cdot)$ denotes the row space of a matrix, applying Theorem A1 in Appendix-A from Harville (2008), and by following Özkale (2014), the g-inverse of D can be written as

$$\begin{aligned} D^{-} &= \frac{1}{a(\phi)}[(M_0^{(0)})^{-} + (M_0^{(0)})^{-}M_1^{(0)}((S_d^{(0)})^{-1}S^{(0)}(S_d^{(0)})^{-1} \\ &\quad -M_1^{(0)}(M_0^{(0)})^{-}M_1^{(0)})^{-}M_1^{(0)}(M_0^{(0)})^{-}]. \end{aligned}$$

Furthermore, $-(d-1)M_1^{(0)}\beta \in \mathbb{C}(D)$ if and only if $-(d-1)M_1^{(0)}\beta = -(d-1)DD^{-}M_1^{(0)}\beta$. The proof of the equality $-(d-1)M_1^{(0)}\beta = -(d-1)DD^{-}M_1^{(0)}\beta$ can be followed from Özkale (2014).

Case 2. The restrictions are not true, i.e. $R\beta \neq r$

In this case, the performances of the $\hat{\beta}_r(d)^{(1)}$ and $\hat{\beta}_r^{(1)}$ estimators depend upon $R\beta \neq r$ and both of them are biased estimators of β. We consider the difference

$$\begin{aligned} \Delta_7 &= \mathrm{MSE}[\hat{\beta}_r^{(1)}] - \mathrm{MSE}[\hat{\beta}_r(d)^{(1)}] \\ &= a(\phi)(M_0^{(0)} - M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)}) + (S^{(0)})^{-1}R^{\top}(R(S^{(0)})^{-1}R^{\top})^{-1} \\ &\quad \times (R\beta - r)(R\beta - r)^{\top}(R(S^{(0)})^{-1}R^{\top})^{-1}R^{\top}(S^{(0)})^{-1} \\ &\quad -(d-1)^2[M_1^{(0)}\beta - (S_1^{(0)})^{-1}R^{\top}(R(S_1^{(0)})^{-1}R^{\top})^{-1}](R\beta - r) \\ &\quad \times (R\beta - r)^{\top}[M_1^{(0)}\beta - (S_1^{(0)})^{-1}R^{\top}(R(S_1^{(0)})^{-1}R^{\top})^{-1}]^{\top} \end{aligned}$$

to compare the $\hat{\beta}_r^{(1)}$ estimator and the $\hat{\beta}_r(d)^{(1)}$ estimator. Then, we give Theorem 3.17.

Theorem 3.17 Under the condition that $a(\phi)(M_0^{(0)} - M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)})$ is a nnd matrix, the $\hat{\beta}_r(d)^{(1)}$ estimator dominates the $\hat{\beta}_r^{(1)}$ estimator in terms of the matrix

MSE criterion if and only if

$$[\beta-(RR^{\top})^{-1}r]^{\top}(M_1^{(0)}S_d^{(0)}-M_0^{(0)}S^{(0)})^{\top}(M_0^{(0)}-M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)})^{-}$$
$$\times(M_1^{(0)}S_d^{(0)}-M_0^{(0)}S^{(0)})[\beta-(RR^{\top})^{-1}r]=0$$

when the restrictions are incorrect.

Proof: Δ_7 can be written as $\Delta_7 = D+d_1d_1^{\top}-d_2d_2^{\top}$ where d_1 and d_2 are the biases of the $\hat{\beta}_r^{(1)}$ and $\hat{\beta}_r(d)^{(1)}$ estimators, respectively. To the end of the comparison, it is important whether $d_i \in \mathbb{C}(D)$, $i=1,2$ or not. From Theorem A3 in Appendix-A $d_1 \notin \mathbb{C}(D)$, $d_2 \notin \mathbb{C}(D)$ and $d_2 \in \mathbb{C}(D:d_1)$. d_1 and d_2 are linearly independent. Thus when D is nnd, Δ_7 is nnd. Because of similarity between the $\hat{\beta}_r(d)^{(1)}$ and $\hat{\beta}_r^{(1)}$ estimators in GLMs and restricted Liu and restricted ML estimators in linear model, the proof can be followed from Özkale (2014).

3.3.3.3. The Comparison Between the First-Order Approximated Restricted Liu Estimator and the First-Order Approximated ML Estimator

Case 1. The restrictions are true, i.e. $R\beta = r$

We consider the difference $\Delta_8 = \text{MSE}[\hat{\beta}^{(1)}] - \text{MSE}[\hat{\beta}_r(d)^{(1)}] = a(\phi)(S^{(0)})^{-1} - a(\phi)M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)} - (d-1)^2M_1^{(0)}\beta\beta^{\top}M_1^{(0)}$ to compare the first-order approximated ML estimator and the first-order approximated restricted Liu estimator. Then, we propose Theorem 3.18.

Theorem 3.18 Under the condition that $a(\phi)((S^{(0)})^{-1}-M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)})$ is nnd, $\hat{\beta}_r(d)^{(1)}$ is superior to $\hat{\beta}^{(1)}$ in the sense of matrix MSE criterion if and only if

$$(d-1)^2\beta^{\top}M_1^{(0)}((S^{(0)})^{-1}-M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)})^{-}M_1^{(0)}\beta \leq a(\phi)$$

when the restrictions are true.

Proof: Let $\mathbb{C}(A)$ denotes the column space of A where $A = a(\phi)[(S^{(0)})^{-1} - M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)}]$. Then if $(1-d)M_1^{(0)}\beta \in \mathbb{C}(A)$, Δ_8 is nnd if and only if $(d-1)^2\beta^\top M_1^{(0)}A^-M_1^{(0)}\beta \leq a(\phi)$ where A^- is a g-inverse of A under the condition that A is nnd.

Since $\mathbb{C}(M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)}) \subset \mathbb{C}((S^{(0)})^{-1})$ and $\mathfrak{R}(M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)})$ $\subset \mathfrak{R}((S^{(0)})^{-1})$ where $\mathfrak{R}(\cdot)$ denotes the row space of a matrix, applying Theorem A1 in Appendix-A from Harville (2008), and by following Özkale (2014), the g-inverse of A can be written as

$$A^- = \frac{1}{a(\phi)}[S^{(0)} + S^{(0)}M_1^{(0)}((S_d^{(0)})^{-1}S^{(0)}(S_d^{(0)})^{-1} - M_1^{(0)}S^{(0)}M_1^{(0)})^-M_1^{(0)}S^{(0)}].$$

Furthermore, $-(d-1)M_1^{(0)}\beta \in \mathbb{C}(A)$ if and only if $-(d-1)M_1^{(0)}\beta = -(d-1)AA^-M_1^{(0)}\beta$. The proof of the equality $-(d-1)M_1^{(0)}\beta = -(d-1)AA^-M_1^{(0)}\beta$ can be followed from Özkale (2014).

Case 2. The restrictions are not true, i.e. $R\beta \neq r$

In this case, Δ_9 equals to

$$\begin{aligned}\Delta_9 &= \text{MSE}[\hat{\beta}^{(1)}] - \text{MSE}[\hat{\beta}_r(d)^{(1)}] \\ &= a(\phi)(S^{(0)})^{-1} - a(\phi)M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)} - (M_1^{(0)}S_d^{(0)} - I) \\ &\quad \times[\beta - R^\top(RR^\top)^{-1}r][\beta - R^\top(RR^\top)^{-1}r]^\top(M_1^{(0)}S_d^{(0)} - I)^\top.\end{aligned}$$

to compare the $\hat{\beta}_r(d)^{(1)}$ and $\hat{\beta}^{(1)}$ estimators. Then, we give Theorem 3.19.

Theorem 3.19 Under the condition that $a(\phi)((S^{(0)})^{-1} - M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)})$ is nnd matrix, the $\hat{\beta}_r(d)^{(1)}$ estimator is better than the $\hat{\beta}^{(1)}$ estimator in the sense of the matrix MSE criterion if and only if

$$\begin{aligned}[\beta - R^\top(RR^\top)^{-1}r]^\top(M_1^{(0)}S_d^{(0)} - I)^\top((S^{(0)})^{-1} - M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)})^- \\ \times(M_1^{(0)}S_d^{(0)} - I)[\beta - R^\top(RR^\top)^{-1}r] \leq a(\phi)\end{aligned}$$

Table 3.1. MSE comparisons of estimators, restrictions and theorems

MSE comparisons	Restrictions	Theorems
$\text{MSE}[\hat{\beta}_r(k)^{(1)}] < \text{MSE}[\hat{\beta}^{(1)}]$	$R\beta \neq r$	Theorem 3.3
$\text{MSE}[\hat{\beta}_r(k)^{(1)}] < \text{MSE}[\hat{\beta}^{(1)}]$	$R\beta = r$	Theorem 3.4
$\text{MSE}[\hat{\beta}_r(k)^{(1)}] < \text{MSE}[\hat{\beta}_r^{(1)}]$	$R\beta \neq r$	Theorem 3.5
$\text{MSE}[\hat{\beta}_r(k)^{(1)}] < \text{MSE}[\hat{\beta}_r^{(1)}]$	$R\beta = r$	Theorem 3.6
$\text{MSE}[\hat{\beta}_r(k)^{(1)}] < \text{MSE}[\hat{\beta}_k^{(1)}]$	$R\beta = r$	Theorem 3.7
$\text{MSE}[\hat{\beta}_d^{(1)}] < \text{MSE}[\hat{\beta}^{(1)}]$	-	Theorem 3.8
$\text{sMSE}[\hat{\alpha}_d^{(1)}] < \text{sMSE}[\hat{\alpha}^{(1)}]$	-	Theorem 3.9
$\text{sMSE}[\hat{\alpha}_{\text{D}}^{(1)}] < \text{sMSE}[\hat{\alpha}^{(1)}]$	-	Theorem 3.10
$\text{MSE}[\hat{\beta}_r(d)^{(1)}] < \text{MSE}[\hat{\beta}_d^{(1)}]$	$R\beta = r$	Theorem 3.15
$\text{MSE}[\hat{\beta}_r(d)^{(1)}] < \text{MSE}[\hat{\beta}_r^{(1)}]$	$R\beta = r$	Theorem 3.16
$\text{MSE}[\hat{\beta}_r(d)^{(1)}] < \text{MSE}[\hat{\beta}_r^{(1)}]$	$R\beta \neq r$	Theorem 3.17
$\text{MSE}[\hat{\beta}_r(d)^{(1)}] < \text{MSE}[\hat{\beta}^{(1)}]$	$R\beta = r$	Theorem 3.18
$\text{MSE}[\hat{\beta}_r(d)^{(1)}] < \text{MSE}[\hat{\beta}^{(1)}]$	$R\beta \neq r$	Theorem 3.19

when the restrictions are not true.

Proof: If $(M_1^{(0)}S_d^{(0)} - I)[\beta - R^\top(RR^\top)^{-1}r] \in \mathbb{C}(A)$ where $A = a(\phi)[(S^{(0)})^{-1} - M_1^{(0)}S_d^{(0)}(S^{(0)})^{-1}S_d^{(0)}M_1^{(0)}]$, Δ_9 is nnd if and only if $AA^-(M_1^{(0)}S_d^{(0)} - I) = M_1^{(0)}S_d^{(0)} - I$. The proof of the equality $AA^-(M_1^{(0)}S_d^{(0)} - I) = M_1^{(0)}S_d^{(0)} - I$ can be followed from Özkale (2014).

We proposed some MSE results on theorems, according to the restrictions. To briefly summarize the theorems for comparing estimators, we give the Table 3.1.

4. NUMERICAL EXAMPLES

We illustrate the theoretical results of the proposed estimator on two different real life data sets.

4.1. Example 1: Mine Data Set

In this numerical example, the response has Poisson distribution with log link. The data set was originally given by Myers (1990) and was also used by Marx (1992). Myers (1990) presented 44 observations on mines in the coal fields of the Appalachian region of western Virginia. There are four continuous explanatory variables. These variables were analyzed for roles contributing to the number of injuries or fractures that occur in the upper seams of the mines.

In this study, as Myers (1990) and Marx (1992) did, a generalized linear regression assuming observations y_i are from Poisson distribution and the log link function is considered. The regression model is

$$\log \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i}$$

where μ_i is the expected number of upper seam injuries or fractures in the ith coal mine area, x_{1i} is the corresponding inner burden thickness (feet), which is the shortest distance between seam floor and lower seam, x_{2i} is the percent extraction of the lower previously mined seam, x_{3i} is the lower seam height (in feet) and x_{4i} is the time (years) that the mine has been opened.

Our computations here are performed by using R (see, Wickham 2009). Marx (1992) investigated the condition indices of the information matrix (with the inclusion of an intercept term) and he stated that the data suffer from the multicollinearity problem. The eigenvalues of $X^\top \hat{W} X$ at final iteration are obtained as $t_1 = 416468$, $t_2 = 338966$, $t_3 = 30607.8$, $t_4 = 3824.86$, $t_5 = 0.95040$. Thus, the con-

dition indices of the information matrix are 1.0000, 3.5052, 11.6647, 32.9976 and 2093.3225, indicating severe ill conditioning and justifying an optional estimation technique.

4.1.1. Restricted Ridge Estimation on Mine Data Set

To solve the multicollinearity problems, restricted ridge estimator from alternative estimators are used in this section. The ridge parameter used is $k = p/\hat{\beta}^\top \hat{\beta}$ which was originally proposed by Hoerl et al (1975) and modified to GLMs by Mackinnon (1986). The ridge parameter k at final iteration is computed as 0.3871427.

By considering the initial values of the estimators, we choose four alternative restrictions

(i) $2\beta_1 + \beta_3 = 0$ where the sum of the effects of the third explanatory variable and the two times of the first explanatory variable on the linear predictor is zero ($R^\top = \begin{bmatrix} 0 & 2 & 0 & 1 & 0 \end{bmatrix}$, $r = 0$),

(ii) $2\beta_1 + \beta_3 = 0.01$ where the sum of the effects of the third explanatory variable and the two times of the first explanatory variable on the linear predictor is 0.01. Here, the number 0.01 is arbitrarily chosen to make difference from the restriction given by (i) ($R^\top = \begin{bmatrix} 0 & 2 & 0 & 1 & 0 \end{bmatrix}$, $r = 0.01$),

(iii) $\beta_0 + \beta_1 + \beta_4 = -4$ which means that the sum of the effects of the constant, first and fourth explanatory variables on the linear predictor is -4 ($R^\top = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \end{bmatrix}$, $r = -4$),

(iv) $\beta_0 + \beta_1 + \beta_4 = -4.5$ which means that sum of the effects of the constant, first and fourth explanatory variables on the linear predictor is -4.5 where the number -4.5 is arbitrarily chosen to make difference

from the restriction given by (iii) ($R^\top = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \end{bmatrix}, r = -4.5$).

All the estimators are obtained iteratively. The OLS estimator $\hat{\beta}_{ols} = (X^\top X)^{-1} X^\top y$ which is calculated as $\hat{\beta}^{(0)} = (-4.579885,\ -0.001896,\ 0.104574,\ -0.007993,\ -0.050250)^\top$ is used as an initial value of β in computing IRLS estimator. The ridge estimator $\hat{\beta}_{ridge} = (X^\top X + kI)^{-1} X^\top y$ is used as an initial value of β in computing ridge estimator in GLMs and calculated as $\hat{\beta}(k)^{(0)} = (-3.128176,\ -0.001924,\ 0.088088,\ -0.011036,\ -0.049607)^\top$. Furthermore, the restricted ML and restricted ridge estimators

$$\hat{\beta}_{rest} = (X^\top X)^{-1} X^\top y + (X^\top X)^{-1} R^\top \left\{ R(X^\top X)^{-1} R^\top \right\}^{-1} \left\{ r - R(X^\top X)^{-1} X^\top y \right\},$$

$$\begin{aligned} \hat{\beta}_{rest.ridge} &= (X^\top X + kI)^{-1} X^\top y + (X^\top X + kI)^{-1} R^\top \left\{ R(X^\top X + kI)^{-1} R^\top \right\}^{-1} \\ &\quad \times \left\{ r - R(X^\top X + kI)^{-1} X^\top y \right\} \end{aligned}$$

are used as initial values of β, respectively in computing restricted ML and restricted ridge estimators in GLMs. Since these initial values change according to the restrictions, their initial values are not given. We use 1×10^{-6} (sufficiently close to zero) as a convergence criterion. If the sum of absolute difference for the parameter estimates between the iterations is smaller than 1×10^{-6}, the iterations stop. All methods converged at 7th iterations. There may be moderate differences between R calculations and other calculations.

Remark. When we change the initial values, i.e. all the initial values are taken as the OLS, the results did not changed. This is because all the estimators are iteratively estimated.

Parameter estimates, corresponding sMSE values and percentage change in coefficients are given in Table 4.1. Table 4.1 indicates that estimators behave dif-

Table 4.1. Parameter and sMSE estimates and percentage change in coefficients for different restrictions (Mine Data Set)

Estimator	Rest.	β_0	β_1	β_2	β_3	β_4	sMSE
$\hat{\beta}$	—	-3.593090	-0.001407	0.062346	-0.002080	-0.030813	1.052477
$\hat{\beta}_k$	—	-2.597776	-0.001326	0.050990	-0.003694	-0.028931	1.414388
$\hat{\beta}_r$	(i)	-3.858815	-0.001285	0.062424	0.002570	-0.033506	1.042726
	(ii)	-4.414579	-0.000928	0.062362	0.011857	-0.041991	1.715279
	(iii)	-3.966935	-0.001442	0.066644	-0.001540	-0.031622	0.140334
	(iv)	-4.465799	-0.001494	0.072401	-0.000870	-0.032706	0.762265
$\hat{\beta}_r(k)$	(i)	-2.864119	-0.001152	0.050146	0.002305	-0.032354	0.924172
	(ii)	-3.278353	-0.000752	0.048342	0.011505	-0.041021	0.539154
	(iii)	-3.966539	-0.001441	0.066659	-0.001529	-0.032019	0.140038
	(iv)	-4.465348	-0.001493	0.072418	-0.000857	-0.033158	0.761466
Change%							sMSE Diff.
b and $b(k)$	—	$\downarrow$ 27.7008%	$\downarrow$5.7518%	$\downarrow$18.2140%	$\uparrow$77.5628%	$\downarrow$6.1096%	0.361911
b_R and $b_R(k)$	(i)	$\downarrow$25.7772%	$\downarrow$10.3303%	$\downarrow$19.6678%	$\downarrow$10.3303%	$\downarrow$3.4387%	0.118554
	(ii)	$\downarrow$25.7380%	$\downarrow$18.9227%	$\downarrow$22.4809%	$\downarrow$2.9641%	$\downarrow$2.3110%	1.176125
	(iii)	$\downarrow$0.0099%	$\downarrow$0.0733%	$\uparrow$0.0222%	$\downarrow$0.7446%	$\uparrow$1.2542%	0.000296
	(iv)	$\downarrow$0.0100%	$\downarrow$0.0644%	$\uparrow$0.0239%	$\downarrow$1.4993%	$\uparrow$1.3806%	0.000799

ferently with respect to the model parameters and restrictions. The restricted ridge estimator is superior to the other estimators for all restrictions.

The superiority of the restricted ridge estimator over restricted ML estimator depends on the restrictions. The maximum difference between sMSE values of restricted ridge and restricted ML estimators exists in the restriction (ii). Under restrictions (iii) and (iv), restricted ridge estimator is slightly better than the restricted ML estimator in sense of sMSE criterion. Under all restrictions, restricted ridge estimator performs better than the other estimators.

Table 4.2. The change in percentage of the coefficients between restrictions for the restricted ML and restricted ridge estimators corresponding to different restrictions (Mine Data Set)

		Change % in Coef.				
Estimator	Rest.	β_0	β_1	β_2	β_3	β_4
$\hat{\beta}_r$	(i) and (ii)	↑14.4024%	↓27.7543%	↓0.0997%	↑361.2074%	↑25.3237%
	(iii) and (iv)	↑2.8018%	↑12.2055%	↑6.7609%	↓159.9313%	↓5.6229%
$\hat{\beta}_r(k)$	(i) and (ii)	↑14.4628%	↓34.6771%	↓3.5980%	↑399.0948%	↑26.7873%
	(iii) and (iv)	↑38.4907%	↑25.0403%	↑32.9290%	↓166.3379%	↓1.0360%

The percentage changes in the coefficients are also given in Table 4.1. The percentage change in $\hat{\beta}$s are calculated by the formula $\hat{\beta}_{change} = [(\hat{\beta}(new)_i - \hat{\beta}(old)_i)/\hat{\beta}(old)_i)] \times 100$ where $\hat{\beta}$ shows any estimator. The maximum change in the percentage of coefficients is between the IRLS and ridge estimators (approximately 77.6%) and the minimum change in the percentage of coefficients is between the restricted ML and restricted ridge estimators under restriction (iii) (approximately 0.0099). The maximum percentage change is between restricted ML and restricted ridge estimators under restrictions (i) and (ii).

The changes in percentage of the coefficients between restrictions for the restricted ML and restricted ridge estimators for the restrictions (i)-(ii) and (iii)-(iv) are given in Table 4.2. The change in the percentage of coefficients under the restrictions (i) and (ii) is maximum for the restricted ML and restricted ridge estimator. The change in the percentage of coefficients under the restrictions (i) and (ii) is minimum for the restricted ML and restricted ridge estimators. The sMSE values of the estimators corresponding to k values in the interval $(0,1)$ are given in Figure 4.1. Figure 4.1 suggests that the ridge estimator has smaller sMSE value than the IRLS estimator when k is smaller than approximately 0.24 under the restriction (i). The restricted

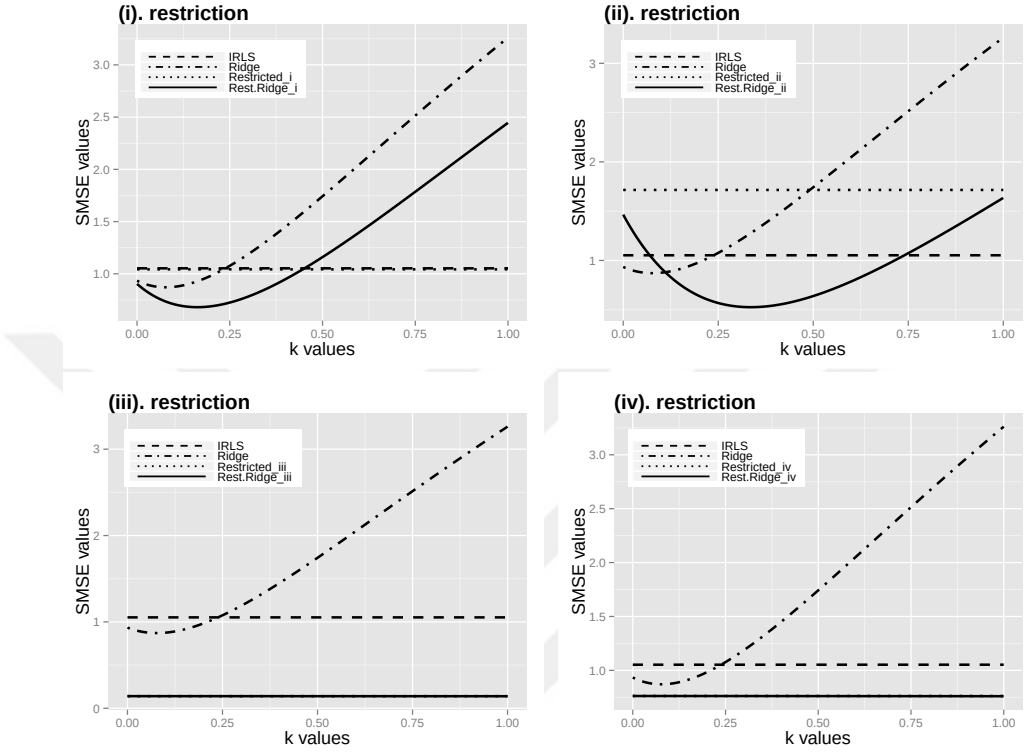

Figure 4.1. sMSE values of the IRLS, ridge, restricted ML and restricted ridge estimators versus k for different restrictions (Mine Data Set)

ridge estimator has smaller sMSE value than the IRLS estimator when k is smaller than approximately 0.45 under the restriction (i). On the other hand, as k moves away from 0.24, the difference between the sMSE values of the ridge and IRLS estimators increase. Figure 4.1 also supports the results of Table 4.1. In other words, since $k = 0.3871427$ is larger than 0.24, the IRLS estimator has smaller sMSE value than the ridge estimator. Furthermore, under the restriction (ii), the restricted ridge estimator has smaller sMSE value than the IRLS estimator when k is approximately between 0.07 and 0.7375. As k moves away from 0.7375, the difference between the sMSE values of the restricted ridge and IRLS estimators increase. Finally, under the

restrictions (iii) and (iv), the restricted ML and restricted ridge estimators give almost the sMSE values and these estimators are superior to the IRLS and ridge estimators for all k values under the restrictions (iii) and (iv).

4.1.2. Restricted Liu Estimation on Mine Data Set

To make comparisons, all the estimators are obtained by the first-order approximation. The OLS estimator $\hat{\beta}_{ols} = (X^\top X)^{-1} X^\top y$ is used as an initial value of β in computing the estimators and calculated as $\hat{\beta}^{(0)} = (-4.579885,\ -0.001896,\ 0.104574,\ -0.007993,\ -0.050250)^\top$. The Liu-biasing parameter d is computed as $\hat{d}_h$ which is proposed in Section (3.3.2.3). Therefore, according to the conditions given by Theorem 3.13, $\hat{d}_h$ value is arbitrarily chosen as 0.95.

The values of β affect the value of $R\beta - r$ which measures the relative performances of the estimators. Therefore, to see how restrictions affect the sMSE values, four different types of restrictions are used corresponding to:

i) $4\beta_1 - \beta_3 = 0$ where the effect of the third explanatory variable on the linear predictor is four times the first explanatory variable ($R^\top = \begin{bmatrix} 0 & 4 & 0 & -1 & 0 \end{bmatrix}$, $r = 0$),

ii) $4\beta_1 - \beta_3 = 0.01$ where the effect of the difference between four times the first explanatory variable and the third explanatory variable on the linear predictor is 0.01. Here, the number 0.01 is arbitrarily chosen to make difference from the restriction given by (i) ($R^\top = \begin{bmatrix} 0 & 4 & 0 & -1 & 0 \end{bmatrix}$, $r = 0.01$),

iii) $\beta_0 + \beta_1 + \beta_4 = -5$ which means that sum of the effects of the constant, first and fourth explanatory variables on the linear predictor is -5 ($R^\top = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \end{bmatrix}$, $r = -5$),

iv) $\beta_0+\beta_1+\beta_4=-5.5$ which means that sum of the effects of the constant, first and fourth explanatory variables on the linear predictor is -5.5 where the number -5.5 is arbitrarily chosen to make difference from the restriction given by (iii) ($R^\top=\begin{bmatrix}1&1&0&0&1\end{bmatrix}$, $r=-5.5$).

The estimated parameter values, sMSE values and the change percentage in coefficients are given in Table 4.3. The percentage change in βs are calculated as $\hat{\beta}_{change}=[\{\hat{\beta}(new)_i-\hat{\beta}(old)_i\}/\hat{\beta}(old)_i]\times 100\%$ where $\hat{\beta}$ shows any estimators. In computing the sMSE values, unknown β parameter is replaced by $\hat{\beta}^{(1)}$ which is approximately unbiased (this idea is similar to those which are done in linear regression model). The estimators behave differently with respect to the model parameters and restrictions. For example, when the estimators are compared under restrictions (i)-(iii), it is observed that $\text{sMSE}[\hat{\beta}_r(d)^{(1)}]<\text{sMSE}[\hat{\beta}_r^{(1)}]<\text{sMSE}[\hat{\beta}_d^{(1)}]<\text{sMSE}[\hat{\beta}^{(1)}]$ under restriction (i), $\text{sMSE}[\hat{\beta}_d^{(1)}]<\text{sMSE}[\hat{\beta}^{(1)}]<\text{sMSE}[\hat{\beta}_r(d)^{(1)}]<\text{sMSE}[\hat{\beta}_r^{(1)}]$ under restriction (ii), $\text{sMSE}[\hat{\beta}_r(d)^{(1)}]<\text{sMSE}[\hat{\beta}_r^{(1)}]<\text{sMSE}[\hat{\beta}_d^{(1)}]<\text{sMSE}[\hat{\beta}^{(1)}]$ under restriction (iii) and $\text{sMSE}[\hat{\beta}_d^{(1)}]<\text{sMSE}[\hat{\beta}^{(1)}]<\text{sMSE}[\hat{\beta}_r^{(1)}]<\text{sMSE}[\hat{\beta}_r(d)^{(1)}]$ under restriction (iv).

There absolutely exists a biased estimator which outperforms $\hat{\beta}^{(1)}$. It is observed that which estimator is superior than the other depends on the restriction and the Liu-biasing parameter used. Under the restrictions (iii) and (iv), the first-order approximated restricted ML and the first-order approximated restricted Liu estimators give almost the same sMSE values. As seen from Table 4.3, under the restrictions (i) and (iii), the first-order approximated restricted ML and first-order approximated restricted Liu estimators have smaller sMSE value than the first-order approximated ML and first-order approximated Liu estimators.

The change in the percentage of coefficients between restrictions (i) and (ii)

Table 4.3. The parameter estimates and sMSE values of the estimators when $d = 0.95$ (Mine Data Set)

Estimator	*Rest.*	β_0	β_1	β_2	β_3	β_4	sMSE
$\hat{\beta}^{(1)}$	-	-4.909674	-0.001789	0.097199	-0.007045	-0.047203	0.232122
$\hat{\beta}_d^{(1)}$	-	-4.863429	-0.001782	0.096665	-0.007101	-0.047090	0.229910
$\hat{\beta}_r^{(1)}$	(i)	-4.903588	-0.001781	0.097159	-0.007124	-0.047166	0.209601
	(ii)	-4.366989	-0.001020	0.093567	-0.014083	-0.043899	0.504145
	(iii)	-4.950893	-0.001796	0.097676	-0.006995	-0.047310	0.001781
	(iv)	-5.449518	-0.001871	0.103437	-0.006389	-0.048609	0.291556
$\hat{\beta}_r(d)^{(1)}$	(i)	-4.862113	-0.001780	0.096658	-0.007122	-0.047080	0.172544
	(ii)	-4.418463	-0.001021	0.094188	-0.014086	-0.044005	0.411639
	(iii)	-4.950882	-0.001796	0.097676	-0.006995	-0.047321	0.001780
	(iv)	-5.449489	-0.001871	0.103438	-0.006387	-0.048639	0.291524
Change %							
$\hat{\beta}^{(1)}$ and $\hat{\beta}_d^{(1)}$	-	↓0.94%	↓0.38%	↓0.54%	↑0.79%	↓0.23%	
$\hat{\beta}_r^{(1)}$	(i)-(ii)	↓10.94%	↓42.68%	↓3.69%	↑97.67%	↓6.92%	
	(iii)-(iv)	↑10.07%	↑4.19%	↑5.89%	↓8.66%	↑2.74%	
$\hat{\beta}_r(d)^{(1)}$	(i)-(ii)	↓9.12%	↓42.62%	↓2.55%	↑97.78%	↓6.53%	
	(iii)-(iv)	↑10.07%	↑4.19%	↑5.89%	↓8.68%	↑2.78%	

is maximum for the first-order approximated restricted ML and first-order approximated restricted Liu estimators (approximately 97%). The change in the percentage of coefficients between restrictions (i) and (ii) is minimum for the first-order approximated ML and first-order approximated Liu estimators (approximately 0.2%). In both the first-order approximated restricted ML and first-order approximated restricted Liu estimators, the changes in percentage of the coefficients for the restrictions (iii)-(iv) give the same amount of change in percentage. To see the sMSE behavior of the first-order approximated Liu, the first-order approximated ML, the first-order approximated restricted ML and first-order approximated restricted Liu estimators in GLMs

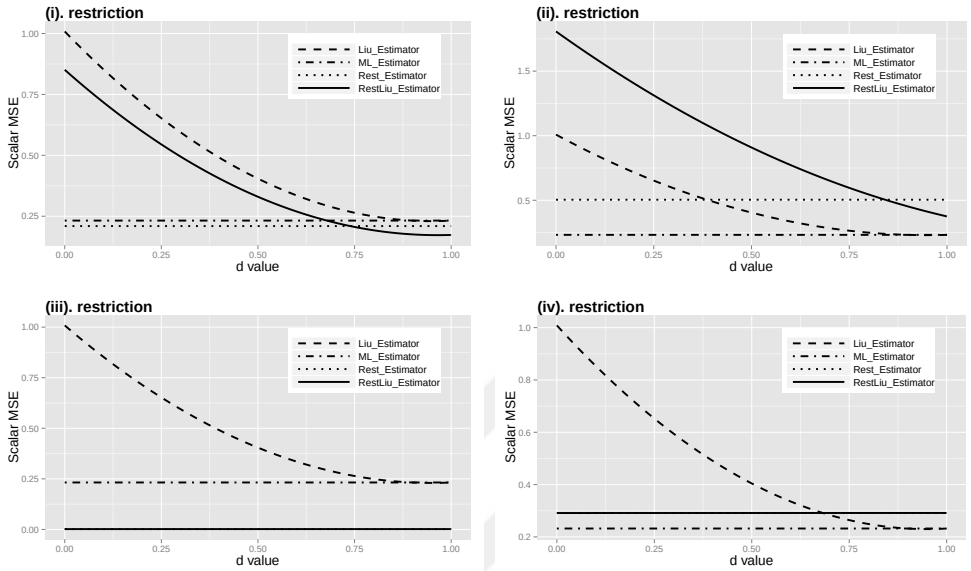

Figure 4.2. sMSE values of the first-order approximated estimators versus d under different restrictions (Mine Data Set)

for values of d in the interval $(0,1)$, Figure 4.2 is given.

To see the changes of the first-order approximated Liu estimator more clearly in the sMSE values under the restriction (i), we have narrowed the interval of d to $(0.7 < d < 1)$ in Figure 4.3. Figure 4.3 clearly indicates that as d moves away from zero, sMSE of the first-order approximated Liu is decreasing up to a d value and then slightly increasing as d approaches to one. Figure 4.3 suggests that the first-order approximated Liu estimator has smaller sMSE value than the first-order approximated ML estimator when d is larger than approximately 0.90. As d moves away from 0.90 to 0, the difference between the sMSE values of the first-order approximated Liu and ML estimators increases where first-order approximated ML estimator gets better than first-order approximated Liu estimator. Figure 4.2 also supports the results of Table 4.3. In other words, since $d = 0.95$ is larger than 0.90, the first-order approximated restricted Liu estimator has less sMSE value than the others. As seen from

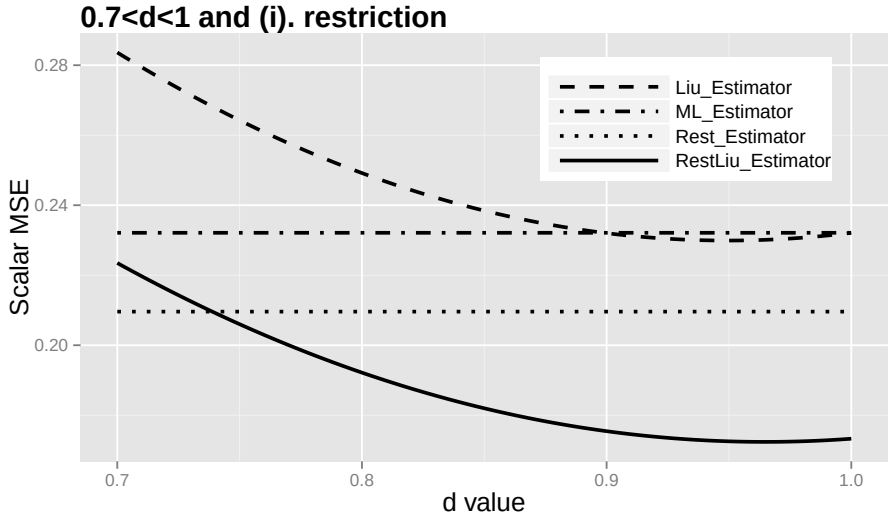

Figure 4.3. sMSE values of the first-order approximated estimators versus $0.7 < d < 1$ under the restriction (i) (Mine Data Set)

Figure 4.2 that under the restrictions (iii) and (iv), sMSE values of the first-order approximated restricted ML and restricted Liu estimators give almost same (see Table 4.3 for $d = 0.95$).

4.2. Example 2: Weather Data Set

In this numerical example, the response has gamma distribution with reciprocal link. The data set was first analyzed by Chatterjee and Hadi (1988). The data set correspond to the weather factors and nitrogen dioxide concentrations (y), in parts per hundred million (p.p.h.m.), for 26 days in September 1984 as measured at a monitoring station in the San Francisco Bay area. There are four explanatory variables. The variables considered in the study are: mean wind speed in miles per hour (x_1) in m.p.h., maximum temperature (x_2) in $^\circ F$, insolation (x_3) in langleys per day and stability factor (x_4) in $^\circ F$. Our computations here are performed by using R. Chatterjee and Hadi (1988) have shown that this data set has a gamma distribution.

The gamma model with reciprocal link is $\mu_i = (\sum_{j=1}^{q} \beta_j x_{ij})^{-1}$. The Pearson estimate of φ is obtained as $\hat{\varphi}^2 = 0.07572852$.

The constant term is added to the X matrix. The eigenvalues of the $X^\top \hat{W} X$ matrix where $\hat{W}$ is obtained at the final iteration of the IRLS method are obtained as $\lambda_1 = 1240.428467$, $\lambda_2 = 72.912574$, $\lambda_3 = 45.938929$, $\lambda_4 = 19.375420$, and $\lambda_5 = 5.801554$. Thus, the condition number which is computed as $\lambda_{\max}/\lambda_{\min} = 213.8097$ indicates that multicollinearity exists.

To compute the IRLS estimator, we assigned Hardin and Hilbe's (2012) initial values $\mu_i^{(0)} = (y + \bar{y})/2$. Then we define the weight matrix W from $(W^{(0)})^{-1} = (d\eta/d\mu)^2 V^{(0)}$ and the working response z with the ith observation $z_i^{(0)} = 1/\mu_i^{(0)} - (y_i - \mu_i^{(0)})/(\mu_i^{(0)})^2$ for the gamma distribution with reciprocal link. Finally, the tolerance is set to 1×10^{-6} for the difference of the current estimated coefficients from previous iteration.

4.2.1. Restricted Ridge Estimation on Weather Data Set

The gamma model with reciprocal link is $\mu_i = 1/(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2j} + \beta_3 x_{3j} + \beta_4 x_{4j})$. The Pearson estimate of $a(\phi)$ is obtained as $\hat{\varphi}^2 = 0.1223818$. Then we define the weight matrix W from $\hat{W}^{-1} = (d\eta/d\mu)^2 V$ and the working-response z with the ith observation $z_i = 1/\mu_i - (y_i - \mu_i)/\mu_i^2$ for the gamma distribution with reciprocal link.

The ridge parameter used is $k = p\hat{\varphi}^2/\hat{\beta}^\top \hat{\beta}$ which was originally proposed by Hoerl et al (1975) and modified to GLMs by Mackinnon (1986). The ridge parameter k at final iteration is computed as 15.65262.

All the estimators are obtained iteratively. The OLS estimator $\hat{\beta}_{ols} = (X^\top X)^{-1} X^\top y$ is used as an initial value of β in computing IRLS estimator. The ridge estimator $\hat{\beta}_{ridge} = (X^\top X + kI)^{-1} X^\top y$ is used as an initial value of β in computing

ridge estimator in GLMs. Furthermore, the restricted

$$\hat{\beta}_{rest} = (X^\top X)^{-1}X^\top y + (X^\top X)^{-1}R^\top \left\{R(X^\top X)^{-1}R^\top\right\}^{-1}\left\{r - R(X^\top X)^{-1}X^\top y\right\},$$

and the restricted ridge estimators

$$\begin{aligned}\hat{\beta}_{rest.ridge} &= (X^\top X + kI)^{-1}X^\top y + (X^\top X + kI)^{-1}R^\top \left\{R(X^\top X + kI)^{-1}R^\top\right\}^{-1} \\ &\quad \times \left\{r - R(X^\top X + kI)^{-1}X^\top y\right\}\end{aligned}$$

are used as initial values of β, respectively in computing restricted ML and restricted ridge estimators in GLMs. In this study, we used $\left\|\hat{\beta}^{(m+1)} - \hat{\beta}^{(m)}\right\| < \xi$ as a convergence criterion where ξ is some prechosen small number. If the sum of absolute difference for the parameter estimates between the iterations is smaller than ξ, the iterations stop. The restrictions are chosen as follows:

(i) $\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$ which means that the sum of the effects of explanatory variables and the constant on the linear predictor is zero ($R^\top = [1\ 1\ 1\ 1\ 1]$, $r = 0$),

(ii) $\beta_1 + \beta_2 + \beta_3 = 0$ which means that sum of the effects of the explanatory variables on the linear predictor is 0 ($R^\top = [0\ 1\ 1\ 1\ 0]$, $r = 0$),

(iii) $\beta_1 + \beta_2 + \beta_3 = 0.01$ which means that sum of the effects of the explanatory variables on the linear predictor is 0.01 where the number 0.01 is arbitrarily chosen to make difference from the restriction given by (ii) ($R^\top = [0\ 1\ 1\ 1\ 0]$, $r = 0.01$).

The estimated parameter values, the corresponding sMSE values and the change percentage in coefficient are given in Table 4.4. Table 4.4 indicates that the estimators behave differently with respect to the model parameters and restrictions. The superiority of the restricted ridge estimator over restricted ML estimator

Table 4.4. The parameter, sMSE estimates, change percentage in coefficient and sMSE difference for different restrictions (Weather Data Set)

Estimator	Rest	β_0	β_1	β_2	β_3	β_4	sMSE
$\hat{\beta}$	-	0.197099	0.015407	-0.002656	0.000102	-0.000625	0.063386
$\hat{\beta}_k$	-	0.022420	0.019970	-0.001111	0.000110	-0.001356	0.031235
$\hat{\beta}_r$	(i)	-0.018852	0.021036	-0.000744	0.000113	-0.001552	0.046682
	(ii)	0.519704	0.005326	-0.005466	0.000140	0.000494	0.115404
	(iii)	0.267775	0.013160	-0.003270	0.000110	-0.000363	0.013164
$\hat{\beta}_r(k)$	(i)	-0.018762	0.020957	-0.000740	0.000114	-0.001568	0.046639
	(ii)	0.243387	0.002164	-0.002571	0.000407	-0.001820	0.003871
	(iii)	0.126128	0.011579	-0.001823	0.000244	-0.001362	0.007502
Change%							mse Diff.
$\hat{\beta}$ and $\hat{\beta}_k$	-	↓88.624558%	↑29.615172%	↓58.169699%	↑7.707434%	↑117.046017%	0.032151
$\hat{\beta}_r$ and $\hat{\beta}_r(k)$	(i)	↓0.4740957%	↓0.3781810%	↓0.5371638%	↑1.4342914%	↑0.9948083%	0.000043
	(ii)	↓53.16803%	↓59.36270%	↓52.95300%	↑190.58765%	↓468.13301%	0.111533
	(iii)	↓52.89766%	↓12.01259%	↓44.24286%	↑121.68284%	↑274.40892%	0.005662

depends on the restrictions. The restricted ridge estimator is superior to the other estimators for the restrictions (ii)-(iii) in the sense of sMSE criterion. Table 4.4 shows that the restrictions affect the sMSE performance of the estimators. When the results in Table 4.4 are considered, it is observed that under the restriction (i), the ordering $\text{sMSE}[\hat{\beta}_k] < \text{sMSE}[\hat{\beta}_r(k)] < \text{sMSE}[\hat{\beta}_r] < \text{sMSE}[\hat{\beta}]$ exists. This ordering changes to $\text{sMSE}[\hat{\beta}_r(k)] < \text{sMSE}[\hat{\beta}_k] < \text{sMSE}[\hat{\beta}] < \text{sMSE}[\hat{\beta}_r]$ under the restriction (ii), and $\text{sMSE}[\hat{\beta}_r(k)] < \text{sMSE}[\hat{\beta}_r] < \text{sMSE}[\hat{\beta}_k] < \text{sMSE}[\hat{\beta}]$ under the restriction (iii). In general, there absolutely exists a biased estimator which outperforms $\hat{\beta}$.

The percentage changes in the coefficients are also given in Table 4.4. The percentage change in $\hat{\beta}$s are calculated as $\hat{\beta}_{change} = [(\hat{\beta}(new)_i - \hat{\beta}(old)_i)/\hat{\beta}(old)_i)] \times 100\%$.[1] The maximum change in the percentage of coefficients is between the restricted ML and restricted ridge estimators (approximately 468.13%) for the re-

[1] $\hat{\beta}$ shows any estimator.

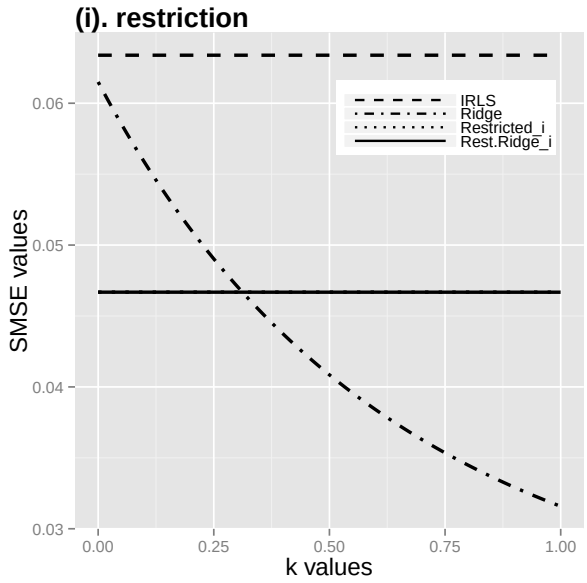

Figure 4.4. sMSE values of the estimators versus k under the restriction (i) (Weather Data Set)

striction (ii) and the minimum change in the percentage of coefficients is between the restricted ML and restricted ridge estimators (approximately 0.37%) under the restriction (i).

To consider the effect of the restrictions and the biasing parameter k on the estimators, the sMSE values of the estimators corresponding to k values in the interval $(0, 1)$ are given in Figures 4.4 - 4.5 - 4.6. As seen from Figure 4.4, the restricted ridge estimator has smaller sMSE value than the ridge estimator when k is smaller than approximately 0.31 under the restriction (i). When k is larger than 0.31, the situation is opposite. On the other hand, as k moves away from 0.31, the difference between the sMSE values of the ridge and restricted ridge estimators increase.

Figure 4.5 suggests that under the restriction (ii), the restricted ridge estimator has smaller sMSE value than the restricted ML estimator when k is larger than

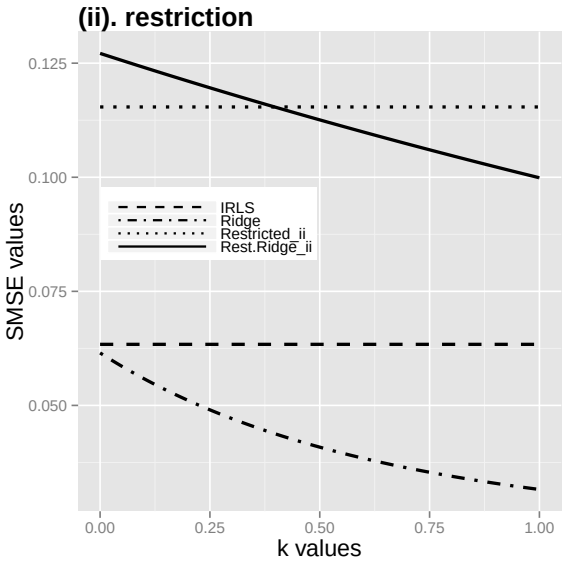

Figure 4.5. sMSE values of the estimators versus k under the restriction (ii) (Weather Data Set)

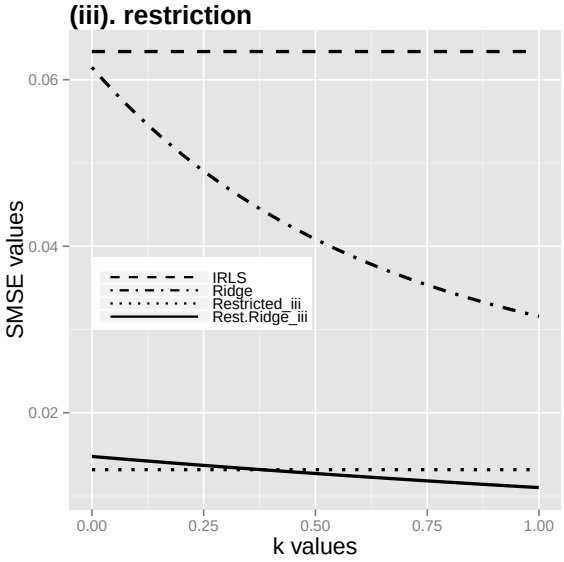

Figure 4.6. sMSE values of the estimators versus k under the restriction (iii) (Weather Data Set)

approximately 0.40 and the ridge estimator outperforms than the other estimators at this restriction. Figure 4.6 suggests that under the restriction (iii), the restricted ridge estimator has smaller sMSE value than the restricted estimator when k is larger than approximately 0.38.

4.2.2. Liu Estimation on Weather Data Set

To solve the multicollinearity problems, Liu estimator from alternative estimators will be used in this section. The sMSE values of the first-order approximated ML and first-order approximated Liu estimators versus d are plotted in Figure 4.7. To see the changes more clearly in the sMSE values, we have narrowed the interval of d on the right side. Figure 4.7 indicates that as d moves away from zero, sMSE$(\hat{\alpha}_d)$ is decreasing up to some value on d and is then slightly increasing as d approaches to one. As seen from Figure 4.7 that there exists a d value in the interval $(0,1)$, say d^*, such that the first-order approximated ML estimator is superior to the first-order approximated Liu estimator in the interval $(0,d^*)$ and the first-order approximated Liu estimator is superior to the first-order approximated ML estimator in the interval $(d^*,1)$. It is observed that this d^* value is 0.9714 for this data set which is the value that sMSE values of the first-order approximated ML and first-order approximated Liu estimators are same. Furthermore, the difference between the sMSE values of the first-order approximated Liu and first-order approximated ML estimators is decreasing as d approaches d^* from zero. Figure 4.7 is given to see the sMSE behavior of the first-order approximated Liu estimator in a GLM for values of d in the interval $(0,1)$.

Tables 4.5-4.6 present the results at a certain d value. The condition $\sum_{j=1}^{q} \lambda_{\min}(\lambda_{\min}+1)/\lambda_j(\lambda_j+1)$ in Theorem 3.13 is computed as 1.137511 which is smaller than 2. Hence, we arbitrarily choose two different d values, as $\hat{d}_1 = 0.5$

Table 4.5. The parameter and sMSE values of the first-order approximated ML and first-order approximated Liu estimators for $\hat{d}_1 = 0.5$ and $\hat{d}_2 = 0.985$ (Weather Data Set)

	First-order approximated estimator		
	ML	Liu ($\hat{d}_1$)	Liu ($\hat{d}_2$)
$\hat{\beta}_0$	0.173577366	0.173306906	0.17356925
$\hat{\beta}_1$	0.181153494	0.179613213	0.18110728
$\hat{\beta}_2$	−0.162092711	−0.160129041	−0.16203380
$\hat{\beta}_3$	0.015342965	0.015522473	0.015348350
$\hat{\beta}_4$	0.001899732	0.000409469	0.001855025
sMSE	0.018952460	0.050781940	0.018924880

Table 4.6. The parameter estimates of the IRLS and iterative Liu estimators for $\hat{d}_1 = 0.5$ and $\hat{d}_2 = 0.985$ (Weather Data Set)

	Iterative		
	IRLS	Liu ($\hat{d}_1$)	Liu ($\hat{d}_2$)
$\hat{\beta}_0$	0.1819061	0.1814188	0.1818913
$\hat{\beta}_1$	0.2099171	0.2049169	0.2097662
$\hat{\beta}_2$	−0.1160440	−0.1180465	−0.1161045
$\hat{\beta}_3$	0.0306132	0.0305095	0.0306101
$\hat{\beta}_4$	−0.0233906	−0.0226831	−0.0233691

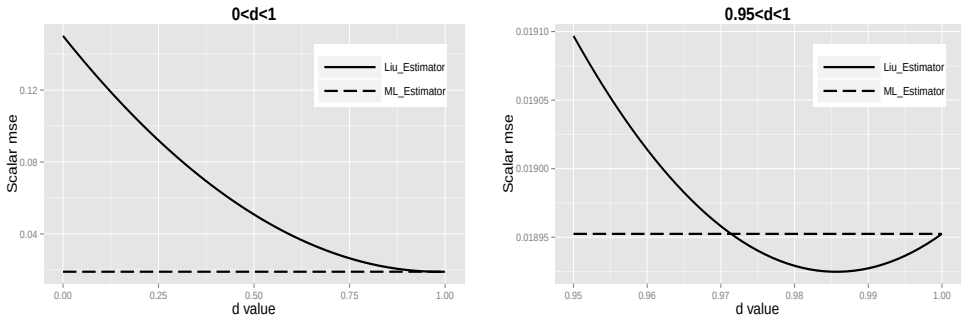

Figure 4.7. sMSE values of the first-order approximated ML and first-order approximated Liu estimators versus d (Weather Data Set)

and $\hat{d}_2 = 0.985$ as stated by Remark 3.14. The estimated parameter values and the corresponding sMSE values are given in Table 4.5 for $\hat{\beta}^{(1)}$ and $\hat{\beta}_d^{(1)}$. As seen from Table 4.5, although coefficients of $\hat{\beta}^{(1)}$, $\hat{\beta}_{\hat{d}_2}^{(1)}$ and $\hat{\beta}_{\hat{d}_1}^{(1)}$ are almost same especially $\hat{\beta}^{(1)}$ and $\hat{\beta}_{\hat{d}_2}^{(1)}$, the sMSE value of $\hat{\beta}_{\hat{d}_2}^{(1)}$ is smaller than that of $\hat{\beta}^{(1)}$ and $\hat{\beta}_{\hat{d}_1}^{(1)}$. The estimated parameter values are given in Table 4.6 for the IRLS and iterative Liu estimators. Since the estimators are iterative so that the weights and the working responses are updated at each iteration, the sMSE values of the estimators are not calculated in Table 4.6. Numerical example shows that the d value affects the performance of the first-order approximated Liu estimator. The first-order approximated Liu estimator gains efficiency over the first-order approximated ML estimator for a properly chosen d value.

4.2.3. Restricted Liu Estimation on Weather Data Set

All the estimators are obtained by the first-order approximation. The OLS estimator $\hat{\beta}_{ols} = (X^\top X)^{-1} X^\top y$ is used as an initial value of β and calculated as $\hat{\beta}^{(0)} = (6.269231, -5.622560, 7.502692, -2.000586, -0.356589)^\top$.

Liu-biasing parameter d is computed as $\hat{d}_h$ which is proposed in Section

3.3.2.3. Therefore, according to the conditions given by Theorem 3.13, two d_h values are arbitrarily chosen as 0.25 and 0.90.

The restrictions are chosen as follows:

i) $\beta_0+\beta_1+\beta_2+\beta_3+\beta_4=0$ which means that the sum of the effects of explanatory variables and the constant on the linear predictor is zero ($R=[1\ 1\ 1\ 1\ 1]$, $r=0$),

ii) $\beta_1+\beta_2+\beta_3=0$ which means that sum of the effects of the explanatory variables on the linear predictor is 0 ($R=[0\ 1\ 1\ 1\ 0]$, $r=0$),

iii) $\beta_1+\beta_2+\beta_3=0.01$ which means that sum of the effects of the explanatory variables on the linear predictor is 0.01 where the number 0.01[2] is arbitrarily chosen to make difference from the restriction given by (ii) ($R=[0\ 1\ 1\ 1\ 0]$, $r=0.01$).

Table 4.7 shows that the restrictions and the $\hat{d}_h$ values affect the MSE performance of the estimators. In computing the sMSE values, unknown β parameter is replaced by $\hat{\beta}^{(1)}$ which is approximately unbiased (this idea is similar to those which are done in linear regression model). When the estimators are compared under the restrictions (i)-(iii), it is observed that $\text{sMSE}[\hat{\beta}(d=0.25)^{(1)}] < \text{sMSE}[\hat{\beta}(d=0.90)^{(1)}] < \text{sMSE}[\hat{\beta}^{(1)}] < \text{sMSE}[\hat{\beta}_r(d=0.90)^{(1)}] < \text{sMSE}[\hat{\beta}_r(d=0.25)^{(1)}] < \text{sMSE}[\hat{\beta}_r^{(1)}]$ under the restriction (i), $\text{sMSE}[\hat{\beta}(d=0.25)^{(1)}] < \text{sMSE}[\hat{\beta}(d=0.90)^{(1)}] < \text{sMSE}[\hat{\beta}^{(1)}] < \text{sMSE}[\hat{\beta}_r(d=0.25)^{(1)}] < \text{sMSE}[\hat{\beta}_r(d=0.90)^{(1)}] < \text{sMSE}[\hat{\beta}_r^{(1)}]$ under the restriction (ii), and $\text{sMSE}[\hat{\beta}_r(d=0.25)^{(1)}] < \text{sMSE}[\hat{\beta}_r^{(1)}] < \text{sMSE}[\hat{\beta}_r(d=0.90)^{(1)}] < \text{sMSE}[\hat{\beta}(d=0.25)^{(1)}] < \text{sMSE}[\hat{\beta}(d=0.90)^{(1)}] < \text{sMSE}[\hat{\beta}^{(1)}]$ under the restriction (iii). In general, there absolutely exists a biased estimator which outperforms $\hat{\beta}^{(1)}$. It

[2] We recognized that as the deviation from 0 increases, sMSE values of the first-order approximated Liu and first-order approximated restricted Liu estimators inflate. Therefore, 0.01 is arbitrarily choosen.

Table 4.7. The parameter estimates and sMSE values of the estimators for different d values and restrictions (Weather Data Set)

Estimator	d	Rest.	β_0	β_1	β_2	β_3	β_4	sMSE
$\hat{\beta}^{(1)}$	-	-	0.3089166	0.0132961	-0.0037102	0.0000512	0.0000507	0.0417548
$\hat{\beta}_d^{(1)}$	0.25	-	0.2267008	0.0150655	-0.0029761	0.0000668	-0.0003206	0.0292329
	0.90	-	0.2979545	0.0135320	-0.0036123	0.0005332	0.0000012	0.0389611
$\hat{\beta}_r^{(1)}$	-	(i)	-0.0181439	0.0202662	-0.0007861	0.0001155	-0.0014516	0.1070521
	-	(ii)	0.5720477	0.0059614	-0.0060111	0.0000497	0.0010802	0.0761032
	-	(iii)	0.2990107	0.0135723	-0.0036235	0.0000513	0.0000120	0.0069032
$\hat{\beta}_r(d)^{(1)}$	0.25	(i)	-0.0181264	0.0202520	-0.0007855	0.0001159	-0.0014558	0.1070405
	0.25	(ii)	0.5312684	0.0055127	-0.0056088	0.0000960	0.0007696	0.0555099
	0.25	(iii)	0.2807248	0.0133711	-0.0034431	0.0000720	-0.0001272	0.0067997
	0.90	(i)	-0.0180952	0.0202273	-0.0007844	0.0001166	-0.0014642	0.1070197
	0.90	(ii)	0.5478234	0.0056949	-0.0057722	0.0000772	0.0008957	0.0640246
	0.90	(iii)	0.2972799	0.0135532	-0.0036065	0.0000532	-0.0000011	0.0070209

is observed that which estimator is superior than the other depends on the restriction and the Liu-biasing parameter used.

To consider the effect of the restrictions and the Liu-biasing parameter on the estimators, the sMSE values of the estimators corresponding to d values in the interval $(0,1)$ are given in Figure 4.8. Because of scaling, the performance of the first-order approximated restricted ML and first-order approximated restricted Liu estimators are not obviously seen from the left side of Figure 4.8. Therefore, the sMSE values of the first-order approximated restricted ML and restricted Liu estimators versus d under all restrictions are plotted separately in the right side of Figure 4.8. As seen from Figure 4.8, as d goes from 0 to 1, the sMSE value of the first-order approximated Liu estimator increases whereas the sMSE value of the first-order approximated restricted Liu estimator decreases in linear manner under the restriction (i), increase in linear manner under the restriction (ii) and decrease and increase in a parabolic manner under the restriction (iii). Figure 4.8 suggests that under the re-

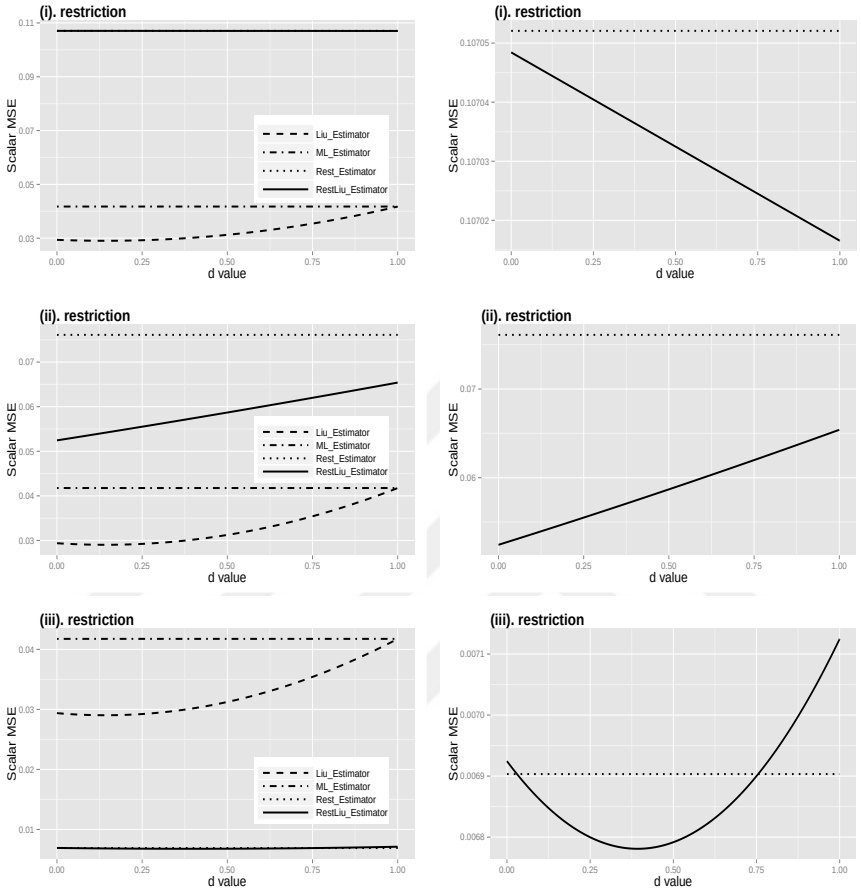

Figure 4.8. sMSE values of the first-order approximated estimators versus d under different restrictions (Weather Data Set)

striction (iii), the first-order approximated restricted ML and restricted Liu estimators perform better than the others. At restrictions (i) and (ii), the first-order approximated Liu estimator shows better performance than the other estimators which is also supported by Table 4.7.

5. MONTE CARLO SIMULATION STUDIES

In this section, we examine the performance of the IRLS, ML, restricted ML, ridge, restricted ridge, Liu and restricted Liu estimators in GLMs via simulation studies by the estimated mean squared error (EMSE) criterion for Poisson, gamma and binomial responses. Since the distribution of response, degrees of multicollinearity, sample size, number of explanatory variables, ridge parameter and Liu-biasing parameter impact the performance of the EMSE, simulation studies are conducted under various choices of these parameters. Then we comprehensively illustrate the theoretical conclusions and performances. Our Monte Carlo simulation studies are performed by using R (see, Wickham 2009) and the seed value of the generator of R is taken as 12345.

Following McDonald and Galarneau (1975), the explanatory variables are generated by

$$x_{ij} = (1-\gamma^2)^{1/2}\vartheta_{ij} + \gamma\vartheta_{i,p+1},\ i = 1,\ldots,n,\ j = 1,\ldots,p$$

where ϑ_{ij} are independent standard normal pseudo-random numbers and γ is specified so that the correlation between any two explanatory variables is given by γ^2.

5.1. Experiment 1: Simulation Study with Poisson Response

These experiments are designed for Poisson distributed response.

5.1.1. Restricted Ridge Estimation

In this section, we examine the performance of the restricted ridge to the restricted ML, ridge and IRLS estimators in GLMs via simulation studies. Since the distribution of response, degrees of multicollinearity, sample size, restriction types, and ridge parameter impact the performance of the estimators, simulation studies are

conducted under various choices of these parameters.

1. Three different sets of correlation are considered corresponding to $\gamma^2 = 0.90, 0.95$ and 0.99.

2. The explanatory variables are then standardized by using unit length scaling so that $X^\top X$ is a matrix of correlations.

3. The sample sizes are chosen as $n = 100$, 200, 300, 400, 500, 600, 700, 800, 900 and 1000, and number of explanatory variables are chosen as $p = 6$ and 10.

4. The parameter vectors corresponding to $p = 6$ and $p = 10$ are arbitrarily chosen as $\beta = (1/\sqrt{p}) \times j_{p\times 1}$ where $j_{p\times 1}$ is a $p \times 1$ vector of ones.

5. The observations Y_i are generated from a Poisson distribution

$$Y_i \sim \mathcal{P}(\mu_i),\ i = 1, \ldots, n$$

where μ_i is determined by $\mu_i = \mathrm{E}(Y_i) = \exp(\beta_1 x_{i1} + \ldots + \beta_p x_{ip})$.

6. The ridge parameter is chosen as $k = p/\hat{\beta}^\top \hat{\beta}$ where $\hat{\beta}$ is the IRLS estimator at final iteration.

7. We use the iterative procedures in computing the estimators. The initial values for the IRLS, restricted, ridge and restricted ridge estimators in GLMs are taken respectively as

$$\begin{aligned}
\hat{\beta}_{ols} &= (X^\top X)^{-1} X^\top y, \\
\hat{\beta}_{rest} &= (X^\top X)^{-1} X^\top y + (X^\top X)^{-1} R^\top \left\{ R(X^\top X)^{-1} R^\top \right\}^{-1} \left\{ r - R(X^\top X)^{-1} X^\top y \right\}, \\
\hat{\beta}_{ridge} &= (X^\top X + kI)^{-1} X^\top y, \\
\hat{\beta}_{rest.ridge} &= (X^\top X + kI)^{-1} X^\top y + (X^\top X + kI)^{-1} R^\top \left\{ R(X^\top X + kI)^{-1} R^\top \right\}^{-1} \\
&\quad \times \left\{ r - R(X^\top X + kI)^{-1} X^\top y \right\}.
\end{aligned}$$

At each iteration, W is computed as a diagonal matrix with elements of $\exp(X\beta^{(m)})$ on the main diagonal where $\beta^{(m)}$ is evaluated at each iteration. The working-response

for Poisson distribution is defined as $z^{(m)} = \log(\mu^{(m)}) + (y_i - \mu^{(m)})/\mu^{(m)}$. Several iterations are then carried out until some convergence is achieved in the resulting values. In estimating the regression parameters, the value 1×10^{-6} is used as a convergence criterion and iterations are stopped if $\left\|\beta^{(m+1)} - \beta^{(m)}\right\| > 1 \times 10^{-6}$.

8. The restrictions are chosen corresponding to $R^\top \beta = \sum_{i=1}^{p} \beta_i$ where $R^\top = j_{p\times 1}$ and $r = \sum_{i=1}^{p} \beta_i \times \kappa$. Since $\sum_{i=1}^{p} \beta_i = \sqrt{p}$ for parameter vectors, r equals $\sqrt{p} \times \kappa$. The usage of κ allows one to assess whether the restriction is true or not. To see how the restrictions affect the EMSE values, five different types of restrictions are used corresponding to $\kappa = 1$ where the restriction is true, $\kappa = 1.05$ and $\kappa = 0.95$ where 5% deviations exist and $\kappa = 0.50$ and $\kappa = 1.50$ where 50% deviations exist (see also Kaçıranlar et al, 2011).

Remark. Since we know the parameter vector in the simulation study, we arbitrarily set up the restrictions on the parameters to see what happens when the restrictions hold true or does not hold true. However, in a real data analysis it should be selected according to the effect of the explanatory variables.

9. For each choices of n, p, γ and restrictions, the experiment is replicated 2000 times by generating Y_i. The performance of the estimators is calculated in terms of EMSE

$$\text{EMSE}(\hat{\beta}) = \frac{1}{2000} \sum_{r=1}^{2000} (\hat{\beta}_{(r)} - \beta)^\top (\hat{\beta}_{(r)} - \beta)$$

where the subscript (r) refers to the rth replicate and $\hat{\beta}_{(r)}$ is the estimate of β in the rth replication of the experiment.

The simulation results are summarized by Figures 5.1-5.6. The main conclusions obtained from the simulation results are as follows:

- When the degree of collinearity becomes larger, the EMSE values of IRLS, ridge, restricted ML and restricted ridge estimators increase at all restrictions when n is fixed.

- As the deviation from $\kappa = 1$ increases, the difference between the EMSE values of the IRLS-restricted ML estimators and ridge-restricted ridge estimators increase.
- Under all the restrictions, the ridge and restricted ridge estimators perform better than the IRLS and restricted ML estimators. When the restriction is true, the IRLS and restricted ML estimators give almost same results.
- The p values have an obvious effect on the results. When p value changes from 6 to 10 and the sample size is small at $\kappa = 1 \pm 0.5$, the difference between the EMSE values of the IRLS-restricted ML estimators and ridge-restricted ridge estimators increases.
- When κ equals to 1 ± 0.5, as the sample size increases, the EMSE value of the IRLS estimator closes to that of restricted ML estimator and the EMSE value of ridge estimator closes to that of restricted ridge estimator.

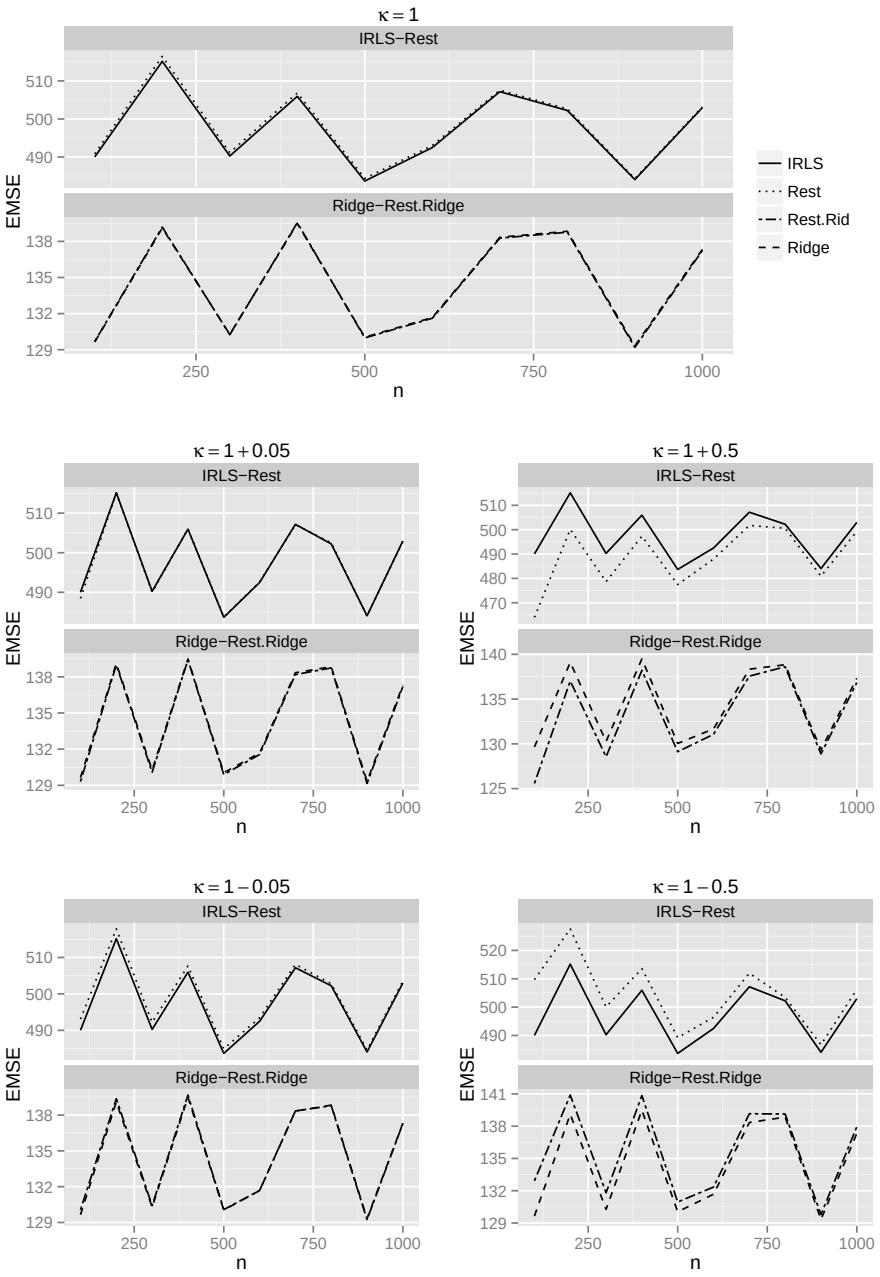

Figure 5.1. Plot of EMSE of the estimators over various n and under the different restrictions when $q = 6$, $\gamma^2 = 0.99$ and observations are from a Poisson distribution

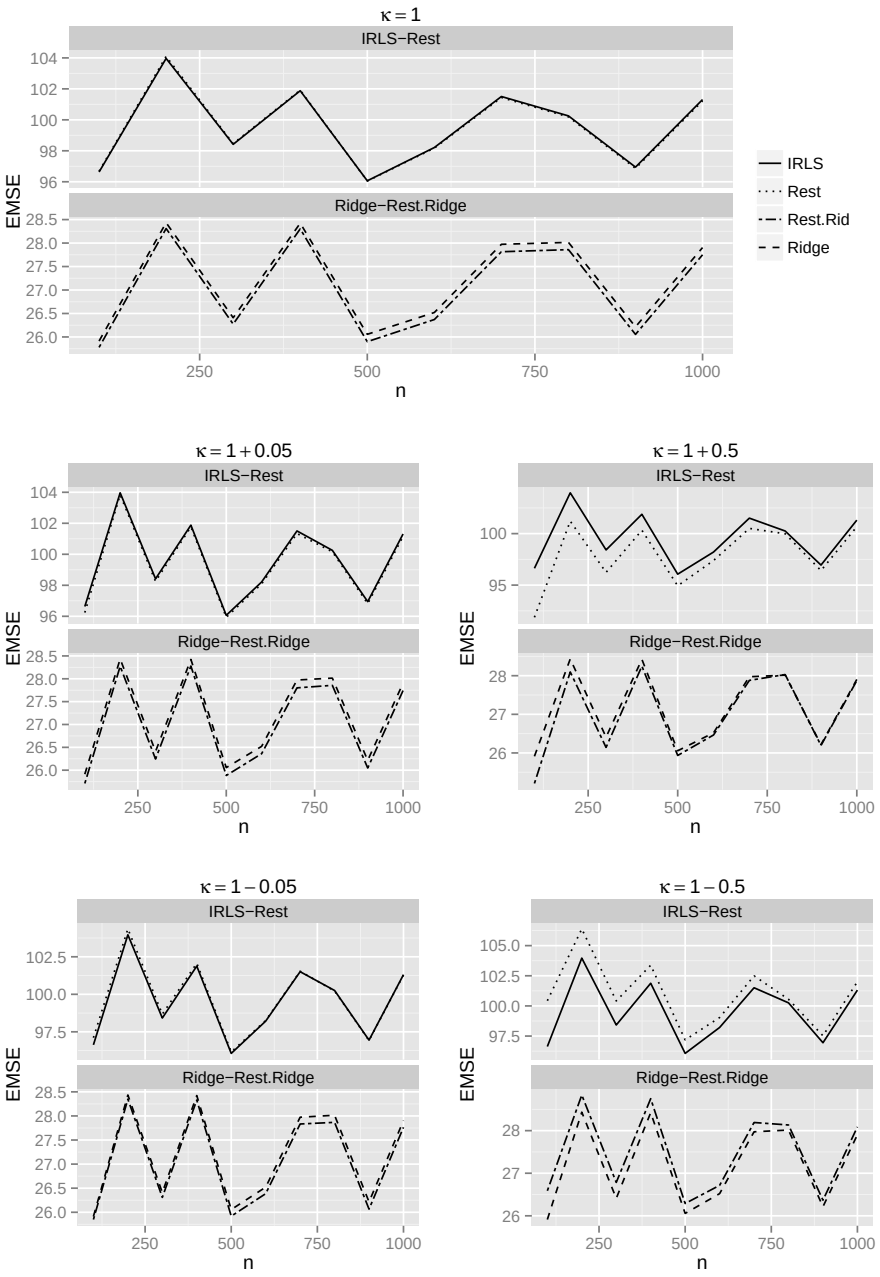

Figure 5.2. Plot of EMSE of the estimators over various n and under the different restrictions when $q = 6$, $\gamma^2 = 0.95$ and observations are from a Poisson distribution

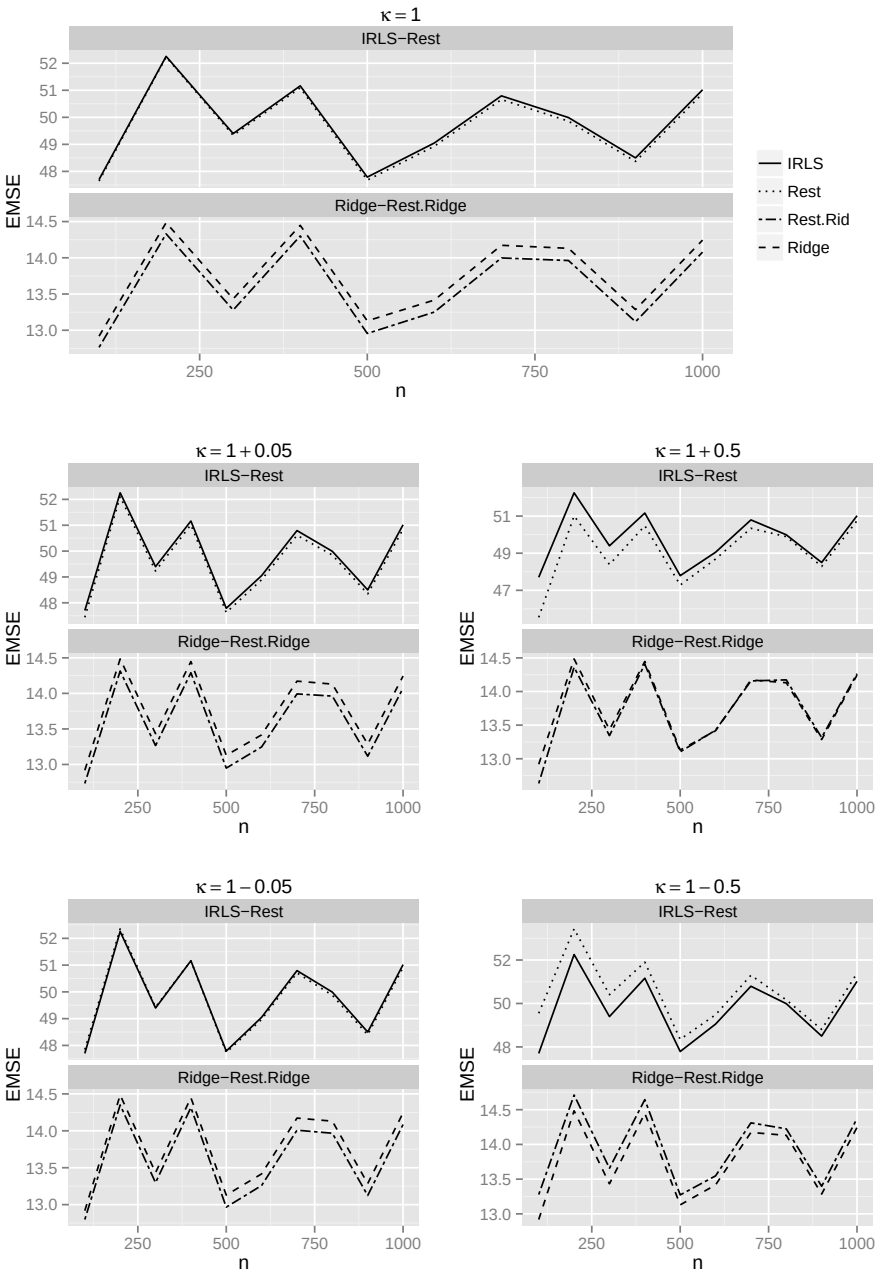

Figure 5.3. Plot of EMSE of the estimators over various n and under the different restrictions when $q = 6$, $\gamma^2 = 0.90$ and observations are from a Poisson distribution

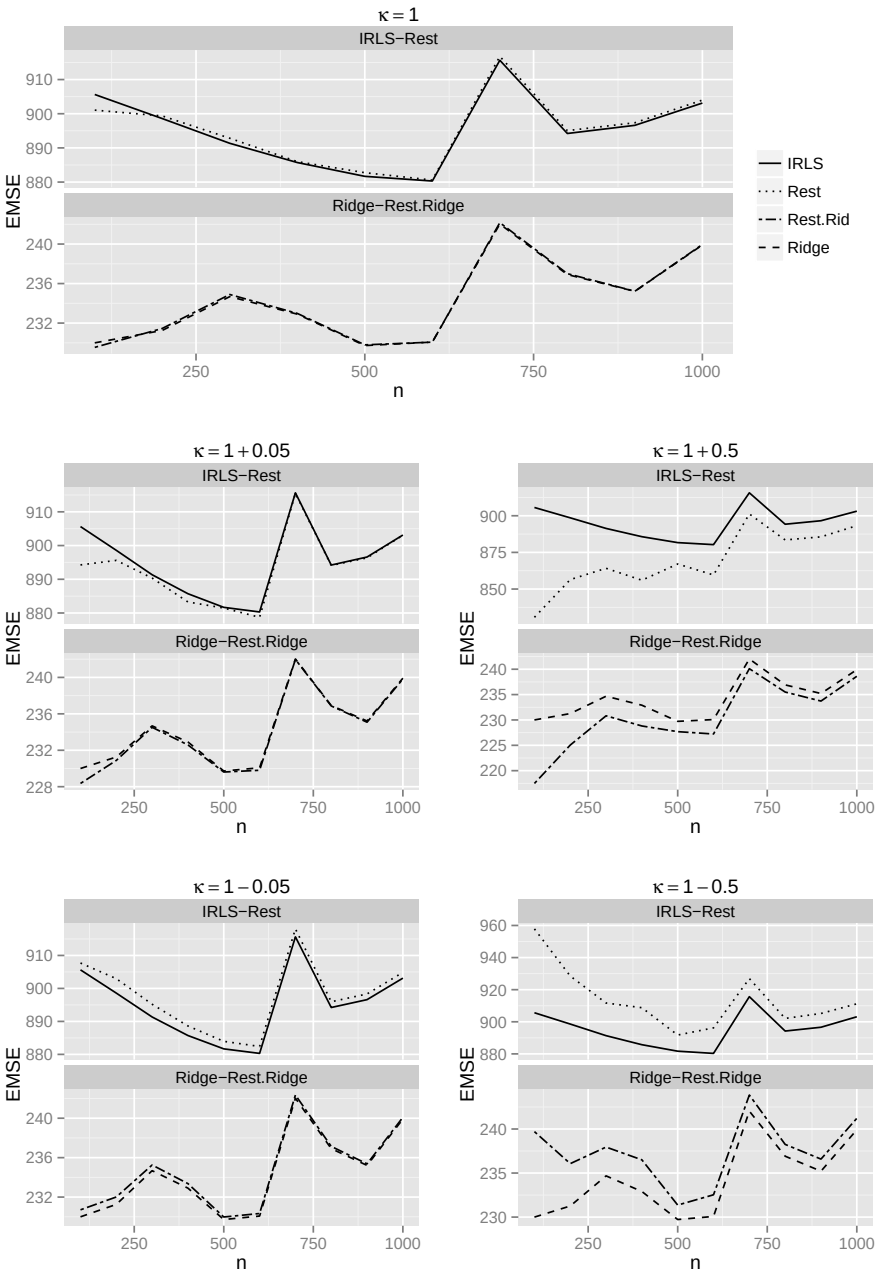

Figure 5.4. Plot of EMSE of the estimators over various n and under the different restrictions when $q = 10$, $\gamma^2 = 0.99$ and observations are from a Poisson distribution

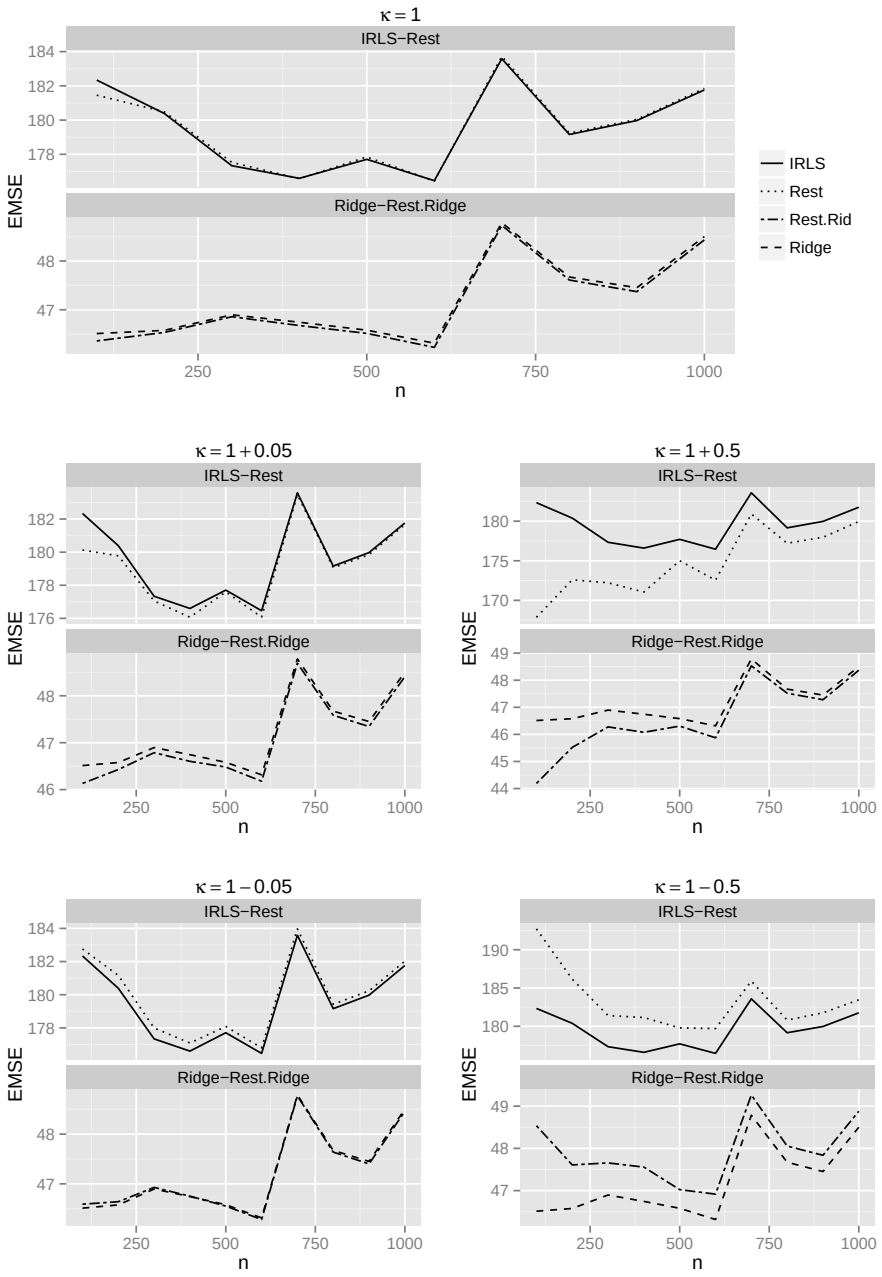

Figure 5.5. Plot of EMSE of the estimators over various n and under the different restrictions when $q = 10$, $\gamma^2 = 0.95$ and observations are from a Poisson distribution

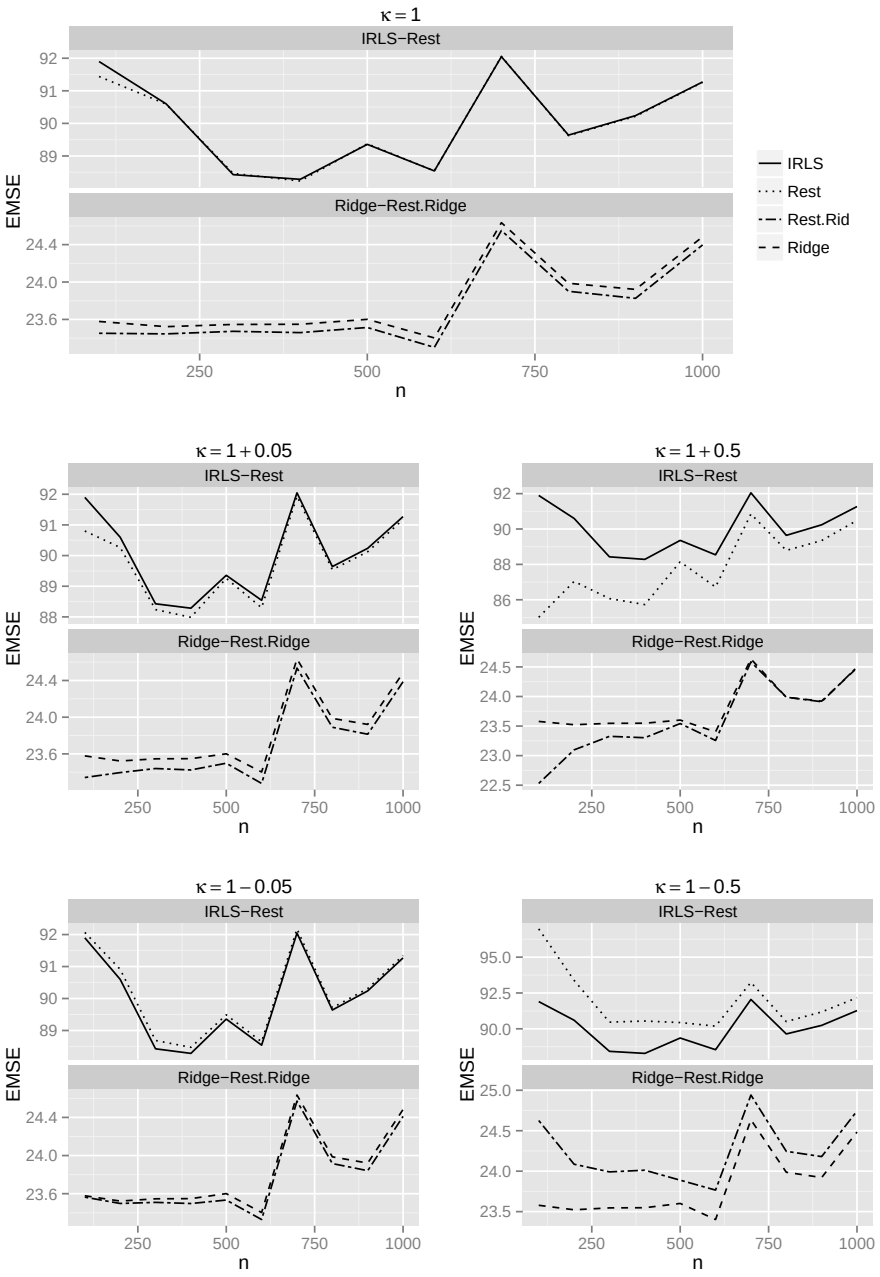

Figure 5.6. Plot of EMSE of the estimators over various n and under the different restrictions when $q = 10$, $\gamma^2 = 0.90$ and observations are from a Poisson distribution

5.1.2. Restricted Liu Estimation

In this section, we examine the performance of the first-order approximated restricted Liu estimator to the first-order approximated restricted ML, Liu and ML estimators in GLMs via simulation studies. Since the distribution of response, degrees of multicollinearity, sample size, restriction types, and Liu-biasing parameter impact the performance of the estimators, simulation studies are conducted under various choices of these parameters. The setting is as follows.

1. To see the effect of the number of observations, the sample sizes are chosen as $n = 100$, 200, 300, 400, 500, 600, 700, 800, 900, 1000 and the number of explanatory variables is chosen as $q = 4$.
2. The explanatory variables are then standardized by using unit length scaling so that $X^\top X$ is a matrix of correlations.
3. To see the correlation effect in the simulation, we consider $\gamma^2 = 0.90$, 0.95, 0.99.
4. For each set of explanatory variables, parameter vector β is chosen as $\beta = (1/2) \times j_{q\times 1}$ where $j_{q\times 1}$ is a $q \times 1$ vector of ones.
5. The values of β affect the value of $R_t^\top \beta - r_t$ which measures the relative performances of the estimators depend on the magnitude of $R_t^\top \beta - r_t$. Therefore, following Özkale (2014) four different scenarios are chosen:
   (i) $\beta_1 - 2\beta_2 = 0$, $\beta_3 - 2\beta_4 = 0$ where the effect of the first explanatory variable on the linear predictor is twice the second explanatory and the effect of the third explanatory variable on the linear predictor is the fourth explanatory variable
   $$\left(R = \begin{bmatrix} 1 & -2 & 0 & 0 \\ 0 & 0 & 1 & -2 \end{bmatrix}, r = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right),$$
   (ii) $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$ where the sum of the effects of the explanatory

variables on the linear predictor is 0 ($R = [1\ 1\ 1\ 1]$, $r = 0$),

(iii) $\beta_1 - \beta_2 = 0$, $\beta_3 - \beta_4 = 0$ where the first and second explanatory variables have same effect on the linear predictor, the third and fourth explanatory variables have the same effect on the linear predictor $\left(R = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, r = \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$,

(iv) $\beta_1 - \beta_2 = 0$, $\beta_3 - \beta_4 = 0$, $\beta_1 - \beta_3 = 0$ where each explanatory variable affect the linear predictor in the same way
$\left(R = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & -1 & 0 \end{bmatrix}, r = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right)$.

6. The observations y_i are generated from a Poisson distribution $y_i \sim \mathcal{P}(\mu_i)$, $i = 1, \ldots, n$ where μ_i is determined by

$$\mu_i = \exp(\beta_1 x_{i1} + \ldots + \beta_q x_{iq}).$$

We calculate the weight matrix $W^{(0)}$ with the diagonal elements $\mu_i^{(0)} = \exp(X\beta^{(0)})$. The working-response for Poisson distribution is defined as $z_i^{(0)} = x_i^\top \beta^{(0)} + (y_i - \mu_i^{(0)})/\mu_i^{(0)}$, $i = 1, \ldots, n$.

7. Model parameters are estimated using the first-order approximated ML, first-order approximated restricted ML and first-order approximated Liu methods.
8. The Liu-biasing parameters are chosen as $d = 0.1$, 0.3, 0.5, 0.7, and 0.9.
9. The number of replications in Monte Carlo simulation study is taken as 2000.

The performance of the estimators is calculated in terms of the EMSE:

$$\text{EMSE}(\tilde{\beta}) = \frac{1}{2000} \sum_{r=1}^{2000} (\tilde{\beta}_{(r)} - \beta)^\top (\tilde{\beta}_{(r)} - \beta)$$

where the subscript (r) refers to the rth replication and $\tilde{\beta}_{(r)}$ is the estimate of β in the rth replication of the experiment.

The simulation results are reported in Tables B21-B30 in Appendix-B. The main conclusions obtained from the simulation results are as follows:

(a) The EMSE values of the first-order approximated ML and restricted ML (except restriction iv) estimators increase as the degree of correlation increases. This case is valid for the first-order approximated Liu and restricted Liu (except restriction iv) estimators for fixed d value; however, the increasement is not as severe as compared to the first-order approximated Liu and restricted Liu estimators, respectively.

(b) Taking account of restrictions, the EMSE values of the first-order approximated restricted ML and restricted Liu estimators is least at the restriction (iv). The largest EMSE value of these estimators is at the restriction (ii).

(c) As d increases from 0.1 to 0.9 when correlation is fixed, the EMSE values of the first-order approximated Liu and restricted Liu estimators increase.

(d) The sample sizes do not have an obvious effect on the EMSE values of all the estimators considered when correlation is fixed.

5.2. Experiment 2: Simulation Study with Gamma Response

5.2.1. Liu Estimation

We consider a simulation example to compare the performance of the proposed first-order approximated Liu and ML estimators in GLMs under different degrees of multicollinearity, different sample sizes and different number of explanatory variables. The steps for the simulation setup are as follows:

1. To see the effect of the number of observations, the sample sizes are chosen as

$n = 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000$ and the number of explanatory variables is chosen as $q = 6$ and 10.

2. The explanatory variables are then standardized by using unit length scaling so that $X^\top X$ is a matrix of correlations and the explanatory matrix is taken as $[1, X]$.

3. Because we are mainly interested in the correlation effect in the simulation, we considered $\gamma^2 = 0.90, 0.95, 0.99$.

4. For each set of explanatory variables, parameter vector β is chosen as $\beta = j_{q\times 1}$ where $j_{q\times 1}$ is a $q \times 1$ vector of ones.

5. The response of the gamma model is generated using pseudo-random numbers from the $Gamma(v = \mu^2/var,\ u = var/\mu)$ distribution, where $\mu = \exp(X\beta)$, var denotes μ^2 and β is as explained in item 4 (Johnson, 2014).

6. We set the initial values $\mu_i^{(0)} = \exp(X\beta^{(0)})$ where $\beta^{(0)}$ is the OLS estimator and calculate the weight matrix W with the diagonal elements $(\mu_i^{(0)})^2$. We assign the working-response z with the ith observation $z_i^{(0)} = \log(\mu_i^{(0)}) + (y_i - \mu_i^{(0)})/\mu_i^{(0)}$ for the gamma distribution. Model parameters are estimated using the first-order approximated ML and first-order approximated Liu methods.

7. In the simulation procedure, the dispersion parameter is estimated to be the Pearson method $\hat{\varphi}_p^2 = (n-q)^{-1}\sum_{i=1}^n (y_i - \hat{\mu}_i^{(0)})^2/(\hat{\mu}_i^{(0)})^2$ for each sample size and explanatory variables.

8. $\hat{d}_h$ given by Equation (3.39.) is used throughout the simulation where h is determined as multiplying the upper bound defined in Theorem 3.13 by 0.99 if $\sum_{j=1}^q (\lambda_{\min}(\lambda_{\min}+1))/\lambda_j(\lambda_j+1) > 2$ and determined arbitrarily as 0.95 if $\sum_{j=1}^q (\lambda_{\min}(\lambda_{\min}+1))/\lambda_j(\lambda_j+1) < 2$ [3].

9. The number of replications in Monte Carlo simulation study is 5000. The perfor-

[3] The condition is obtained greater than 2 in all the simulation replications.

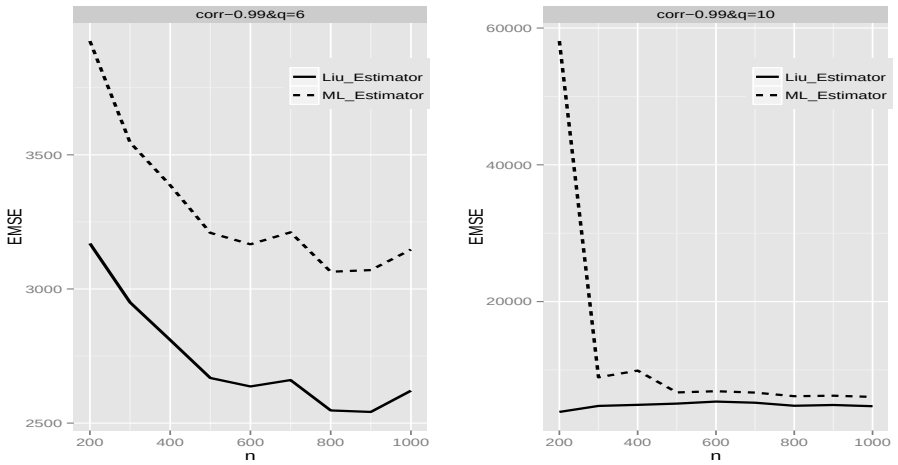

Figure 5.7. Plot of EMSE values of the first-order approximated ML and first-order approximated Liu estimators versus n for $\gamma^2 = 0.99$, $q = 6$ and $q = 10$

mance of the estimators is calculated in terms of the EMSE

$$\text{EMSE}(\tilde{\beta}) = \frac{1}{5000}\sum_{r=1}^{5000}(\tilde{\beta}_{(r)} - \beta)^{\top}(\tilde{\beta}_{(r)} - \beta)$$

where the subscript (r) refers to the rth replication and $\tilde{\beta}_{(r)}$ is the estimate of β in the rth replication of the experiment.

The simulation results are summarized by Figures 5.7-5.9. The main conclusions obtained from the simulation results are as follows:

(i) The EMSE values of both estimators increase with the increase of the number of the explanatory variables in all degree of correlation.

(ii) The first-order approximated Liu estimator is superior to the first-order approximated ML estimator for all cases.

(iii) When the degree of collinearity becomes larger, the EMSE values of the first-order approximated Liu and first-order approximated ML estimators increase for all cases.

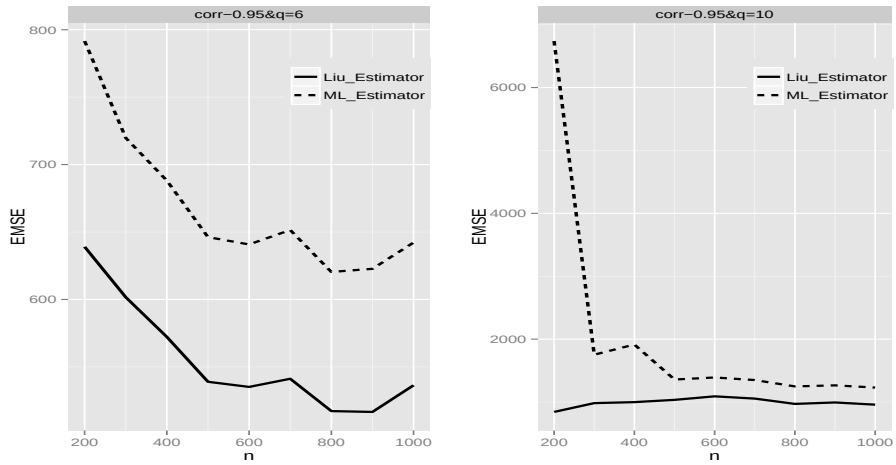

Figure 5.8. Plot of EMSE values of the first-order approximated ML and first-order approximated Liu estimators versus n for $\gamma^2 = 0.95$, $q = 6$ and $q = 10$

(iv) When $q = 6$ as the sample size becomes larger, the difference between the EMSE values of the first-order approximated Liu and ML estimators decreases. [4] This case is clear at high degree of correlation.

5.2.2. Restricted Liu Estimation

The explanatory variables are generated as in Section 5.1.2. and n values, degree of correlation, Liu-biasing parameter d, parameter vector β and restrictions are chosen as in Section 5.1.2. In this experiment, the response has gamma distribution with log link.

The response of the gamma model is generated using pseudo-random numbers from $Gamma(u = \mu_i^2/\text{var}_i, v = \text{var}_i/\mu_i)$ distribution with log link function where $\text{var}_i = \mu_i^2$ and

[4] The EMSE values of the first-order approximated Liu and the first-order approximated ML estimators versus n are plotted without $n = 100$, because of very large EMSE value compared to others.

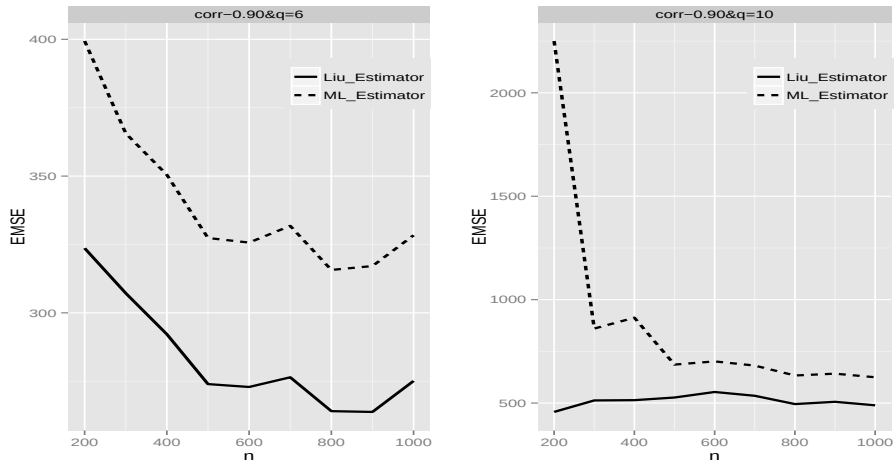

Figure 5.9. Plot of EMSE values of the first-order approximated ML and first-order approximated Liu estimators versus n for $\gamma^2 = 0.90$, $q = 6$ and $q = 10$

$$\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_q x_{iq}). \quad i = 1, \ldots, n,\ j = 1, \ldots, q$$

We assign the working-response z with the ith observation $z_i^{(0)} = \log(\mu_i^{(0)}) + (y_i - \mu_i^{(0)})/\mu_i^{(0)}$ for the gamma distribution. In the simulation procedure, the dispersion parameter is estimated to be the Pearson method $\hat{\varphi}_p^2 = (n-q)^{-1}\sum_{i=1}^{n}(y_i - \hat{\mu}_i^{(0)})^2/(\hat{\mu}_i^{(0)})^2$ for each sample size and explanatory variables.

The simulation results are reported in Tables B11-B20 in Appendix-B. The main conclusions obtained from the simulation results are as follows:

(a) While the degree of collinearity becomes larger at fixed Liu-biasing parameter except for restriction (iv), the EMSE values of the estimators increase.

(b) When the Liu-biasing parameter becomes larger at fixed degree of collinearity and sample size, the EMSE values of the first-order approximated Liu and first-order approximated restricted Liu estimators increase.

(c) The sample sizes do not have an obvious effect on the EMSE values of the estimators at fixed degree of collinearity and Liu-biasing parameter.
(d) Taking account of restrictions, the difference between the EMSE values of the first-order approximated restricted Liu and first-order approximated restricted ML estimators is least at the restrictions (iv). The largest difference of these estimators is at the restriction (ii).
(e) The Liu-biasing parameter, degree of collinearity and sample size do not have an obvious effect on these results.

5.3. Experiment 3: Simulation Study with Binomial Response

5.3.1. Restricted Ridge Estimation

In this experiment, the response has binomial distribution with logit link. Apart from Experiment 1, we manually select several values of the ridge parameter. The setting is as follows.

The explanatory variables are generated as in Experiment 1 and n values, degree of correlation and restrictions are chosen as in Experiment 1. The n observations for the Bernoulli response are generated from a Bernoulli distribution

$$Y_i \sim \mathcal{B}(\mu_i),\ i = 1,\ldots,n$$

where μ_i is determined by $\mu_i = \exp(x_i^\top \beta)/[1+\exp(x_i^\top \beta)]$.

We calculate the weight matrix W with the diagonal element $\mu^{(m)}(1-\mu^{(m)})$. The working-response for Bernoulli distribution is defined as $z^{(m)} = X\beta^{(m)} + (y_i - \mu^{(m)})/\{\mu^{(m)}(1-\mu^{(m)})\}$. The value 1×10^{-6} is used as a convergence criterion and iterations are stopped if $\left\|\beta^{(m+1)} - \beta^{(m)}\right\| > 1\times 10^{-6}$.

We arbitrarily consider three types of the ridge parameter as $k = 0.1,\ 0.3,$

0.9. The simulation results are reported in Tables B1-B10 in Appendix-B[5] . The conclusions from Tables B1-B10 are as follows:

(i) In all cases, the restricted ridge and ridge estimators are better than the IRLS and restricted ML estimators.

(ii) As collinearity increases, the superiority of the restricted ridge and ridge estimators over the IRLS and restricted ML estimators becomes more explicit.

(iii) While the degree of collinearity becomes larger, the EMSE values of the IRLS and restricted ML estimators increase and the EMSE values of the restricted ridge and ridge estimators decrease when the restriction type, the ridge parameter and the sample size are fixed.

(iv) As the ridge parameter increases from 0.1 to 0.9, the EMSE values of the restricted ridge and ridge estimators decrease at fixed correlation and fixed sample size.

(v) At large collinearity and sample size, the superiority of the restricted ridge estimator is more explicit.

[5] In tables (i), (ii), (iii), (v) and (iv) show respectively the restrictions corresponding to $\kappa = 1$, $\kappa = 1.05$, $\kappa = 1.5$, $\kappa = 0.95$ and $\kappa = 0.5$.

6. CONCLUSIONS

GLMs were first introduced by Nelder and Wedderburn (1972) as an extension of the class of linear models. Multicollinearity problem is very common in applied research and it may lead to some of regression coefficients being statistically insignificant and difficulties in interpreting the estimates of an individual coefficient. It is well known that the estimator performance can be adversely affected by multicollinearity. The presence of multicollinearity has serious effects on the ML estimates of the regression coefficients. As in the case in linear regression, collinearity causes difficulty in the interpretation of estimated regression coefficients in GLMs.

In this thesis, several biased estimators are presented as alternatives to the ML estimator when the data are ill conditioned in GLMs. The adverse effects of multicollinearity on parameter estimation in GLMs are also explored by various authors in the case of ML estimation. To overcome multicollinearity in GLMs, ridge and restricted ML estimators were proposed. To cope with the effects of multicollinearity in GLMs, three new estimators, called restricted ridge, Liu and restricted Liu estimators in GLMs are introduced. Using a penalty function approach Fisher scoring method for obtaining the estimators is proposed in GLMs.

Firstly, a restricted ridge estimator is introduced by unifying the ridge estimator and the restricted ML estimator in GLMs and its MSE properties are investigated and the MSE comparisons are done in the context of first-order approximated estimators. We also obtain necessary and sufficient conditions for the superiority of the first-order approximated restricted ridge estimator over the first-order approximated ML, restricted ML and ridge estimators by the approximated MSE criterion. Furthermore, the theoretical results are illustrated by two numerical examples and simulation studies are conducted with Poisson, gamma and binomial responses. Two numerical

examples and simulation studies are conducted in order to see the effect of collinearity, the number of sample size, type of restrictions on the superiority of the restricted ridge estimator to the IRLS, restricted ML, and ridge estimators in GLMs with respect to approximated MSE criterion. Based on the numerical example and the simulation studies, we conclude that the restricted ridge estimator generally outperforms other estimators under certain restrictions.

Secondly, we propose a first-order approximated Liu estimator to combat multicollinearity in GLMs which is an extension of the Liu estimator in the linear regression model. We also obtain necessary and sufficient condition for the superiority of the first-order approximated Liu estimator over the the first-order approximated ML estimator by the approximated MSE criterion. Liu-biasing parameter estimator of the first-order approximated Liu estimator is proposed by minimizing the approximated MSE. The results are illustrated by conducting a numerical example and simulation study with gamma response. A numerical example and a simulation study are conducted in order to see the effect of collinearity, the number of sample size and the number of explanatory variables on the superiority of the first-order approximated Liu estimator to the first-order approximated ML estimator in GLMs with respect to MSE criterion. Based on the numerical example and simulation study, we conclude that the first-order approximated Liu estimator outperforms the first-order approximated ML estimator for model on gamma distributed response. The simulation study indicates that there is a clear gain of using the first-order approximated Liu estimator, especially when the degree of multicollinearity is severe and the sample size is small for GLMs on gamma distributed response.

Thirdly, restricted Liu estimator is introduced by unifying Liu and restricted ML estimators in GLMs. We also obtain necessary and sufficient conditions for the superiority of the first-order approximated restricted Liu estimator over the first-order

approximated ML and Liu estimators by the approximated MSE criterion. The results are illustrated by conducting simulation studies and numerical examples. The superiority of the first-order approximated restricted Liu estimator over the first-order approximated Liu, restricted ML and ML estimators is examined with respect to the approximated MSE criterion. The numerical examples and simulation studies are conducted on the superiority of the first-order approximated restricted Liu estimator over the first-order approximated ML, restricted ML and Liu estimators. Based on the analyses, we conclude that the first-order approximated restricted Liu estimator outperforms the other estimators under certain restrictions. The Liu-biasing parameter affects the performance of the first-order approximated restricted Liu and Liu estimators in GLMs.

The numerical and simulation results verify that collinearity seriously affects the ML estimator in GLMs, producing large and unstable estimates. In this situation, our proposed estimators both performed well in reducing the MSE and producing less variable squared errors.

The advantages of these alternative estimators are that they can reduce the effect of the collinearity and improve the precision and accuracy of the estimated independent variable effects. Based on the analyses, we conclude that these estimators outperform other estimators under certain conditions. The studies reported in this thesis give evidence that the our proposed estimators in GLMs work well in certain cases. For further research, different methods to cope with the effects of multicollinearity in GLMs can be given or the performance of our proposed estimators over other estimators can be compared.

REFERENCES

Agresti, A., 2002. Categorical Data Analysis. Wiley, Hoboken, 734p.

Agresti, A., 2015. Foundations of Linear and Generalized Linear Models. Wiley, Hoboken, 472p.

Belsley, D. A., Kuh, E., and Welsch, R. E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley, New York, 292p.

Ben-Israel, A., and Greville, T. N. E., 1974. Generalized Inverses: Theory and Applications. Wiley, New York, 395p.

Chatterjee, S., and Hadi, A. S., 1988. Sensitivity Analysis in Linear Regression. Wiley, Toronto, 315p.

Cook, R. J., 2005. Generalized Linear Model. Encyclopedia of Biostatistics, Wiley, DOI: 10.1002/0470011815.b2a09019.

Dobson, A. J., and Barnett, A. G., 2008. An Introduction to Generalized Linear Models. CRC Press, Boca Raton, 320p.

Draper, N., and Smith, H., 1981. Applied Regression Analysis. Wiley, New York, 736p.

Fahrmeir, L., and Tutz, G., 1994. Multivariate Statistical Modelling Based on Generalized Linear Models. Springer, New York, 426p.

Farebrother, R. W., 1976. Further Results on the Mean Square Error of Ridge Regression. The Journal of the Royal Statistical Society Series B, 38(3):248-250.

Firth, D., 1991. Generalized Linear Models (D. V. Hinkley, N. Reid, and E. J. Snell editors). In Statistical Theory and Modelling: In Honour of Sir David Cox, FRS, CRC Press, London, p.55-82.

Gbur, E. E., Stroup, W. W., McCarter, K. S., Durham, S., Young, L. J., Christman, M., West, M., and Kramer, M., 2012. Analysis of Generalized Linear Mixed

Models in the Agricultural and Natural Resources Sciences. ASA, SSSA, and CSSA, Madison, 298p.

Gill, J., 2001. Generalized Linear Models: A Unified Approach. A Sage University Paper:Quantitative Applications in the Social Sciences, California, 112p.

Graybill, F. A., 1976. Theory and Application of the Linear Model. Duxbury Press, Belmont, 704p.

Groß, J., 2003. Restricted Ridge Estimation. Statistics and Probability Letters, 65(1):57-64.

Hardin, J. W., and Hilbe, J. M., 2012. Generalized Linear Models and Extensions. Stata Press, Texas, 479p.

Harville, D. A., 2008. Matrix Algebra from a Statistician's Perspective. Springer, New York, 634p.

Hoerl, A. E., and Kennard, R. W., 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12:55-67.

Hoerl, A. E., Kennard, R. W., and Baldwin, K. F., 1975. Ridge Regression: Some Simulations. Communications in Statistics-Theory and Methods, 4:105-123.

Huang, C. C. L., Jou, Y. J., and Cho, H. J., 2016. A New Multicolliearity Diagnostic for Generalized Linear Models. Journal of Applied Statistics, 43(11):2029-2043.

Johnson, P. E., 2014. GLM with a Gamma-Distributed Dependent Variable. http://pj.freefaculty.org/guides/stat/Regression-GLM/Gamma/GammaGLM-01.pdf. Accessed 5 June 2016.

Kaçıranlar, S., Sakallıoğlu, S., Özkale, M. R., and Güler, H., 2011. More on the Restricted Ridge Regression Estimation. Journal of Statistical Computation and Simulation, 81(11):1433-1448.

Khuri, A. I., 2009. Linear Model Methodology. CRC Press, Boca Raton, 562p.

Kibria, B. G., and Saleh, A. M. E., 2012. Improving the Estimators of the Parameters of a Probit Regression Model: A Ridge Regression Approach. Journal of Statistical Planning and Inference, 142(6):1421-1435.

Le Cessie, S., and Van Houwelingen, J. C., 1992. Ridge Estimators in Logistic Regression. Applied statistics, 41(1):191-201.

Lee, A. H., and Silvapulle, M. J., 1988. Ridge Estimation in Logistic Regression. Communications in Statistics-Simulation and Computation, 17(4):1231-1257.

Lesaffre, E., and Marx, B. D., 1993. Collinearity in Generalized Linear Regression. Communications in Statistics-Theory and Methods, 22(7):1933-1952.

Lindsey, J. K., 1997. Applying Generalized Linear Models. Springer, New York, 256p.

Liu, K., 1993. A New Class of Biased Estimate in Linear Regression. Communications in Statistics-Theory and Methods, 22(2):393-402.

Mackinnon, M. J., 1986. Collinearity in Generalized Linear Models. Dissertation, The University of British Columbia, 86p.

Mackinnon, M. J., and Puterman, M. L., 1989. Collinearity in Generalized Linear Models. Communications in Statistics-Theory and Methods, 18(9):3463-3472.

Madsen, H., Thyregod, P., 2010. Introduction to General and Generalized Linear Models. CRC Press, Boca Raton, 316p.

Månsson, K., and Shukur, G., 2011. On Ridge Parameters in Logistic Regression. Communications in Statistics-Theory and Methods, 40(18):3366-3381.

Marx, B. D., 1992. A Continuum of Principal Component Generalized Linear Regressions. Computational Statistics and Data Analysis 13(4):385-393.

Marx, B. D., and Smith, E. P., 1990a. Ill-Conditioned Information Matrices, Gen-

eralized Linear Models and Estimation of the Effects of Acid Rain. Environmetrics, 1(1):57-71.

_____, 1990b. Principal Component Estimation for Generalized Linear Regression. Biometrika, 77(1):23-31.

McCullagh, P., and Nelder, J. A., 1989. Generalized Linear Models. CRC Press, London, 532p.

McDonald, J. W., and Diamond, I. D., 1990. On the Fitting of Generalized Linear Models with Nonnegativity Parameter Constraints. Biometrics, 46(1):201-206.

McDonald, G. C., and Galarneau, D. I., 1975. A Monte Carlo Evaluation of Ridge-Type Estimators. Journal of the American Statistical Association, 70:407-416.

Millar, R. B., 2011. Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB. Wiley, Chichester, 357p.

Montgomery, D., Peck, E. A., and Vining, G. G., 2001. Introduction to Linear Regression Analysis. Wiley, New York, 672p.

Myers, R. H., 1990. Classical and Modern Regression with Applications. Duxbury Press, Boston, 488p.

Myers, R. H., Montgomery, D. C., Vining, G. G., and Robinsin, T. J., 2010. Generalized Linear Models with Applications in Engineering and the Sciences. Wiley, New York, 496p.

Nelder, J. A., and Wedderburn, R. W. M., 1972. Generalized Linear Models. Journal of the Royal Statistical Society Series A, 135(3):370-384.

Nyquist, H., 1991. Restricted Estimation of Generalized Linear Models. Journal of the Royal Statistical Society Series C, 40(1):133-141.

Paula, G. A., 1993. Assessing Local Influence in Restricted Regression Models.

Computational Statistics and Data Analysis, 16(1):63-79.

Olsson, U., 2002. Generalized Linear Models: An Applied Approach. Studentlitteratur, Lund, 244p.

Özkale, M. R., 2014. The Relative Efficiency of the Restricted Estimators in Linear Regression Models. Journal of Applied Statistics, 41(5):998-1027.

Sarkar, N., 1992. A New Estimator Combining the Ridge Regression and the Restricted Least Squares Methods of Estimation. Communications in Statistics-Theory and Methods, 21:1987-2000.

Segerstedt, B., 1992. On Ordinary Ridge Regression in Generalized Linear Models. Communications in Statistics-Theory and Methods, 21(8):2227-2246.

Schaefer, R. L., 1986. Alternative Estimators in Logistic Regression When the Data are Collinear. Journal of Statistical Computation and Simulation, 25(1-2):75-91.

Schaefer, R. L., Roi, L. D., and Wolfe, R. A., 1984. A Ridge Logistic Estimator. Communications in Statistics-Theory and Methods, 13(1):99-113.

Stein, C. M., 1981. Estimation of the Mean of a Multivariate Normal Distribution. The Annals of Statistics, 1135-1151.

Trenkler, G., and Toutenburg, H., 1990. Mean Squared Error Matrix Comparisons Between Biased Estimators? An Overview of Recent Results. Statistical Papers, 31(1):165-179.

Wedderburn, R. W. M., 1976. On the Existence and Uniqueness of the Maximum Likelihood Estimates for Certain Generalized Linear Models. Biometrika, 63(1):27-32.

Weissfeld, L. A., Sereika, S. M., 1991. A Multicollinearity Diagnostic for Generalized Linear Models. Communications in Statistics-Theory and Methods, 20(4):1183-1198.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer, New York, 213p.

CURRICULUM VITAE

Fikriye KURTOĞLU was born February 1985, in Adana, Turkey. She graduated from Çukurova University in 2008 with a Bachelor of Science (BSc) degree in Department of Statistics. After graduation, she continued her graduate studies in the same department from which she received her Master of Science degree (MSc) in Statistics in 2011. She has been working as a research assistant in Department of Statistics, Çukurova University, since 2011. She began a PhD degree in Department of Statistics Çukurova University in 2011.

APPENDICES

1. APPENDIX-A

1.1. Theorems

Theorem A1 (Harville 2008) Let R represent an $n \times q$ matrix, S an $n \times m$ matrix, T an $m \times p$ matrix, and U a $p \times q$ matrix. If $\Re(STU) \subset \Re(R)$ and $\mathbb{C}(STU) \subset \mathbb{C}(R)$, then the matrix $R^- - R^-STT^-(T^- + T^-TUR^-STT^-)^-T^-TUR^-$ is a g-inverse of the matrix $R+STU$.

Theorem A2 (Graybill 1976) Let A and B be $n \times n$ symmetric matrices

1. If A is pd, there exists a nonsingular matrix Q such that $Q^\top AQ = I$ and $Q^\top BQ = D$, where D is a diagonal matrix and the diagonal elements of D are roots λ of the polynomial equation $|B - \lambda A| = 0$.

2. If A and B are both nonnegative (neither has to be positive definite), there exists a nonsingular matrix Q such that $Q^\top AQ$ and $Q^\top BQ$ are each diagonal.

Theorem A3 (Ben-Israel and Greville 1974) Let $\varsigma_{n\times p}$ be the set of $n \times p$ complex matrices and let H_n be the subset of $\varsigma_{n\times n}$ consisting of Hermitian matrices. Further, given $L \in \varsigma_{n\times p}$, the symbols L^*, $R(L)$ and $\varsigma(L)$ stand for the conjugate transpose, the range and set of all g-inverses, respectively of L. Now, let $A \in H_n$, a_1 and $a_2 \in \varsigma_{n\times 1}$ be linearly independent, $f_{ij} = a_i^* A^- a_j$, $i, j = 1,2$ for $A^- \in \varsigma(A)$, and $s = [a_1^*(I_n - AA^-)^*(I_n - AA^-)a_2]/[a_1^*(I_n - AA^-)^*(I_n - AA^-)a_1]$, provided that $a_1 \notin R(A)$. Then $A + a_1a_1^* - a_2a_2^*$ is nnd if and only if any one of the following sets of conditions hold.

(a) A is nnd, $a_1 \in R(A)$, $a_2 \in R(A)$ and $(f_{11}+1)(f_{22}-1) \leq |f_{12}|^2$;

(b) A is nnd, $a_1 \notin R(A)$, $a_2 \in R(A : a_1)$ and $(a_2 - sa_1)^* A^- (a_2 - sa_1) \leq 1 - |s|^2$;

(c) $A = U\Delta U^* - \delta vv^*$, $a_1 \in R(A)$, $a_2 \in R(A)$, $v^*a_1 \neq 0$ and $(f_{11}+1) \leq 0$, $(f_{22}-1) \leq 0$, $(f_{11}+1)(f_{22}-1) \geq |f_{12}|^2$ where $(U : v)$ (with U possibly absent) is a subunitary matrix, Δ is a pd diagonal matrix (occuring when U is present) and δ is a positive scalar. Further, the conditions (a)-(c) are all independent of the choice of $A^- \in \varsigma(A)$.

2. APPENDIX-B

2.1. Tables of Simulation Studies for Restricted Ridge Estimator

Table B1. Simulation results with binomial response and n = 100

k	corr	EMSEs of Estimators											
		IRLS	Restricted					Ridge	Restricted ridge				
			Restrictions						Restrictions				
			(i)	(ii)	(iii)	(iv)	(v)		(i)	(ii)	(iii)	(iv)	(v)
	0.90	253.1106	246.6541	247.2221	253.8099	246.1191	242.7964	8.9491	8.2532	8.2593	8.5463	8.2524	8.4779
0.1	0.95	513.1905	500.4911	501.6962	515.3872	499.3496	491.9534	5.7583	5.0950	5.0988	5.3605	5.0963	5.3356
	0.99	2547.1080	2486.4370	2492.5710	2561.2860	2480.6050	2441.7700	1.8447	1.1999	1.2024	1.4506	1.2023	1.4494
	0.90	253.1106	246.6541	247.2221	253.8099	246.1191	242.7964	1.7793	1.2407	1.2435	1.4933	1.2431	1.4894
0.3	0.95	513.1905	500.4911	501.6962	515.3872	499.3496	491.9534	1.1908	0.6668	0.6694	0.9176	0.6693	0.9164
	0.99	2547.1080	2486.4370	2492.5710	2561.2860	2480.6050	2441.7700	0.6534	0.1380	0.1406	0.3881	0.1405	0.3880
	0.90	253.1106	246.6541	247.2221	253.8099	246.1191	242.7964	0.5789	0.1546	0.1571	0.4047	0.1571	0.4045
0.9	0.95	513.1905	500.4911	501.6962	515.3872	499.3496	491.9534	0.4928	0.0786	0.0811	0.3286	0.0810	0.3285
	0.99	2547.1080	2486.4370	2492.5710	2561.2860	2480.6050	2441.7700	0.4232	0.0155	0.0180	0.2655	0.0180	0.2655

Table B2. Simulation results with binomial response and n = 200

k	corr	EMSEs of Estimators											
		IRLS	Restricted					Ridge	Restricted ridge				
			Restrictions						Restrictions				
			(i)	(ii)	(iii)	(iv)	(v)		(i)	(ii)	(iii)	(iv)	(v)
	0.90	237.0984	233.9629	234.2776	238.0107	233.6683	231.9265	8.3233	7.6776	7.6822	7.9533	7.6780	7.9118
0.1	0.95	472.7312	467.2567	467.9081	475.3978	466.6415	462.7487	5.3419	4.7269	4.7302	4.9863	4.7287	4.9711
	0.99	2341.8600	2316.9210	2320.2330	2357.4080	2313.7730	2292.8890	1.7287	1.1303	1.1329	1.3808	1.1328	1.3800
	0.90	237.0984	233.9629	234.2776	238.0107	233.6683	231.9265	1.6487	1.1271	1.1297	1.3786	1.1294	1.3761
0.3	0.95	472.7312	467.2567	467.9081	475.3978	466.6415	462.7487	1.1155	0.6106	0.6131	0.8611	0.6131	0.8603
	0.99	2341.8600	2316.9210	2320.2330	2357.4080	2313.7730	2292.8890	0.6247	0.1297	0.1322	0.3797	0.1322	0.3797
	0.90	237.0984	233.9629	234.2776	238.0107	233.6683	231.9265	0.5675	0.1388	0.1413	0.3889	0.1413	0.3888
0.9	0.95	472.7312	467.2567	467.9081	475.3978	466.6415	462.7487	0.4894	0.0715	0.0740	0.3216	0.0740	0.3215
	0.99	2341.8600	2316.9210	2320.2330	2357.4080	2313.7730	2292.8890	0.4244	0.0146	0.0171	0.2646	0.0171	0.2646

Table B3. Simulation results with binomial response and n = 300

k	corr	EMSEs of Estimators											
		IRLS	Restricted					Ridge	Restricted ridge				
			Restrictions						Restrictions				
			(i)	(ii)	(iii)	(iv)	(v)		(i)	(ii)	(iii)	(iv)	(v)
	0.90	222.3729	219.9038	220.1237	222.7917	219.6992	218.5489	8.7120	8.0562	8.0604	8.3272	8.0570	8.2929
0.1	0.95	443.5255	439.3050	439.7563	445.0037	438.8803	436.2485	5.6020	4.9756	4.9788	5.2332	4.9775	5.2208
	0.99	2201.2040	2182.7130	2184.9960	2210.6820	2180.5450	2166.1980	1.8059	1.1985	1.2010	1.4490	1.2010	1.4482
	0.90	222.3729	219.9038	220.1237	222.7917	219.6992	218.5489	1.7195	1.1904	1.1930	1.4417	1.1927	1.4395
0.3	0.95	443.5255	439.3050	439.7563	445.0037	438.8803	436.2485	1.1563	0.6448	0.6474	0.8952	0.6473	0.8946
	0.99	2201.2040	2182.7130	2184.9960	2210.6820	2180.5450	2166.1980	0.6376	0.1376	0.1401	0.3877	0.1401	0.3876
	0.90	222.3729	219.9038	220.1237	222.7917	219.6992	218.5489	0.5750	0.1470	0.1495	0.3970	0.1494	0.3969
0.9	0.95	443.5255	439.3050	439.7560	445.0037	438.8803	436.2485	0.4909	0.0756	0.0781	0.3257	0.0781	0.3256
	0.99	2201.2040	2182.7130	2184.9960	2210.6820	2180.5450	2166.1980	0.4224	0.0155	0.0180	0.2655	0.0180	0.26545

Table B4. Simulation results with binomial response and n = 400

k	corr	EMSEs of Estimators											
		IRLS	Restricted					Ridge	Restricted ridge				
			Restrictions						Restrictions				
			(i)	(ii)	(iii)	(iv)	(v)		(i)	(ii)	(iii)	(iv)	(v)
	0.90	216.6480	214.7188	214.8748	216.8360	214.5752	213.8401	8.6157	7.9640	7.9677	8.2287	7.9653	8.2049
0.1	0.95	432.3340	429.2285	429.5509	433.3676	428.9264	427.1235	5.5416	4.9174	4.9204	5.1728	4.9195	5.1640
	0.99	2146.9680	2134.3880	2136.0270	2154.5420	2132.8320	2122.5980	1.7914	1.1822	1.1847	1.4325	1.1847	1.4320
	0.90	216.6480	214.7188	214.8748	216.8360	214.5752	213.8401	1.7007	1.1728	1.1754	1.4238	1.1753	1.4222
0.3	0.95	432.3340	429.2285	429.5509	433.3676	428.9264	427.1235	1.1481	0.6360	0.6385	0.8863	0.6385	0.8858
	0.99	2146.9680	2134.3880	2136.0270	2154.5420	2132.8320	2122.5980	0.6397	0.1357	0.1382	0.3857	0.1382	0.3857
	0.90	216.6480	214.7188	214.8748	216.8360	214.5752	213.8401	0.5719	0.1446	0.1471	0.3946	0.1471	0.3945
0.9	0.95	432.3340	429.2285	429.5509	433.3676	428.9264	427.1235	0.4913	0.0745	0.0770	0.3246	0.0770	0.3245
	0.99	2146.9680	2134.3880	2136.0270	2154.5420	2132.8320	2122.5980	0.4263	0.0152	0.0177	0.2652	0.0177	0.2652

Table B5. Simulation results with binomial response and n = 500

k	corr	EMSEs of Estimators											
		IRLS	Restricted					Ridge	Restricted ridge				
			Restrictions						Restrictions				
			(i)	(ii)	(iii)	(iv)	(v)		(i)	(ii)	(iii)	(iv)	(v)
	0.90	201.6655	200.0394	200.1540	201.6530	199.9352	199.4673	8.7557	8.1174	8.1208	8.3791	8.1189	8.3602
0.1	0.95	406.6036	404.1232	404.3616	407.2410	403.9012	402.6409	5.6410	5.0323	5.0352	5.2866	5.0345	5.2797
	0.99	2033.7440	2024.3580	2025.5770	2039.4020	2023.2020	2015.6760	1.7842	1.1982	1.2007	1.4484	1.2007	1.4481
	0.90	201.6655	200.0394	200.1540	201.6530	199.9352	199.4673	1.7303	1.2107	1.2132	1.4614	1.2131	1.4602
0.3	0.95	406.6036	404.1232	404.3616	407.2410	403.9012	402.6409	1.1568	0.6548	0.6573	0.9050	0.6573	0.9046
	0.99	2033.7440	2024.3580	2025.5770	2039.4020	2023.2020	2015.6760	0.6247	0.1377	0.1402	0.3877	0.1402	0.3876
	0.90	201.6655	200.0394	200.1540	201.6530	199.9352	199.4673	0.5762	0.1500	0.1525	0.4000	0.1525	0.4000
0.9	0.95	406.6036	404.1232	404.3616	407.2410	403.9012	402.6409	0.4912	0.0769	0.0794	0.3269	0.0794	0.3269
	0.99	2033.7440	2024.3580	2025.5770	2039.4020	2023.2020	2015.6760	0.4190	0.0155	0.0180	0.2655	0.0180	0.2655

Table B6. Simulation results with binomial response and n = 600

k	corr	EMSEs of Estimators											
		IRLS	Restricted					Ridge	Restricted ridge				
			Restrictions						Restrictions				
			(i)	(ii)	(iii)	(iv)	(v)		(i)	(ii)	(iii)	(iv)	(v)
	0.90	205.6111	204.1114	204.2068	205.4960	204.0255	203.6837	8.6671	8.0079	8.0111	8.2668	8.0097	8.2526
0.1	0.95	409.9816	407.8258	408.0222	410.4428	407.6439	406.6612	5.5538	4.9270	4.9298	5.1801	4.9293	5.1752
	0.99	2053.3050	2045.5130	2046.5290	2058.1090	2044.5530	2038.3570	1.7942	1.1852	1.1877	1.4353	1.18775	1.4351
	0.90	205.6111	204.1114	204.2068	205.4960	204.0255	203.6837	1.7205	1.1879	1.1904	1.4384	1.1904	1.4376
0.3	0.95	409.9816	407.8258	408.0222	410.4428	407.6439	406.6612	1.1524	0.6395	0.6420	0.8897	0.6420	0.8895
	0.99	2053.3050	2045.5130	2046.5290	2058.1090	2044.5530	2038.3570	0.6385	0.1361	0.1386	0.3861	0.1386	0.3861
	0.90	205.6111	204.1114	204.2068	205.4960	204.0255	203.6837	0.5722	0.1468	0.1493	0.3969	0.1493	0.3968
0.9	0.95	409.9816	407.8258	408.0222	410.4428	407.6439	406.6612	0.4884	0.0751	0.0776	0.3251	0.0776	0.3251
	0.99	2053.3050	2045.5130	2046.5290	2058.1090	2044.5530	2038.3570	0.4218	0.0153	0.0178	0.2653	0.0178	0.2653

Table B7. Simulation results with binomial response and n = 700

k	corr	EMSEs of Estimators											
		IRLS	Restricted					Ridge	Restricted ridge				
			Restrictions						Restrictions				
			(i)	(ii)	(iii)	(iv)	(v)		(i)	(ii)	(iii)	(iv)	(v)
	0.90	212.3812	211.0249	211.1095	212.2730	210.9493	210.6718	8.7057	8.0710	8.0742	8.3290	8.0729	8.3162
0.1	0.95	424.9977	423.0482	423.2215	425.3745	422.8880	422.0421	5.5881	4.9786	4.9814	5.2315	4.9809	5.2268
	0.99	2128.1680	2121.2020	2122.0830	2132.1290	2120.3690	2115.0040	1.7904	1.1997	1.2022	1.4499	1.2022	1.4496
	0.90	212.3812	211.0249	211.1095	212.2730	210.9493	210.6718	1.7048	1.1913	1.1939	1.4418	1.1938	1.4410
0.3	0.95	424.9977	423.0482	423.2215	425.3745	422.8880	422.0421	1.1425	0.6444	0.6469	0.8946	0.6469	0.8943
	0.99	2128.1680	2121.2020	2122.0830	2132.1290	2120.3690	2115.0040	0.6236	0.1377	0.1402	0.3877	0.1402	0.3877
	0.90	212.3812	211.0249	211.1095	212.2730	210.9493	210.6718	0.5598	0.1470	0.1495	0.3970	0.1495	0.3970
0.9	0.95	424.9977	423.0482	423.2215	425.3745	422.8880	422.0421	0.4778	0.0755	0.0780	0.3256	0.0780	0.3250
	0.99	2128.1680	2121.2020	2122.0830	2132.1290	2120.3690	2115.0040	0.4093	0.0155	0.0180	0.2655	0.0180	0.2655

Table B8. Simulation results with binomial response and n = 800

k	corr	EMSEs of Estimators											
		IRLS	Restricted					Ridge	Restricted ridge				
			Restrictions						Restrictions				
			(i)	(ii)	(iii)	(iv)	(v)		(i)	(ii)	(iii)	(iv)	(v)
	0.90	204.9565	203.6861	203.7584	204.7908	203.6222	203.4287	8.7550	8.1188	8.1219	8.3756	8.1208	8.3649
0.1	0.95	411.4711	409.6955	409.8453	411.7453	409.5581	408.8738	5.6329	5.0252	5.0279	5.2776	5.0275	5.2738
	0.99	2057.9500	2051.8090	2052.5800	2061.4400	2051.0800	2046.4410	1.7910	1.2034	1.2059	1.4535	1.2059	1.4533
	0.90	204.9565	203.6861	203.7584	204.7908	203.6222	203.4287	1.7230	1.2058	1.2084	1.4562	1.2083	1.4556
0.3	0.95	411.4711	409.6955	409.8453	411.7453	409.5581	408.8738	1.1517	0.6526	0.6551	0.9027	0.6551	0.9025
	0.99	2057.9500	2051.8090	2052.5800	2061.4400	2051.0800	2046.4410	0.6241	0.1382	0.1407	0.3882	0.1407	0.3882
	0.90	204.9565	203.6861	203.7584	204.7908	203.6222	203.4287	0.5670	0.1491	0.1516	0.3991	0.1516	0.3991
0.9	0.95	411.4711	409.6955	409.8453	411.7453	409.5581	408.8738	0.4823	0.0766	0.0791	0.3266	0.0791	0.3266
	0.99	2057.9500	2051.8090	2052.5800	2061.4400	2051.0800	2046.4410	0.4123	0.0155	0.0180	0.2655	0.0180	0.2655

Table B9. Simulation results with binomial response and n = 900

k	corr	EMSEs of Estimators											
		IRLS	Restricted					Ridge	Restricted ridge				
			Restrictions						Restrictions				
			(i)	(ii)	(iii)	(iv)	(v)		(i)	(ii)	(iii)	(iv)	(v)
	0.90	204.7928	203.5756	203.6402	204.5855	203.5191	203.3748	8.7466	8.1085	8.1114	8.3643	8.1105	8.3552
0.1	0.95	409.5172	407.8425	407.9761	409.6933	407.7203	407.1366	5.6267	5.0147	5.0174	5.2668	5.0171	5.2635
	0.99	2041.2280	2035.7180	2036.4010	2044.2600	2035.0750	2031.0050	1.7969	1.2044	1.2069	1.4545	1.2069	1.4544
	0.90	204.7928	203.5756	203.6402	204.5855	203.5191	203.3748	1.7222	1.2032	1.2058	1.4535	1.2057	1.4531
0.3	0.95	409.5172	407.8425	407.9761	409.6933	407.7203	407.1366	1.1535	0.6511	0.6536	0.9012	0.6536	0.9010
	0.99	2041.2280	2035.7180	2036.4010	2044.2600	2035.0750	2031.0050	0.6283	0.1383	0.1408	0.3883	0.1408	0.3883
	0.90	204.7928	203.5756	203.6402	204.5855	203.5191	203.3748	0.5681	0.1488	0.1513	0.3988	0.1513	0.3987
0.9	0.95	409.5172	407.8425	407.9761	409.6933	407.7203	407.1366	0.4840	0.0764	0.0789	0.3264	0.0789	0.3264
	0.99	2041.2280	2035.7180	2036.4010	2044.2600	2035.0750	2031.0050	0.4150	0.0155	0.0180	0.2655	0.0180	0.2655

Table B10. Simulation results with binomial response and n = 1000

k	corr	EMSEs of Estimators											
		IRLS	Restricted					Ridge	Restricted ridge				
			Restrictions						Restrictions				
			(i)	(ii)	(iii)	(iv)	(v)		(i)	(ii)	(iii)	(iv)	(v)
	0.90	209.6366	208.4380	208.4947	209.3499	208.3890	208.2939	8.6210	7.9643	7.9672	8.2192	7.9664	8.2115
0.1	0.95	415.5924	413.9930	414.1082	415.6183	413.8883	413.4201	5.5270	4.8964	4.8991	5.1482	4.899	5.1454
	0.99	2051.9860	2047.1850	2047.7670	2054.4940	2046.6360	2043.1930	1.7869	1.1768	1.1793	1.4269	1.1793	1.4267
	0.90	209.6366	208.4380	208.4947	209.3499	208.3890	208.2939	1.7059	1.1739	1.1765	1.4242	1.1764	1.4238
0.3	0.95	415.5924	413.9930	414.1082	415.6183	413.8883	413.4201	1.1493	0.6336	0.6361	0.8837	0.6361	0.8836
	0.99	2051.9860	2047.1850	2047.7670	2054.4940	2046.6360	2043.1930	0.6381	0.1351	0.1376	0.3851	0.1376	0.3851
	0.90	209.6366	208.4380	208.4947	209.3499	208.3890	208.2939	0.5682	0.1447	0.1472	0.3947	0.1472	0.3947
0.9	0.95	415.5924	413.9930	414.1082	415.6183	413.8883	413.4201	0.4870	0.07427	0.0768	0.3243	0.0768	0.3243
	0.99	2051.9860	2047.1850	2047.7670	2054.4940	2046.6360	2043.1930	0.4201	0.0152	0.0177	0.2652	0.0177	0.2652

2.2. Tables of Simulation Studies for Restricted Liu Estimator

Table B11. Simulation results with gamma response and n = 100

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	29.5422	11.0529	30.3281	11.5347	0.2743	1.1915	0.6463	1.9655	0.5626	0.2267
0.1	0.95	59.4653	21.6902	60.2694	22.9911	0.2647	1.4184	0.7308	2.1978	0.6528	0.2208
	0.99	301.0209	106.8740	301.8662	115.3358	0.2582	3.7647	1.5997	4.5479	1.5706	0.2169
	0.90	29.5422	11.0529	30.3281	11.5347	0.2743	4.0886	1.7324	4.8596	1.6718	0.2297
0.3	0.95	59.4653	21.6902	60.2694	22.9911	0.2647	6.7766	2.7276	7.5534	2.7145	0.2235
	0.99	301.0209	106.8740	301.8662	115.3358	0.2582	28.5083	10.7601	29.2891	11.1787	0.2194
	0.90	29.5422	11.0529	30.3281	11.5347	0.2743	8.9303	3.5675	9.6937	3.5591	0.2370
0.5	0.95	59.4653	21.6902	60.2694	22.9911	0.2647	16.4440	6.3548	17.2141	6.4784	0.2302
	0.99	301.0209	106.8740	301.8662	115.3358	0.2582	76.8595	28.6882	77.6340	30.0105	0.2257
	0.90	29.5422	11.0529	30.3281	11.5347	0.2743	15.7166	6.1515	16.4678	6.2244	0.2487
0.7	0.95	59.4653	21.6902	60.2694	22.9911	0.2647	30.4206	11.6123	31.1798	11.9444	0.2410
	0.99	301.0209	106.8740	301.8662	115.3358	0.2582	148.8183	55.3839	149.5828	58.0658	0.2358
	0.90	29.5422	11.0529	30.3281	11.5347	0.2743	24.4475	9.4846	25.1821	9.6677	0.2648
0.9	0.95	59.4653	21.6902	60.2694	22.9911	0.2647	48.7065	18.5001	49.4506	19.1125	0.2558
	0.99	301.0209	106.8740	301.8662	115.3358	0.2582	244,3848	90,8474	245.1353	95.3447	0.2498

Table B12. Simulation results with gamma response and n = 200

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	30.1531	11.6998	30.8828	11.4509	0.2650	1.1771	0.6343	1.9630	0.5419	0.2140
0.1	0.95	59.7589	22.9170	60.4954	22.5290	0.2552	1.4040	0.7221	2.1958	0.6283	0.2082
	0.99	295.7997	112.4884	296.5275	110.9072	0.2479	3.6976	1.5997	4.4937	1.4941	0.2039
	0.90	30.1531	11.6998	30.8828	11.4509	0.2650	4.1285	1.7485	4.9106	1.6345	0.2175
0.3	0.95	59.7589	22.9170	60.4954	22.5290	0.2552	6.7874	2.7575	7.5758	2.6311	0.2113
	0.99	295.7997	112.4884	296.5275	110.9072	0.2479	28.0228	10.8483	28.8158	10.6216	0.2068
	0.90	30.1531	11.6998	30.8828	11.4509	0.2650	9.0729	3.6384	9.8467	3.4950	0.2255
0.5	0.95	59.7589	22.9170	60.4954	22.5290	0.2552	16.5047	6.4616	17.2856	6.2850	0.2187
	0.99	295.7997	112.4884	296.5275	110.9072	0.2479	75.5401	28.9512	76.3263	28.5005	0.2136
	0.90	30.1531	11.6998	30.8828	11.4509	0.2650	16.0103	6.3038	16.7712	6.1235	0.2380
0.7	0.95	59.7589	22.9170	60.4954	22.5290	0.2552	30.5560	11.8342	31.3251	11.5899	0.2302
	0.99	295.7997	112.4884	296.5275	110.9072	0.2479	146.2498	55.9084	147.0251	55.1307	0.2244
	0.90	30.1531	11.6998	30.8828	11.4509	0.2650	24.9406	9.7449	25.6842	9.5199	0.2549
0.9	0.95	59.7589	22.9170	60.4954	22.5290	0.2552	48.9411	18.8754	49.6943	18.5458	0.2458
	0.99	295.7997	112.4884	296.5275	110.9072	0.2479	240.1517	91.7198	240.9122	90.5122	0.2391

Table B13. Simulation results with gamma response and n = 300

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	29.2874	9.1568	30.0099	9.2613	0.2631	1.1714	0.6014	1.9570	0.5148	0.2142
0.1	0.95	58.7009	18.0131	59.4287	18.3725	0.2517	1.3928	0.6622	2.1859	0.5788	0.2067
	0.99	295.3786	89.3077	296.0891	91.7421	0.2430	3.6899	1.3523	4.4890	1.2934	0.2009
	0.90	29.2874	9.1568	30.0099	9.2613	0.2631	4.0491	1.4915	4.8314	1.4170	0.2173
0.3	0.95	58.7009	18.0131	59.4287	18.3725	0.2517	6.6896	2.2856	7.4798	2.2377	0.2095
	0.99	295.3786	89.3077	296.0891	91.7421	0.2430	27.9808	8.7243	28.7773	8.8954	0.2034
	0.90	29.2874	9.1568	30.0099	9.2613	0.2631	8.8528	2.9642	9.6271	2.9228	0.2248
0.5	0.95	58.7009	18.0131	59.4287	18.3725	0.2517	16.2369	5.1944	17.0199	5.2271	0.2164
	0.99	295.3786	89.3077	296.0891	91.7421	0.2430	75.4310	23.1019	76.2210	23.7442	0.2098
	0.90	29.2874	9.1568	30.0099	9.2613	0.2631	15.5822	5.0195	16.3441	5.0321	0.2368
0.7	0.95	58.7009	18.0131	59.4287	18.3725	0.2517	30.0347	9.3885	30.8063	9.5470	0.2274
	0.99	295.3786	89.3077	296.0891	91.7421	0.2430	146.0406	44.4849	146.8201	45.8398	0.2201
	0.90	29.2874	9.1568	30.0099	9.2613	0.2631	24.2375	7.6575	24.9824	7.7450	0.2533
0.9	0.95	58.7009	18.0131	59.4287	18.3725	0.2517	48.0829	14.8680	48.8390	15.1974	0.2426
	0.99	295.3786	89.3077	296.0891	91.7421	0.2430	239.8095	72.8734	240.5746	75.1823	0.2344

Table B14. Simulation results with gamma response and n = 400

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	30.4989	10.1679	31.2269	10.1423	0.2692	1.1982	0.6309	1.9880	0.5447	0.2102
0.1	0.95	60.7168	19.9452	61.4530	20.0150	0.2591	1.4280	0.7012	2.2239	0.6146	0.2040
	0.99	302.5332	98.3480	303.2697	99.1015	0.2516	3.7765	1.4669	4.5772	1.3820	0.1993
	0.90	30.4989	10.1679	31.2269	10.1423	0.2692	4.1940	1.6256	4.9787	1.5367	0.2152
0.3	0.95	60.7168	19.9452	61.4530	20.0150	0.2591	6.9077	2.5100	7.6989	2.4243	0.2087
	0.99	302.5332	98.3480	303.2697	99.1015	0.2516	28.6622	9.6375	29.4584	9.5716	0.2038
	0.90	30.4989	10.1679	31.2269	10.1423	0.2692	9.1987	3.2757	9.9735	3.1866	0.2248
0.5	0.95	60.7168	19.9452	61.4530	20.0150	0.2591	16.7849	5.7552	17.5669	5.6759	0.2177
	0.99	302.5332	98.3480	303.2697	99.1015	0.2516	77.2648	25.5768	78.0524	25.5526	0.2123
	0.90	30.4989	10.1679	31.2269	10.1423	0.2692	16.2122	5.5810	16.9725	5.4944	0.2391
0.7	0.95	60.7168	19.9452	61.4530	20.0150	0.2591	31.0595	10.4370	31.8281	10.3695	0.2310
	0.99	302.5332	98.3480	303.2697	99.1015	0.2516	149.5844	49.2849	150.3594	49.3249	0.2249
	0.90	30.4989	10.1679	31.2269	10.1423	0.2692	25.2345	8.5417	25.9756	8.4600	0.2580
0.9	0.95	60.7168	19.9452	61.4530	20.0150	0.2591	49.7317	16.5552	50.4824	16.5050	0.2487
	0.99	302.5332	98.3480	303.2697	99.1015	0.2516	245.6210	80.7618	246.3792	80.8885	0.2417

Table B15. Simulation results with gamma response and n = 500

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	29.6030	10.4164	30.3499	10.0334	0.2697	1.1968	0.6330	1.9752	0.5455	0.2218
0.1	0.95	59.1960	20.5059	59.9570	19.8935	0.2588	1.4185	0.7065	2.2041	0.6151	0.2145
	0.99	296.8833	101.5791	297.6659	99.1856	0.2506	3.7243	1.4973	4.5155	1.3830	0.2088
	0.90	29.6030	10.4164	30.3499	10.0334	0.2697	4.1111	1.6422	4.8867	1.5207	0.2248
0.3	0.95	59.1960	20.5059	59.9570	19.8935	0.2588	6.7674	2.5513	7.5505	2.4043	0.2172
	0.99	296.8833	101.5791	297.6659	99.1856	0.2506	28.1492	9.8955	28.9380	9.5523	0.2113
	0.90	29.6030	10.4164	30.3499	10.0334	0.2697	8.9672	3.3284	9.7358	3.1490	0.2321
0.5	0.95	59.1960	20.5059	59.9570	19.8935	0.2588	16.3966	5.8776	17.1731	5.6288	0.2240
	0.99	296.8833	101.5791	297.6659	99.1856	0.2506	75.8436	26.2998	76.6260	25.5085	0.2177
	0.90	29.6030	10.4164	30.3499	10.0334	0.2697	15.7651	5.6915	16.5222	5.4306	0.2438
0.7	0.95	59.1960	20.5059	59.9570	19.8935	0.2588	30.3062	10.6853	31.0719	10.2887	0.2348
	0.99	296.8833	101.5791	297.6659	99.1856	0.2506	146.8074	50.7102	147.5796	49.2516	0.2280
	0.90	29.6030	10.4164	30.3499	10.0334	0.2697	24.5049	8.7316	25.2461	8.3652	0.2600
0.9	0.95	59.1960	20.5059	59.9570	19.8935	0.2588	48.4960	16.9745	49.2470	16.3840	0.2498
	0.99	296.8833	101.5791	297.6659	99.1856	0.2506	241.0406	83.1267	241.7988	80.7817	0.2421

Table B16. Simulation results with gamma response and n = 600

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	31.0492	9.4549	31.7829	9.7342	0.2644	1.2202	0.6271	2.0056	0.5519	0.2145
0.1	0.95	61.6927	18.4864	62.4344	19.1613	0.2527	1.4518	0.6873	2.2449	0.6155	0.2068
	0.99	306.5041	90.7966	307.2401	94.5599	0.2433	3.8274	1.3867	4.6269	1.3474	0.2005
	0.90	31.0492	9.4549	31.7829	9.7342	0.2644	4.2684	1.5579	5.0504	1.5158	0.2178
0.3	0.95	61.6927	18.4864	62.4344	19.1613	0.2527	7.0191	2.3710	7.8092	2.3675	0.2097
	0.99	306.5041	90.7966	307.2401	94.5599	0.2433	29.0418	8.9219	29.8386	9.2231	0.2031
	0.90	31.0492	9.4549	31.7829	9.7342	0.2644	9.3625	3.0874	10.1367	3.1082	0.2255
0.5	0.95	61.6927	18.4864	62.4344	19.1613	0.2527	17.0548	5.3713	17.8377	5.5007	0.2168
	0.99	306.5041	90.7966	307.2401	94.5599	0.2433	78.2831	23.5922	79.0733	24.5711	0.2097
	0.90	31.0492	9.4549	31.7829	9.7342	0.2644	16.5027	5.2156	17.2645	5.3291	0.2377
0.7	0.95	61.6927	18.4864	62.4344	19.1613	0.2527	31.5587	9.6885	32.3303	10.0153	0.2281
	0.99	306.5041	90.7966	307.2401	94.5599	0.2433	151.5513	45.3974	152.3309	47.3913	0.2202
	0.90	31.0492	9.4549	31.7829	9.7342	0.2644	25.6889	7.9426	26.4337	8.1786	0.2544
0.9	0.95	61.6927	18.4864	62.4344	19.1613	0.2527	50.5310	15.3225	51.2870	15.9110	0.2435
	0.99	306.5041	90.7966	307.2401	94.5599	0.2433	248.8465	74.3378	249.6116	77.6837	0.2346

Table B17. Simulation results with gamma response and n = 700

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	30.6508	10.6061	31.4007	10.6001	0.2682	1.2114	0.6376	1.9914	0.5519	0.2202
0.1	0.95	61.1418	20.8514	61.9077	20.9660	0.2581	1.4429	0.7122	2.2298	0.6270	0.2132
	0.99	305.5251	103.0444	306.3232	104.1052	0.2507	3.8167	1.5126	4.6087	1.4331	0.2080
	0.90	30.6508	10.6061	31.4007	10.6001	0.2682	4.2199	1.6633	4.9972	1.5751	0.2232
0.3	0.95	61.1418	20.8514	61.9077	20.9660	0.2581	6.9573	2.5850	7.7416	2.5029	0.2160
	0.99	305.5251	103.0444	306.3232	104.1052	0.2507	28.9423	10.0209	29.7317	9.9892	0.2107
	0.90	30.6508	10.6061	31.4007	10.6001	0.2682	9.2476	3.3764	10.0177	3.2915	0.2305
0.5	0.95	61.1418	20.8514	61.9077	20.9660	0.2581	16.9015	5.9606	17.6790	5.8935	0.2229
	0.99	305.5251	103.0444	306.3232	104.1052	0.2507	78.0226	26.6400	78.8055	26.7119	0.2173
	0.90	30.6508	10.6061	31.4007	10.6001	0.2682	16.2945	5.7769	17.0531	5.7012	0.2423
0.7	0.95	61.1418	20.8514	61.9077	20.9660	0.2581	31.2753	10.8391	32.0420	10.7987	0.2339
	0.99	305.5251	103.0444	306.3232	104.1052	0.2507	151.0576	51.3701	151.8301	51.6011	0.2277
	0.90	30.6508	10.6061	31.4007	10.6001	0.2682	25.3605	8.8647	26.1033	8.8041	0.2585
0.9	0.95	61.1418	20.8514	61.9077	20.9660	0.2581	50.0788	17.2204	50.8307	17.2186	0.2490
	0.99	305.5251	103.0444	306.3232	104.1052	0.2507	248.0473	84.2112	248.8056	84.6569	0.2421

Table B18. Simulation results with gamma response and n = 800

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	28.7676	9.2882	29.4890	8.9809	0.2787	1.1544	0.6029	1.9432	0.5108	0.2112
0.1	0.95	57.2782	18.1590	58.0090	17.6806	0.2683	1.3691	0.6659	2.1643	0.5704	0.2048
	0.99	285.5347	89.2477	286.2683	87.3721	0.2607	3.5840	1.3566	4.3840	1.2416	0.2000
	0.90	28.7676	9.2882	29.4890	8.9809	0.2787	3.9868	1.5065	4.7690	1.3868	0.2177
0.3	0.95	57.2782	18.1590	58.0090	17.6806	0.2683	6.5468	2.3050	7.3357	2.1641	0.2111
	0.99	285.5347	89.2477	286.2683	87.3721	0.2607	27.0810	8.7428	27.8750	8.4313	0.2060
	0.90	28.7676	9.2882	29.4890	8.9809	0.2787	8.7072	3.0032	9.4780	2.8377	0.2292
0.5	0.95	57.2782	18.1590	58.0090	17.6806	0.2683	15.8654	5.2434	16.6435	5.0196	0.2218
	0.99	285.5347	89.2477	286.2683	87.3721	0.2607	72.9545	23.1494	73.7381	22.4508	0.2163
	0.90	28.7676	9.2882	29.4890	8.9809	0.2787	15.3154	5.0931	16.0700	4.8634	0.2454
0.7	0.95	57.2782	18.1590	58.0090	17.6806	0.2683	29.3249	9.4810	30.0877	9.1370	0.2370
	0.99	285.5347	89.2477	286.2683	87.3721	0.2607	141.2043	44.5766	141.9734	43.3001	0.2309
	0.90	28.7676	9.2882	29.4890	8.9809	0.2787	23.8116	7.7763	24.5451	7.4638	0.2664
0.9	0.95	57.2782	18.1590	58.0090	17.6806	0.2683	46.9252	15.0178	47.6683	14.5163	0.2568
	0.99	285.5347	89.2477	286.2683	87.3721	0.2607	231.8305	73.0242	232.5807	70.9791	0.2497

Table B19. Simulation results with gamma response and n = 900

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	30.8776	10.1850	31.6047	10.4930	0.2745	1.2322	0.6455	2.0123	0.5616	0.2199
0.1	0.95	61.5356	19.9866	62.2734	20.7300	0.2637	1.4632	0.7147	2.2502	0.6343	0.2130
	0.99	307.0856	98.6128	307.8316	102.7954	0.2557	3.8459	1.4826	4.6382	1.4311	0.2078
	0.90	30.8776	10.1850	31.6047	10.4930	0.2745	4.2710	1.6455	5.0469	1.5851	0.2241
0.3	0.95	61.5356	19.9866	62.2734	20.7300	0.2637	7.0238	2.5329	7.8069	2.5050	0.2169
	0.99	307.0856	98.6128	307.8316	102.7954	0.2557	29.1138	9.6898	29.9023	9.9218	0.2114
	0.90	30.8776	10.1850	31.6047	10.4930	0.2745	9.3379	3.2999	10.1050	3.2911	0.2328
0.5	0.95	61.5356	19.9866	62.2734	20.7300	0.2637	17.0352	5.7889	17.8102	5.8715	0.2250
	0.99	307.0856	98.6128	307.8316	102.7954	0.2557	78.4495	25.6918	79.2304	26.4979	0.2191
	0.90	30.8776	10.1850	31.6047	10.4930	0.2745	16.4327	5.6087	17.1865	5.6796	0.2461
0.7	0.95	61.5356	19.9866	62.2734	20.7300	0.2637	31.4973	10.4830	32.2599	10.7339	0.2373
	0.99	307.0856	98.6128	307.8316	102.7954	0.2557	151.8530	49.4887	152.6223	51.1593	0.2308
	0.90	30.8776	10.1850	31.6047	10.4930	0.2745	25.5556	8.5719	26.2915	8.7505	0.2639
0.9	0.95	61.5356	19.9866	62.2734	20.7300	0.2637	50.4101	16.6150	51.1562	17.0921	0.2539
	0.99	307.0856	98.6128	307.8316	102.7954	0.2557	249.3244	81.0805	250.0780	83.9060	0.2464

Table B20. Simulation results with gamma response and n = 1000

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	29.0214	9.3668	29.7541	9.2088	0.2720	1.1730	0.6068	1.9594	0.5176	0.2137
0.1	0.95	57.9086	18.3709	58.6554	18.1790	0.2600	1.3884	0.6688	2.1824	0.5785	0.2060
	0.99	289.5640	90.6743	290.3296	90.1641	0.2506	3.6342	1.3676	4.4345	1.2717	0.1998
	0.90	29.0214	9.3668	29.7541	9.2088	0.2720	4.0338	1.5161	4.8154	1.4136	0.2185
0.3	0.95	57.9086	18.3709	58.6554	18.1790	0.2600	6.6265	2.3231	7.4161	2.2142	0.2105
	0.99	289.5640	90.6743	290.3296	90.1641	0.2506	27.4651	8.8525	28.2612	8.6892	0.2039
	0.90	29.0214	9.3668	29.7541	9.2088	0.2720	8.7962	3.0220	9.5684	2.9015	0.2280
0.5	0.95	57.9086	18.3709	58.6554	18.1790	0.2600	16.0485	5.2890	16.8294	5.1510	0.2193
	0.99	289.5640	90.6743	290.3296	90.1641	0.2506	73.9868	23.4544	74.7747	23.1618	0.2122
	0.90	29.0214	9.3668	29.7541	9.2088	0.2720	15.4601	5.1245	16.2183	4.9815	0.2421
0.7	0.95	57.9086	18.3709	58.6554	18.1790	0.2600	29.6545	9.5664	30.4224	9.3889	0.2323
	0.99	289.5640	90.6743	290.3296	90.1641	0.2506	143.1995	45.1733	143.9751	44.6895	0.2245
	0.90	29.0214	9.3668	29.7541	9.2088	0.2720	24.0256	7.8236	24.7651	7.6534	0.2609
0.9	0.95	57.9086	18.3709	58.6554	18.1790	0.2600	47.4446	15.1552	48.1952	14.9279	0.2497
	0.99	289.5640	90.6743	290.3296	90.1641	0.2506	235.1031	74.0092	235.8623	73.2722	0.2409

Table B21. Simulation results with Poisson response and n = 100

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	28.5934	11.1609	31.2301	11.4054	0.2844	1.2374	0.6653	2.0402	0.5681	0.2295
0.1	0.95	57.4946	21.9826	62.2156	22.7807	0.2741	1.4450	0.7474	2.2483	0.6511	0.2236
	0.99	290.8867	108.3680	312.8578	114.2393	0.2661	3.6979	1.5955	4.5010	1.5155	0.2180
	0.90	28.5934	11.1609	31.2301	11.4054	0.2844	4.0791	1.7422	4.9075	1.6391	0.2344
0.3	0.95	57.4946	21.9826	62.2156	22.7807	0.2741	6.6750	2.7243	7.5087	2.6312	0.2281
	0.99	290.8867	108.3680	312.8578	114.2393	0.2661	27.6698	10.5985	28.5078	10.6483	0.2224
	0.90	28.5934	11.1609	31.2301	11.4054	0.2844	8.7707	3.5469	9.6233	3.4438	0.2431
0.5	0.95	57.4946	21.9826	62.2156	22.7807	0.2741	16.0338	6.2959	16.8969	6.2235	0.2362
	0.99	290.8867	108.3680	312.8578	114.2393	0.2661	74.4118	28.1940	75.2842	28.5201	0.2301
	0.90	28.5934	11.1609	31.2301	11.4054	0.2844	15.3123	6.0792	16.1877	5.9822	0.2558
0.7	0.95	57.4946	21.9826	62.2156	22.7807	0.2741	29.5215	11.4622	30.4131	11.4278	0.2479
	0.99	290.8867	108.3680	312.8578	114.2393	0.2661	143.9241	54.3819	144.8302	55.1308	0.2412
	0.90	28.5934	11.1609	31.2301	11.4054	0.2844	23.7039	9.3393	24.6007	9.2544	0.2723
0.9	0.95	57.4946	21.9826	62.2156	22.7807	0.2741	47.1380	18.2231	48.0572	18.2443	0.2631
	0.99	290.8867	108.3680	312.8578	114.2393	0.2661	236.2066	89.1623	237.1457	90.4804	0.2557

Table B22. Simulation results with Poisson response and n = 200

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	30.7105	12.2138	32.2011	11.8859	0.2752	1.2254	0.6582	2.0225	0.5619	0.2139
0.1	0.95	60.7527	24.0159	63.0798	23.4800	0.2640	1.4497	0.7482	2.2511	0.6499	0.2077
	0.99	301.0278	118.7868	310.0946	116.2995	0.2526	3.7785	1.6578	4.5842	1.5424	0.2015
	0.90	30.7105	12.2138	32.2011	11.8859	0.2752	4.2499	1.8099	5.0521	1.6878	0.2193
0.3	0.95	60.7527	24.0159	63.0798	23.4800	0.2640	6.9454	2.8543	7.7539	2.7156	0.2126
	0.99	301.0278	118.7868	310.0946	116.2995	0.2526	28.5607	11.2675	29.3755	10.9749	0.2059
	0.90	30.7105	12.2138	32.2011	11.8859	0.2752	9.2902	3.7526	10.0940	3.5923	0.2292
0.5	0.95	60.7527	24.0159	63.0798	23.4800	0.2640	16.8313	6.6741	17.6435	6.4677	0.2217
	0.99	301.0278	118.7868	310.0946	116.2995	0.2526	76.9276	30.0649	77.7486	29.4308	0.2141
	0.90	30.7105	12.2138	32.2011	11.8859	0.2752	16.3464	6.4863	17.1481	6.2752	0.2436
0.7	0.95	60.7527	24.0159	63.0798	23.4800	0.2640	31.1072	12.2077	31.9200	11.9062	0.2349
	0.99	301.0278	118.7868	310.0946	116.2995	0.2526	148.8792	58.0501	149.7035	56.9103	0.2261
	0.90	30.7105	12.2138	32.2011	11.8859	0.2752	25.4185	10.0110	26.2145	9.7366	0.2624
0.9	0.95	60.7527	24.0159	63.0798	23.4800	0.2640	49.7733	19.4551	50.5834	19.0312	0.2523
	0.99	301.0278	118.7868	310.0946	116.2995	0.2526	244.4154	95.2229	245.2401	93.4134	0.2419

Table B23. Simulation results with Poisson response and n = 300

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	29.6153	10.0287	31.7611	10.0938	0.2783	1.2322	0.6549	2.0277	0.5615	0.2240
0.1	0.95	59.3247	19.6938	62.9979	19.9774	0.2664	1.4498	0.7180	2.2499	0.6266	0.2162
	0.99	297.9529	97.6277	313.8923	99.6760	0.2581	3.7596	1.4641	4.5615	1.3865	0.2112
	0.90	29.6153	10.0287	31.7611	10.0938	0.2783	4.1650	1.6319	4.9763	1.5379	0.2285
0.3	0.95	59.3247	19.6938	62.9979	19.9774	0.2664	6.8319	2.4889	7.6509	2.4094	0.2203
	0.99	297.9529	97.6277	313.8923	99.6760	0.2581	28.2971	9.4620	29.1206	9.5031	0.2149
	0.90	29.6153	10.0287	31.7611	10.0938	0.2783	9.0261	3.2366	9.8511	3.1538	0.2372
0.5	0.95	59.3247	19.6938	62.9979	19.9774	0.2664	16.4877	5.6463	17.3237	5.6044	0.2283
	0.99	297.9529	97.6277	313.8923	99.6760	0.2581	76.1710	25.0394	77.0148	25.3345	0.2224
	0.90	29.6153	10.0287	31.7611	10.0938	0.2783	15.8155	5.4689	16.6519	5.4092	0.2501
0.7	0.95	59.3247	19.6938	62.9979	19.9774	0.2664	30.4172	10.1900	31.2685	10.2115	0.2402
	0.99	297.9529	97.6277	313.8923	99.6760	0.2581	147.3814	48.1965	148.2440	48.8806	0.2335
	0.90	29.6153	10.0287	31.7611	10.0938	0.2783	24.5333	8.3289	25.3788	8.3041	0.2671
0.9	0.95	59.3247	19.6938	62.9979	19.9774	0.2664	48.6204	16.1202	49.4851	16.2308	0.2560
	0.99	297.9529	97.6277	313.8923	99.6760	0.2581	241.9283	78.9330	242.8081	80.1415	0.2484

Table B24. Simulation results with Poisson response and n = 400

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	30.7848	10.1645	32.3040	10.0751	0.2847	1.2492	0.6552	2.0313	0.5654	0.2292
0.1	0.95	61.3707	19.9339	63.7764	19.8870	0.2734	1.4794	0.7232	2.2669	0.6329	0.2222
	0.99	305.0298	97.5975	314.5162	97.9331	0.2663	3.8416	1.4750	4.6314	1.3826	0.2181
	0.90	30.7848	10.1645	32.3040	10.0751	0.2847	4.2842	1.6473	5.0739	1.5452	0.2338
0.3	0.95	61.3707	19.9339	63.7764	19.8870	0.2734	7.0330	2.5256	7.8301	2.4207	0.2264
	0.99	305.0298	97.5975	314.5162	97.9331	0.2663	28.9513	9.5400	29.7526	9.4166	0.2221
	0.90	30.7848	10.1645	32.3040	10.0751	0.2847	9.3355	3.2860	10.1296	3.1692	0.2427
0.5	0.95	61.3707	19.9339	63.7764	19.8870	0.2734	17.0183	5.7497	17.8224	5.6255	0.2346
	0.99	305.0298	97.5975	314.5162	97.9331	0.2663	77.9587	25.2596	78.7692	25.0840	0.2298
	0.90	30.7848	10.1645	32.3040	10.0751	0.2847	16.4030	5.5711	17.1985	5.4374	0.2559
0.7	0.95	61.3707	19.9339	63.7764	19.8870	0.2734	31.4355	10.3957	32.2439	10.2473	0.2468
	0.99	305.0298	97.5975	314.5162	97.9331	0.2663	150.8639	48.6337	151.6813	48.3848	0.2413
	0.90	30.7848	10.1645	32.3040	10.0751	0.2847	25.4868	8.5028	26.2807	8.3499	0.2734
0.9	0.95	61.3707	19.9339	63.7764	19.8870	0.2734	50.2843	16.4634	51.0945	16.2863	0.2630
	0.99	305.0298	97.5975	314.5162	97.9331	0.2663	247.6667	79.6623	248.4888	79.3191	0.2566

Table B25. Simulation results with Poisson response and n = 500

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	29.1453	10.4823	30.2451	10.2481	0.2666	1.1867	0.6286	1.9790	0.5464	0.2130
0.1	0.95	58.3401	20.6028	59.8329	20.2990	0.2566	1.4050	0.7008	2.2027	0.6162	0.2067
	0.99	292.0949	101.8702	296.6694	101.1129	0.2499	3.6703	1.4824	4.4718	1.3897	0.2021
	0.90	29.1453	10.4823	30.2451	10.2481	0.2666	4.0644	1.6324	4.8586	1.5351	0.2169
0.3	0.95	58.3401	20.6028	59.8329	20.2990	0.2566	6.6869	2.5312	7.4875	2.4249	0.2104
	0.99	292.0949	101.8702	296.6694	101.1129	0.2499	27.7135	9.7979	28.5185	9.6280	0.2057
	0.90	29.1453	10.4823	30.2451	10.2481	0.2666	8.8480	3.3050	9.6401	3.1798	0.2254
0.5	0.95	58.3401	20.6028	59.8329	20.2990	0.2566	16.1805	5.8267	16.9803	5.6770	0.2183
	0.99	292.0949	101.8702	296.6694	101.1129	0.2499	74.6430	26.0374	75.4483	25.7097	0.2132
	0.90	29.1453	10.4823	30.2451	10.2481	0.2666	15.5375	5.6465	16.3237	5.4806	0.2383
0.7	0.95	58.3401	20.6028	59.8329	20.2990	0.2566	29.8856	10.5873	30.6812	10.3724	0.2303
	0.99	292.0949	101.8702	296.6694	101.1129	0.2499	144.4590	50.2010	145.2612	49.6347	0.2247
	0.90	29.1453	10.4823	30.2451	10.2481	0.2666	24.1329	8.6568	24.9093	8.4374	0.2557
0.9	0.95	58.3401	20.6028	59.8329	20.2990	0.2566	47.8024	16.8129	48.5902	16.5112	0.2465
	0.99	292.0949	101.8702	296.6694	101.1129	0.2499	237.1613	82.2886	237.9574	81.4031	0.2402

Table B26. Simulation results with Poisson response and n = 600

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	29.7498	9.1261	31.2383	9.1824	0.2779	1.1963	0.6252	1.9901	0.5368	0.2166
0.1	0.95	59.0678	17.7827	61.3714	18.0406	0.2659	1.4155	0.6817	2.2151	0.5954	0.2091
	0.99	293.6343	86.8990	302.4162	88.5403	0.2560	3.6892	1.3474	4.4937	1.2754	0.2026
	0.90	29.7498	9.1261	31.2383	9.1824	0.2779	4.1267	1.5238	4.9256	1.4446	0.2223
0.3	0.95	59.0678	17.7827	61.3714	18.0406	0.2659	6.7585	2.3003	7.5651	2.2411	0.2144
	0.99	293.6343	86.8990	302.4162	88.5403	0.2560	27.8600	8.5508	28.6733	8.6270	0.2076
	0.90	29.7498	9.1261	31.2383	9.1824	0.2779	9.0084	2.9950	9.8090	2.9396	0.2325
0.5	0.95	59.0678	17.7827	61.3714	18.0406	0.2659	16.3693	5.1780	17.1800	5.1781	0.2238
	0.99	293.6343	86.8990	302.4162	88.5403	0.2560	75.0374	22.5665	75.8570	22.9457	0.2164
	0.90	29.7498	9.1261	31.2383	9.1824	0.2779	15.8414	5.0389	16.6405	5.0220	0.2471
0.7	0.95	59.0678	17.7827	61.3714	18.0406	0.2659	30.2478	9.3149	31.0597	9.4064	0.2373
	0.99	293.6343	86.8990	302.4162	88.5403	0.2560	145.2213	43.3944	146.0445	44.2314	0.2291
	0.90	29.7498	9.1261	31.2383	9.1824	0.2779	24.6258	7.6554	25.4200	7.6916	0.2661
0.9	0.95	59.0678	17.7827	61.3714	18.0406	0.2659	48.3942	14.7109	49.2044	14.9259	0.2549
	0.99	293.6343	86.8990	302.4162	88.5403	0.2560	238.4117	71.0346	239.2361	72.4843	0.2457

Table B27. Simulation results with Poisson response and n = 700

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	30.4922	10.1295	31.5624	10.3847	0.2690	1.2144	0.6259	2.0009	0.5451	0.2183
0.1	0.95	60.7094	19.9315	62.1660	20.5798	0.2590	1.4416	0.6975	2.2339	0.6194	0.2119
	0.99	303.3710	97.9897	307.8700	101.8113	0.2507	3.7972	1.4554	4.5940	1.4042	0.2066
	0.90	30.4922	10.1295	31.5624	10.3847	0.2690	4.2134	1.6067	5.0014	1.5479	0.2219
0.3	0.95	60.7094	19.9315	62.1660	20.5798	0.2590	6.9246	2.4889	7.7194	2.4602	0.2152
	0.99	303.3710	97.9897	307.8700	101.8113	0.2507	28.7549	9.5415	29.5553	9.7594	0.2097
	0.90	30.4922	10.1295	31.5624	10.3847	0.2690	9.2165	3.2409	10.0023	3.2276	0.2299
0.5	0.95	60.7094	19.9315	62.1660	20.5798	0.2590	16.8005	5.7125	17.5945	5.7842	0.2226
	0.99	303.3710	97.9897	307.8700	101.8113	0.2507	77.4921	25.3281	78.2930	26.0856	0.2166
	0.90	30.4922	10.1295	31.5624	10.3847	0.2690	16.2237	5.5283	17.0035	5.5841	0.2422
0.7	0.95	60.7094	19.9315	62.1660	20.5798	0.2590	31.0694	10.3684	31.8592	10.5912	0.2340
	0.99	303.3710	97.9897	307.8700	101.8113	0.2507	150.0090	48.8153	150.8072	50.3828	0.2273
	0.90	30.4922	10.1295	31.5624	10.3847	0.2690	25.2350	8.4690	26.0050	8.6176	0.2587
0.9	0.95	60.7094	19.9315	62.1660	20.5798	0.2590	49.7312	16.4565	50.5134	16.8814	0.2494
	0.99	303.3710	97.9897	307.8700	101.8113	0.2507	246.3054	80.0030	247.0979	82.6510	0.2417

Table B28. Simulation results with Poisson response and n = 800

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	30.1563	10.2279	30.9995	9.9267	0.2666	1.2082	0.6329	1.9987	0.5441	0.2113
0.1	0.95	60.0867	20.0413	61.0733	19.5458	0.2575	1.4335	0.7025	2.2299	0.6101	0.2052
	0.99	299.6464	98.7883	301.7316	96.7418	0.2499	3.7574	1.4645	4.5581	1.3524	0.2006
	0.90	30.1563	10.2279	30.9995	9.9267	0.2666	4.1814	1.6256	4.9695	1.5113	0.2157
0.3	0.95	60.0867	20.0413	61.0733	19.5458	0.2575	6.8701	2.5048	7.6645	2.3682	0.2095
	0.99	299.6464	98.7883	301.7316	96.7418	0.2499	28.4215	9.6027	29.2209	9.2942	0.2046
	0.90	30.1563	10.2279	30.9995	9.9267	0.2666	9.1316	3.2706	9.9129	3.1131	0.2245
0.5	0.95	60.0867	20.0413	61.0733	19.5458	0.2575	16.6480	5.7361	17.4366	5.5184	0.2179
	0.99	299.6464	98.7883	301.7316	96.7418	0.2499	76.5650	25.4768	77.3595	24.7803	0.2125
	0.90	30.1563	10.2279	30.9995	9.9267	0.2666	16.0588	5.5680	16.8290	5.3495	0.2379
0.7	0.95	60.0867	20.0413	61.0733	19.5458	0.2575	30.7674	10.3964	31.5461	10.0606	0.2304
	0.99	299.6464	98.7883	301.7316	96.7418	0.2499	148.1880	49.0869	148.9738	47.8107	0.2244
	0.90	30.1563	10.2279	30.9995	9.9267	0.2666	24.9629	8.5178	25.7177	8.2207	0.2556
0.9	0.95	60.0867	20.0413	61.0733	19.5458	0.2575	49.2282	16.4859	49.9931	15.9949	0.2472
	0.99	299.6464	98.7883	301.7316	96.7418	0.2499	243.2904	80.4329	244.0638	78.3853	0.2402

Table B29. Simulation results with Poisson response and n = 900

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	29.3950	9.3950	30.4309	9.5818	0.2676	1.1864	0.6134	1.9774	0.5261	0.2133
0.1	0.95	58.6737	18.3986	60.0636	18.8864	0.2570	1.4065	0.6754	2.2041	0.5905	0.2062
	0.99	292.5257	90.3224	296.7109	93.5075	0.2492	3.6737	1.3723	4.4757	1.3122	0.2012
	0.90	29.3950	9.3950	30.4309	9.5818	0.2676	4.0840	1.5300	4.8756	1.4577	0.2175
0.3	0.95	58.6737	18.3986	60.0636	18.8864	0.2570	6.7150	2.3391	7.5142	2.2896	0.2101
	0.99	292.5257	90.3224	296.7109	93.5075	0.2492	27.7514	8.8481	28.5559	9.0142	0.2049
	0.90	29.3950	9.3950	30.4309	9.5818	0.2676	8.9080	3.0442	9.6960	3.0086	0.2261
0.5	0.95	58.6737	18.3986	60.0636	18.8864	0.2570	16.2621	5.3155	17.0592	5.3449	0.2182
	0.99	292.5257	90.3224	296.7109	93.5075	0.2492	74.7500	23.4210	75.5537	24.0473	0.2125
	0.90	29.3950	9.3950	30.4309	9.5818	0.2676	15.6581	5.1560	16.4387	5.1791	0.2392
0.7	0.95	58.6737	18.3986	60.0636	18.8864	0.2570	30.0478	9.6047	30.8392	9.7564	0.2304
	0.99	292.5257	90.3224	296.7109	93.5075	0.2492	144.6696	45.0912	145.4692	46.4113	0.2241
	0.90	29.3950	9.3950	30.4309	9.5818	0.2676	24.3345	7.8655	25.1037	7.9689	0.2568
0.9	0.95	58.6737	18.3986	60.0636	18.8864	0.2570	48.0721	15.2065	48.8542	15.5242	0.2468
	0.99	292.5257	90.3224	296.7109	93.5075	0.2492	237.5101	73.8585	238.3024	76.1064	0.2396

Table B30. Simulation results with Poisson response and n = 1000

d	corr	EMSEs of Estimators									
		ML	Restricted ML				Liu	Restricted Liu			
			Restrictions					Restrictions			
			(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	0.90	30.0800	10.1238	31.0495	10.1334	0.2765	1.2220	0.6406	2.0077	0.5543	0.2178
0.1	0.95	59.9603	19.9813	61.2093	20.1167	0.2667	1.4447	0.7114	2.2360	0.6250	0.2116
	0.99	299.5451	98.5618	302.9818	99.6335	0.2589	3.7670	1.4701	4.5626	1.3887	0.2068
	0.90	30.0800	10.1238	31.0495	10.1334	0.2765	4.1932	1.6263	4.9776	1.5408	0.2229
0.3	0.95	59.9603	19.9813	61.2093	20.1167	0.2667	6.8758	2.5116	7.6666	2.4326	0.2164
	0.99	299.5451	98.5618	302.9818	99.6335	0.2589	28.4283	9.5852	29.2242	9.5561	0.2114
	0.90	30.0800	10.1238	31.0495	10.1334	0.2765	9.1310	3.2556	9.9101	3.1767	0.2324
0.5	0.95	59.9603	19.9813	61.2093	20.1167	0.2667	16.6339	5.7342	17.4206	5.6747	0.2255
	0.99	299.5451	98.5618	302.9818	99.6335	0.2589	76.5565	25.4075	77.3493	25.4878	0.2199
	0.90	30.0800	10.1238	31.0495	10.1334	0.2765	16.0356	5.5285	16.8051	5.4621	0.2465
0.7	0.95	59.9603	19.9813	61.2093	20.1167	0.2667	30.7192	10.3795	31.4978	10.3515	0.2387
	0.99	299.5451	98.5618	302.9818	99.6335	0.2589	148.1518	48.9371	148.9378	49.1836	0.2324
	0.90	30.0800	10.1238	31.0495	10.1334	0.2765	24.9068	8.4450	25.6628	8.3969	0.2651
0.9	0.95	59.9603	19.9813	61.2093	20.1167	0.2667	49.1315	16.4472	49.8984	16.4629	0.2561
	0.99	299.5451	98.5618	302.9818	99.6335	0.2589	243.2139	80.1738	243.9898	80.6436	0.2489