Generalized Minimum Penalized Hellinger Distance Estimation
and Generalized Penalized Hellinger Deviance Testing for
Generalized Linear Models: The Discrete Case

Huey Yan
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd
Part of the Mathematics Commons, and the Statistics and Probability Commons

Recommended Citation
Yan, Huey, "Generalized Minimum Penalized Hellinger Distance Estimation and Generalized Penalized
Hellinger Deviance Testing for Generalized Linear Models: The Discrete Case" (2001). All Graduate Theses
and Dissertations. 7066.
https://digitalcommons.usu.edu/etd/7066
GENERALIZED MINIMUM PENALIZED HELLINGER DISTANCE ESTIMATION
AND GENERALIZED PENALIZED HELLINGER DEVIANCE TESTING
FOR GENERALIZED LINEAR MODELS: THE DISCRETE CASE

by

Hucy Yan

A dissertation submitted in partial fulfillment
of the requirements for the degree
of
DOCTOR OF PHILOSOPHY
in
Mathematical Sciences

UTAH STATE UNIVERSITY
Logan, Utah
2001
ABSTRACT

Generalized Minimum Penalized Hellinger Distance Estimation
and Generalized Penalized Hellinger Deviance Testing
for Generalized Linear Models: The Discrete Case

by

Huey Yan, Doctor of Philosophy
Utah State University, 2001

Major Professor: Dr. Richard Cutler
Department: Mathematics and Statistics

In this dissertation, robust and efficient alternatives to quasi-likelihood estimation and likelihood ratio tests are developed for discrete generalized linear models. The estimation method considered is a penalized minimum Hellinger distance procedure that generalizes a procedure developed by Harris and Basu for estimating parameters of a single discrete probability distribution from a random sample. A bootstrap algorithm is proposed to select the weight of the penalty term. Simulations are carried out to compare the new estimators with quasi-likelihood estimation. The robustness of the estimation procedure is demonstrated by simulation work and by Hampel's α-influence curve. Penalized minimum Hellinger deviance tests for goodness-of-fit and for testing nested linear hypotheses are proposed and simulated. A nonparametric bootstrap algorithm is proposed to obtain critical values for the testing procedure.
ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my major professor, Dr. Richard Cutler. His clear vision of the topics, effective guidance, patience, careful proofreading of this dissertation, and most importantly the confidence he had in me throughout this research made this dissertation possible.

I also would like to express my sincere appreciation to my committee member, Dr. Adele Cutler, for her expert help in statistical computing, kind proofreading of this dissertation, and friendship. This dissertation would never be finished without her consistent encouragement.

My thanks certainly go to my committee members, Dr. Dan Coster, Dr. Michael Minnotte, and Dr. Thomas Edwards, for their precious time in reading this dissertation and for their valuable comments and suggestions.

Special acknowledgment is given to Dr. Ayanedranath Basu at the Indian Institute of Statistics, India, to Dr. Marianthi Markatou at Columbia University, and to Mr. Chansoek Park at the University of Texas at Austin for kindly sending me their preprints or technical reports when I needed them during this research. I also thank Dr. Bruce Lindsay at Pennsylvania State University for referring me to these authors.

I am grateful to my best friend, Miss Shu-fang Ting, for the unceasing encouragement and support she gave me during the writing of this dissertation.

Finally, I dedicate this dissertation to my parents. I can never repay them enough for the endless love they have given me since the day I was born. Their unconditional support, understanding, and trust throughout my entire graduate education provided me the strength to complete this dissertation. I would like to let them know that I am blessed to be their daughter and that I love them very much.

Huey Yan
CONTENTS

ABSTRACT ................................................................. iii

ACKNOWLEDGMENTS ....................................................... iv

LIST OF TABLES .......................................................... viii

LIST OF FIGURES ........................................................... ix

1 INTRODUCTION AND BACKGROUND ................................... 1
  1.1 Introduction ....................................................... 1
  1.2 Minimum Distance Estimation ...................................... 2
  1.3 Minimum Hellinger Distance Estimation ........................... 4
  1.4 Minimum Penalized Hellinger Distance Estimation for Discrete Models . 7
  1.5 Hellinger Deviance Testing ........................................ 12
  1.6 Penalized Hellinger Deviance Testing for Discrete Models .......... 15
  1.7 Generalized Linear Models ......................................... 16
  1.8 Generalized Maximum Likelihood Estimation ....................... 18

2 GENERALIZED MINIMUM PENALIZED HELLINGER DISTANCE ESTIMATION AND GENERALIZED PENALIZED HELLINGER DEVIANCE TESTING FOR DISCRETE GENERALIZED LINEAR MODELS ................................................ 21
  2.1 Generalized Minimum Hellinger Distance Estimation and Its Application in GLMs ......................................................... 21
  2.2 Generalized Minimum Hellinger Distance Estimation for Discrete GLMs . 24
  2.3 Generalized Minimum Penalized Hellinger Distance Estimation for Discrete GLMs ......................................................... 25
    2.3.1 Examples ....................................................... 26
    2.3.2 The Choice of the Penalty Factor ............................. 27
  2.4 Generalized Hellinger Deviance Testing for Discrete GLMs ............ 33
    2.4.1 The GHDV Test for Goodness-of-Fit ........................... 33
    2.4.2 The GHDV Test for Comparing Two Nested Models ............... 35
  2.5 Generalized Penalized Hellinger Deviance Testing for Discrete GLMs ....... 37
    2.5.1 The GPHDV Test for Goodness-of-Fit ........................... 37
    2.5.2 The GPHDV Test for Comparing Two Nested Models ............... 39
  2.6 Nonparametric Bootstrap GPHDV Testing for Discrete GLMs ............ 40
    2.6.1 The NPB-GPHDV Test for Goodness-of-Fit ........................ 41
    2.6.2 The NPB-GPHDV Test for Comparing Two Nested Models .......... 42
B  SIMULATION RESULTS FOR HYPOTHESIS TESTING of 5% AND 1%
    NOMINAL LEVELS OF SECTION 4.2.3.2 ........................................ 206
C  SIMULATION RESULTS FOR HYPOTHESIS TESTING of 5% AND 1%
    NOMINAL LEVELS OF SECTION 4.3.3.1 ........................................ 217
D  SIMULATION RESULTS FOR HYPOTHESIS TESTING of 5% AND 1%
    NOMINAL LEVELS OF SECTION 4.3.3.2 ........................................ 242
CURRICULUM VITAE ................................................................. 263


LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Some Commonly Used Metrics on Distributions</td>
<td>3</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>4.1</td>
<td>Plots of penalty weights $h$ where the optimal MPHD estimators occur (according to MSE) vs. sample sizes $n$ for iid Poisson models.</td>
</tr>
<tr>
<td>4.2</td>
<td>Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\tilde{\mu}_i)}{\mu_i^2}$) vs. common sample sizes $n_c$ for Poisson GLMs. (No contamination)</td>
</tr>
<tr>
<td>4.3</td>
<td>Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\tilde{\mu}_i)}{\mu_i^2}$) vs. common sample sizes $n_c$ for Poisson GLMs. (Type I contamination)</td>
</tr>
<tr>
<td>4.4</td>
<td>Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\tilde{\mu}_i)}{\mu_i^2}$) vs. common sample sizes $n_c$ for Poisson GLMs. (Type II contamination)</td>
</tr>
<tr>
<td>4.5</td>
<td>Plots of adjusted total MSEs, $n_c * \sum_{i=1}^{m} \frac{\text{MSE}(\tilde{\mu}_i)}{\mu_i^2}$, of the QL estimates, the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, and the ordinary GMHD estimates vs. the common sample sizes $n_c$ (of 1 to 20) for Poisson GLMs. (No contamination)</td>
</tr>
<tr>
<td>4.6</td>
<td>Plots of adjusted total MSEs, $n_c * \sum_{i=1}^{m} \frac{\text{MSE}(\tilde{\mu}_i)}{\mu_i^2}$, of the QL estimates, the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, and the ordinary GMHD estimates vs. the common sample sizes $n_c$ (of 1 to 1000) for Poisson GLMs. (No contamination)</td>
</tr>
</tbody>
</table>
4.7 Plots of adjusted total MSEs, $n_c \star \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$, of the QL estimates, the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, the GMPHD estimates using PBoot optimal $h$, and the GMPHD estimates using NPBoot optimal $h$ vs. the common sample sizes $n_c$ for Poisson GLMs. (No contamination) .................................. 69

4.8 Plots of total MSEs, $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$, and the adjusted total MSEs, $n_c \star \text{MSE}(\hat{\mu})$, of the QL estimates (for plots on the left only), the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, the GMPHD estimates using PBoot optimal $h$, and the GMPHD estimates using NPBoot optimal $h$ vs. the common sample sizes $n_c$ for Poisson GLMs. (Type I contamination) .................................. 70

4.9 Plots of total MSEs, $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$, and the adjusted total MSEs, $n_c \star \text{MSE}(\hat{\mu})$, of the QL estimates (for plots on the left only), the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, the GMPHD estimates using PBoot optimal $h$, and the GMPHD estimates using NPBoot optimal $h$ vs. the common sample sizes $n_c$ for Poisson GLMs. (Type II contamination) .................................. 71

4.10 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, and the ordinary GHDV tests vs. common sample sizes $n_c$ (of 1 to 100) for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.1) .................................. 73

4.11 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, and the ordinary GHDV tests vs. common sample sizes $n_c$ (of 1 to 100) for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.1) .................................. 74
4.12 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.1) ........................................ 75

4.13 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.1) ...................... 76

4.14 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type I contamination; Nominal level 0.1) ...................... 77

4.15 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type I contamination; Nominal level 0.1) ...................... 78

4.16 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type II contamination; Nominal level 0.1) ...................... 79
4.17 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type II contamination; Nominal level 0.1) ........................................ 80

4.18 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.1) ........................................ 84

4.19 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.1) ........................................ 85

4.20 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type I contamination; Nominal level 0.1) ........................................ 86

4.21 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type I contamination; Nominal level 0.1) ........................................ 87
4.22 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type II contamination; Nominal level 0.1)

4.23 Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i}$) vs. common sample sizes $n_c$ for binomial GLMs. (B1, P, and N models; No contamination)

4.24 Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i}$) vs. common sample sizes $n_c$ for binomial GLMs. (B2 and B3 models; No contamination)

4.25 Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i}$) vs. common sample sizes $n_c$ for binomial GLMs. (B1, P, and N models; Type I contamination)

4.26 Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i}$) vs. common sample sizes $n_c$ for binomial GLMs. (B2 and B3 models; Type I contamination)

4.27 Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i}$) vs. common sample sizes $n_c$ for binomial GLMs. (B1, P, and N models; Type II contamination)

4.28 Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i}$) vs. common sample sizes $n_c$ for binomial GLMs. (B2 and B3 models; Type II contamination)
4.29 Plots of adjusted total MSEs, $n_c \sum_{i=1}^m \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$, of the QL estimates, the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, and the GMPHD estimates using PBoot optimal $h$ vs. common sample sizes $n_c$ for binomial GLMs. (B1, P, and N models; No contamination) .................................................. 97

4.30 Plots of adjusted total MSEs, $n_c \sum_{i=1}^m \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$, of the QL estimates, the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, and the GMPHD estimates using PBoot optimal $h$ vs. common sample sizes $n_c$ for binomial GLMs. (B2 and B3 models; No contamination) .................................................. 98

4.31 Plots of total MSEs, $\text{MSE}(\mu) = \sum_{i=1}^m \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$, and the adjusted total MSEs, $n_c \text{MSE}(\mu)$, of the QL estimates (for plots on the left only), the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, and the GMPHD estimates using PBoot optimal $h$ vs. common sample sizes $n_c$ for binomial GLMs. (B1, P, and N models; Type I contamination) .................................................. 99

4.32 Plots of total MSEs, $\text{MSE}(\mu) = \sum_{i=1}^m \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$, and the adjusted total MSEs, $n_c \text{MSE}(\mu)$, of the QL estimates (for plots on the left only), the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, and the GMPHD estimates using PBoot optimal $h$ vs. common sample sizes $n_c$ for binomial GLMs. (B2 and B3 models; Type I contamination) 100

4.33 Plots of total MSEs, $\text{MSE}(\mu) = \sum_{i=1}^m \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$, and the adjusted total MSEs, $n_c \text{MSE}(\mu)$, of the QL estimates (for plots on the left only), the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, and the GMPHD estimates using PBoot optimal $h$ vs. common sample sizes $n_c$ for binomial GLMs. (B1, P, and N models; Type II contamination) 101
4.34 Plots of total MSEs, $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{n_c} \frac{\text{MSE}(\hat{\mu}_i)}{\hat{\mu}_i^2}$, and the adjusted total MSEs, $n_c \cdot \text{MSE}(\hat{\mu})$, of the QL estimates (for plots on the left only), the optimal GM-PHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, and the GMPHD estimates using PBoot optimal $h$ vs. common sample sizes $n_c$ for binomial GLMs. (B2 and B3 models; Type II contamination) ........................................... 102

4.35 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.1) ........................................... 104

4.36 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.1) ........................................... 105

4.37 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.1) ........................................... 106

4.38 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.1) ........................................... 107
4.39 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.1) ............................................. 108

4.40 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.1) ............................................. 109

4.41 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.1) ............................................. 110

4.42 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.1) ............................................. 111

4.43 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.1) ............................................. 112
4.44 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.1) .................................................. 113

4.45 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.1) .................................................. 114

4.46 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.1) .................................................. 115

4.47 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.1) .................................................. 117

4.48 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.1) .................................................. 118
4.49 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.1) .................................................. 119

4.50 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.1) .................................................. 120

4.51 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.1) .................................................. 121

4.52 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.1) .................................................. 122

4.53 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.1) .................................................. 123
4.54 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.1) ................................. 124

4.55 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.1) .................................................. 125

4.56 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.1) .................................................. 126

5.1 $\alpha$-IC of QL and GMHD estimators for Poisson GLMs under Type I contamination ................................................................. 172

5.2 $\alpha$-IC of QL and GMHD estimators for binomial GLMs under Type I contamination (B1, P, and N models) ........................................ 174

5.3 $\alpha$-IC of QL and GMHD estimators for binomial GLMs under Type I contamination (B2 and B3 models) ........................................ 175

5.4 Empirical $\alpha$-IC before eliminating the bias caused by estimation of QL estimator, the optimal GMPHD estimator, the GMPHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLM with means approximately equal to 2 under Type I contamination for $n_c$ of sizes 1, 20, and 100 178
5.5 Empirical $\alpha$-IC of QL estimator, the optimal GM-PHD estimator, the GM-PHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLM with means approximately equal to 2 under Type I contamination for $n_c$ of sizes 1, 20, and 100 ............................................. 179

5.6 Empirical $\alpha$-IC of QL estimator, the optimal GM-PHD estimator, the GM-PHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLMs under Type I contamination for $n_c$ of size 100 ............................. 180

5.7 Empirical $\alpha$-IC of QL estimator, the optimal GM-PHD estimator, the GM-PHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLMs under Type II contamination for $n_c$ of size 100 ............................. 181

5.8 Empirical $\alpha$-IC of QL estimator, the optimal GM-PHD estimator, the GM-PHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLMs under Type I contamination for $n_c$ of size 100 (B1, P, and N models) ................................................................. 182

5.9 Empirical $\alpha$-IC of QL estimator, the optimal GM-PHD estimator, the GM-PHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLMs under Type I contamination for $n_c$ of size 100 (B2 and B3 models) ................................................................. 183

5.10 Empirical $\alpha$-IC of QL estimator, the optimal GM-PHD estimator, the GM-PHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLMs under Type II contamination for $n_c$ of size 100 (B1, P, and N models) ................................................................. 184

5.11 Empirical $\alpha$-IC of QL estimator, the optimal GM-PHD estimator, the GM-PHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLMs under Type II contamination for $n_c$ of size 100 (B2 and B3 models) ................................................................. 185
A.1 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.05) 194

A.2 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.05) 195

A.3 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type I contamination; Nominal level 0.05) 196

A.4 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type I contamination; Nominal level 0.05) 197

A.5 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type II contamination; Nominal level 0.05) 198
A.6 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type II contamination; Nominal level 0.05) ........................................ 199

A.7 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.01) ........................................ 200

A.8 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.01) ........................................ 201

A.9 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type I contamination; Nominal level 0.01) ........................................ 202

A.10 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type I contamination; Nominal level 0.01) ........................................ 203
A.11 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type II contamination; Nominal level 0.01) 204

A.12 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type II contamination; Nominal level 0.01) 205

B.1 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.05) 207

B.2 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.05) 208

B.3 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type I contamination; Nominal level 0.05) 209
B.4 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for comparing two nested Poisson GLMs. (Type I contamination; Nominal level 0.05) .......................................................... 210

B.5 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for comparing two nested Poisson GLMs. (Type II contamination; Nominal level 0.05) .......................................................... 211

B.6 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.01) .......................................................... 212

B.7 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.01) .......................................................... 213

B.8 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for testing goodness-of-fit of Poisson GLMs. (Type I contamination; Nominal level 0.01) .......................................................... 214
B.9 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type I contamination; Nominal level 0.01) ........................................... 215

B.10 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type II contamination; Nominal level 0.01) ........................................... 216

C.1 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.05) ........................................... 218

C.2 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.05) ........................................... 219

C.3 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.05) ........................................... 220
C.4 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.05) ................................. 221

C.5 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.05) ................................. 222

C.6 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.05) ................................. 223

C.7 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.05) ................................. 224

C.8 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.05) ................................. 225
C.9 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.05) ......................................................... 226

C.10 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.05) .......................................................... 227

C.11 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.05) .......................................................... 228

C.12 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.05) .......................................................... 229

C.13 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.01) .......................................................... 230
C.14 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.01) .................. 231

C.15 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.01) .................. 232

C.16 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.01) .................. 233

C.17 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.01) .................. 234

C.18 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.01) .................. 235
C.19 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.01) .......................................................... 236

C.20 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.01) .......................................................... 237

C.21 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.01) .......................................................... 238

C.22 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.01) .......................................................... 239

C.23 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.01) .......................................................... 240
C.24 Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.01) ................................................................. 241

D.1 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.05) ................................................................. 243

D.2 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.05) ................................................................. 244

D.3 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.05) ................................................................. 245

D.4 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.05) ................................................................. 246
D.5 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.05) .............................................. 247

D.6 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.05) .............................................. 248

D.7 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.05) .............................................. 249

D.8 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.05) .............................................. 250

D.9 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.05) .............................................. 251
D.10 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.05) .......................... 252

D.11 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.01) .......................................................... 253

D.12 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.01) .......................................................... 254

D.13 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.01) .......................................................... 255

D.14 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.01) .......................................................... 256
D.15 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.01) .................................................... 257

D.16 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.01) .................................................... 258

D.17 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.01) .................................................... 259

D.18 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.01) .................................................... 260

D.19 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.01) .................................................... 261
D.20 Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.01) .................. 262
CHAPTER 1
INTRODUCTION AND BACKGROUND

1.1 Introduction

In parametric inference, most classical estimation procedures (e.g., maximum likelihood [ML], method-of-moments [MOM], UMVU, Bayes') are very sensitive to tiny deviations from the assumed probability model and often require that the data are distributed exactly according to the assumed model. However, in reality, data frequently suffer from contamination, discretization and many other kinds of perturbations. There may be outliers caused by recording errors or rounding errors, there may be legitimate outlying observations, or the data may come from a model which is different from the one specified. Any of the above departures from the assumed model may invalidate traditional analyses. The standard outlier detection and model diagnostic techniques are often not sensitive enough and are limited in their scope. In the last few decades, tremendous efforts have been made to develop robust estimation procedures for independently and identically distributed (iid) data, and regression model data. The aim of this dissertation is to develop robust and efficient estimation and testing procedures for generalized linear models (GLMs). The procedures studied are generalized minimum Hellinger distance (GMHD) estimation, generalized minimum penalized Hellinger distance (GMPHD) estimation, generalized Hellinger deviance (GHDV) tests, and generalized penalized Hellinger deviance (GPHDV) tests.

The organization of this dissertation is as follows. The rest of the Chapter 1 contains a review of the background studies that motivate the research presented in this dissertation. Chapter 2 introduces the new estimation and testing procedures for GLMs, including the GMHD estimation, GMPHD estimation, GHDV tests, and GPHDV tests. In Chapter 3, a general fitting algorithm for calculating the GMHD estimators is derived. These estimates may be used as robust starting values of the parameters for the GMPHD estimation algo-
rithm. Chapter 4 contains the results of an extensive simulation study of the performance of the generalized procedures for GLMs, while the asymptotic and robustness properties of the procedures are discussed in Chapter 5. Chapter 6 contains concluding remarks.

1.2 Minimum Distance Estimation

The modern development of minimum distance (MD) methods dates back to a series of papers by Wolfowitz (1952, 1954, and 1957) in the 1950’s. In his seminal 1957 paper, Wolfowitz outlined the early work on MD methodology, provided some interesting examples of its use, and proved some basic results concerning consistency of MD estimates. Over the next four decades there have been a substantial number of papers on MD methods. For references prior to 1981 see the bibliography provided by Parr (1981).

MD estimation may be formulated as follows. Let \( \mathcal{F} = \{ F_\theta : \theta \in \Theta \} \) be a parametric family of distributions, and assume that \( G \) is a distribution that is either in \( \mathcal{F} \) or at least close to one of the members of \( \mathcal{F} \). Ignoring some technicalities for the moment, the object of MD estimation is to identify the \( F_\theta \in \mathcal{F} \) which is closest to \( G \) in some metric, \( u \), on distributions. That is, we want to find

\[
\theta_0 = \arg \min_{\theta \in \Theta} u(F_\theta, G).
\]

Let \( Y_1, Y_2, \ldots, Y_n \) be a random sample from \( G \) and let \( \hat{G}_n \) be an empirical representation (typically a cumulative distribution function or density estimate) of \( G \) based on the data. The MD estimator \( \theta \) of \( \theta_0 \) based on \( Y_1, Y_2, \ldots, Y_n \) is given by

\[
\hat{\theta} = \arg \min_{\theta \in \Theta} u(F_\theta, \hat{G}_n).
\]

It should be mentioned that even in the cases where \( G \notin \mathcal{F} \), the MD estimation (under regularity conditions) is associated with the best approximation to \( G \) in \( \mathcal{F} \). Parr and DeWet (1981) referring to Parr and Schucany (1980) gave more detailed discussion of this point.

Some commonly used metrics on distributions are given in Table 1.1. Upper case letters \((F, G)\) denote cumulative distribution functions; lower case letters \((f, g)\) denote density
<table>
<thead>
<tr>
<th>Distance Name</th>
<th>Distance Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>$u(F, G) = \sup_t</td>
</tr>
<tr>
<td>Kuiper</td>
<td>$u(F, G) = \sup_{A=[a,b]}</td>
</tr>
<tr>
<td>Variation</td>
<td>$u(F, G) = \sup_{\text{measurable } A}</td>
</tr>
<tr>
<td>Levy</td>
<td>$u(F, G) = \inf {\varepsilon &gt; 0 : F(t) \leq G(t + \varepsilon) + \varepsilon}$</td>
</tr>
<tr>
<td>Prohorov</td>
<td>$u(F, G) = \inf {\varepsilon &gt; 0 : F(A) \leq G(A^\varepsilon) + \varepsilon \text{ for all measurable } A}$</td>
</tr>
<tr>
<td>Cramer-Von Mises</td>
<td>$u(F, G) = \left[\int (F(t) - G(t))^2 dF(t)\right]^{1/2}$</td>
</tr>
<tr>
<td>$L_2$</td>
<td>$u(f, g) = \left[\int (f(t) - g(t))^2 dt\right]^{1/2}$</td>
</tr>
<tr>
<td>$L_1$</td>
<td>$u(f, g) = \int</td>
</tr>
<tr>
<td>Hellinger</td>
<td>$u(f, g) = \left[\int (f^{1/2}(t) - g^{1/2}(t))^2 dt\right]^{1/2}$</td>
</tr>
</tbody>
</table>

Donoho and Liu (1988) indicated an attractive feature of the MD estimator, namely that it is “automatically robust” over contamination neighborhoods defined by the metric $u$, as opposed to the ML estimator’s attractive “automatically efficient” feature. On the subject of stability of variance (one of the two quantitative robustness notions indicated by Donoho and Liu in their 1988 paper), they pointed out that some MD estimators can have arbitrarily large asymptotic variance or even be inconsistent at some distributions arbitrarily near the model. However, Donoho and Liu (1988) also showed that a subclass of the MD estimators, those defined by metrics based on quadratic measures of deviation (called Hilbertian metrics because they generate Hilbert spaces of distributions), avoid this problem. The distance measure that is the topic of this dissertation, Hellinger distance, is
Hilbertian, as are Cramer-Von Mises type distance and $L_2$ distance.

1.3 Minimum Hellinger Distance Estimation

Minimum Hellinger distance (MHD) estimation is an example of MD estimation. Let $Y_1, Y_2, \ldots, Y_n$ be iid from a distribution with probability distribution function $g$ and let $\mathcal{F} = \{f_{\theta} : \theta \in \Theta\}$ be a parametric family of distributions. Assume that $g$ is either in $\mathcal{F}$ or at least close to a member $f_{\theta_0}$ of $\mathcal{F}$. Then, the MHD estimator of $\theta_0$ based on $Y_1, Y_2, \ldots, Y_n$ is any $\theta$ in $\Theta$ that minimizes

$$
\mu_n(f_{\theta}, \hat{g}_n) := \|f_{\theta}^{1/2} - \hat{g}_n^{1/2}\|_2
= \int (f_{\theta}^{1/2}(t) - \hat{g}_n^{1/2}(t))^2 dt)^{1/2}
= [2 - 2 \int f_{\theta}^{1/2}(t) \hat{g}_n^{1/2}(t) dt]^{1/2}
$$

(1.1)
or, equivalently, maximizes

$$
\int f_{\theta}^{1/2}(t) \hat{g}_n^{1/2}(t) dt,
$$

where $\hat{g}_n$ is a nonparametric estimate of the true probability distribution function of the data, $g$, based on $Y_1, Y_2, \ldots, Y_n$.

For discrete models in which the density function has countable support, one can simply take $\hat{g}_n$ to be the empirical density function such that for each $t$ value in the sample space

$$
\hat{g}_n(t) = \frac{N_t}{n},
$$

(1.2)
where $N_t$ is the number of sample observations equal to the value $t$. Then, the MHD estimator is the maximizer of

$$
\sum_t f_{\theta}^{1/2}(t) \hat{g}_n^{1/2}(t).
$$

(1.3)
No continuous nonparametric density estimation is needed.

MHD estimators have a second attractive feature beyond other MD estimators in that, under regularity conditions, they may be shown to be asymptotically efficient when the assumed parametric model is true. Beran (1977) introduced MHD estimation as an efficient
and robust parametric estimation procedure. He concentrated on estimation for continuous distributions with compact support and established the consistency as well as asymptotic efficiency under the assumed model of MHD estimates.

Beran (1977) gave an interesting explanation of why MHD estimates should be asymptotically efficient under the assumed model by showing, in a heuristic way, that the MHD estimator is asymptotically equivalent to the ML estimator if $g$ is in fact a member of $\mathcal{F}$ (but not otherwise!). Here is a review of the explanation. With the assumption that $g = f_{\theta_0} \in \mathcal{F}$ and $n$ is sufficiently large, the ML estimate $\hat{\theta}_{mle}$ of $\theta$ should be close to $\theta_0$ and $\hat{g}_n$ should be close to $f_{\theta_0} = g$. Then, according to Beran (1977), finding the ML estimator amounts to maximizing

$$\int \log f_{\theta}(t)d\hat{G}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \log f_{\theta}(y_i) \quad \text{over} \quad \theta \in \Theta,$$

where $\hat{G}_n$ is the empirical cdf of the data, and this procedure should be nearly the same as maximizing

$$\int \log[f_{\theta}(t)/\hat{g}_n(t)]\hat{g}_n(t)dt = 2 \int \log[1 + (f_{\theta}^{1/2}(t)/\hat{g}_n^{1/2}(t) - 1)]\hat{g}_n(t)dt$$

$$\approx 2 \int \left[(f_{\theta}^{1/2}(t)/\hat{g}_n^{1/2}(t) - 1) - \frac{1}{2}(f_{\theta}^{1/2}(t)/\hat{g}_n^{1/2}(t) - 1)^2 \right] \hat{g}_n(t)dt$$

$$= -2\|f_{\theta}^{1/2} - \hat{g}_n^{1/2}\|_2^2$$

(1.4)

where $\approx$ stands for “is asymptotically equivalent to.” Thus, if $g \in \mathcal{F}$, the MHD estimator should be asymptotically as efficient as the ML estimator.

One can view the ML estimator as a MD estimator. Define two related measures of divergence or disparity between distributions $f$ and $g$ as follows. The Kullback-Leibler divergence is given by

$$KL(f, g) = \int \log[g(t)/f(t)]g(t)dt,$$

and the likelihood disparity is given by

$$LD(f, g) = \int \log[f(t)/g(t)]f(t)dt,$$
(Lindsay 1994, and Basu and Lindsay 1994). Obviously, $KL(f_\theta, g) = LD(g, f_\theta)$.

First observe that the left-hand side of the Equation (1.4) is equal to $-KL(f_\theta, \hat{g}_n)$ or equivalently $-LD(\hat{g}_n, f_\theta)$. Thus, the ML estimator can be interpreted as a MD estimator in the sense that it minimizes the Kullback-Leibler distance (or equivalently the likelihood disparity) between the model density and the density estimate based on data. Note that for discrete models with $\hat{g}_n$ being the empirical density function, the maximizer $\hat{\theta}_{mle}$ of $\frac{1}{n} \sum_{i=1}^{n} \log f_\theta(y_i) = \sum_t \log [f_\theta(t)] \hat{g}_n(t)$ is exactly the minimizer of $KL(f_\theta, \hat{g}_n) = LD(\hat{g}_n, f_\theta) = \sum_t \log [\hat{g}_n(t)/f_\theta(t)] \hat{g}_n(t)$. This minimum distance interpretation of ML estimation provides the basis for developing the Hellinger deviance (HDV) tests (Simpson 1989) because it connects the classical likelihood ratio (LR) tests to the distance-based deviance tests (see Section 1.5). For convenience, only the $KL$ representation is used in discussions of the MD nature of ML estimation in the rest of this dissertation.

Equation (1.4) also shows that $KL(f_\theta, \hat{g}_n)$ and $2n^{-2} (f_\theta, \hat{g}_n)$ are asymptotically equivalent if $g$ is in fact some $f_\theta_0$. This allows us to establish the heuristic relationship between the GMHD estimator and the generalized maximum likelihood (GML) estimator under the assumed model for non-iid data.

The study of MHD estimation for the analysis of count data was begun by Simpson (1987), who established the consistency and asymptotic efficiency of the MHD estimator. The conditions are much less restrictive than that for continuous data in large part because the assumptions imposed on the density estimates are no longer necessary. (See Simpson 1987 for the details.)

An intuitive discussion based on the estimating equations of ML estimation and MHD estimation was carried out by Simpson (1987) to explain the robustness of the MHD estimator over the ML estimator. A review of this explanation is given below. Suppose that $\hat{g}_n(t)$ is defined by Equation (1.2). Then, the ML estimate maximizes $\sum_t \log [f_\theta(t)] \hat{g}_n(t)$ and the MHD estimate maximizes $\rho_{\theta,n} = \sum_t f_{\theta}^{1/2}(t) \hat{g}_n^{1/2}(t)$. Let $l_\theta(t)$ be the gradient of $\log f_\theta(t)$
with respect to \( \theta \). Then, ML estimation has the (standardized) estimating equation

\[
\sum_t \hat{g}_n(t)l'_\theta(t) = 0,
\]

and MHD estimation has the standardized estimating equation

\[
\rho_{\theta,n}^{-1} \sum_t \hat{g}_n^{1/2}(t) f_{\theta}^{1/2}(t) l'_\theta(t) = 0.
\]

One can see that, if the assumed model is true, Equations (1.5) and (1.6) agree in the limit as \( n \to \infty \), but they treat an outlier quite differently. In Equation (1.5), the expectation of \( l'_\theta(t) \) is with respect to \( \hat{g}_n \) whereas in Equation (1.6), the expectation is with respect to \( \rho_{\theta,n}^{-1} \hat{g}_n^{1/2} f_{\theta}^{1/2} \). An outlying observation is clearly downweighted in the latter case. If, in particular, \( f_{\theta} \) has finite Fisher information, then \( f_{\theta}^{1/2}(t) l'_\theta(t) \to 0 \) as \( t \to \infty \), implying that an improbable count has very little impact on the MHD estimator.

### 1.4 Minimum Penalized Hellinger Distance
Estimation for Discrete Models

In discrete parametric models, Harris and Basu (1994) studied the MHD estimator by expressing the Hellinger distance in the form of a penalized log-likelihood, in which the model itself is modified according to the data to make the correspondence exact. The penalty is the sum of the model probabilities over the \textit{empty cells}, the non-observed cells of the sample space. Adjustment of the amount of the penalty leads to the class of minimum penalized Hellinger distance (MPHD) estimators. Harris and Basu established that by suitably reducing the weight of the penalty, the MPHD estimation can perform substantially better than the ordinary MHD estimation in terms of small sample efficiency at the model, without sacrificing the asymptotic and the robustness properties of the latter. A brief review based on Harris and Basu’s study is given here.

Consider a parametric family of distributions \( \{f_{\theta} : \theta \in \Theta\} \) with countable support and let \( \hat{g}_n(t) \) be the proportion of sample observations having the value \( t \). Recall that the MHD
estimator of $\theta$ minimizes the distance

$$2u_n^2(f_\theta, \hat{g}_n) = 2 \sum_t (f_\theta^{1/2}(t) - \hat{g}_n^{1/2}(t))^2$$

over $\Theta$, and the ML estimator of $\theta$ minimizes the divergence

$$KL(f_\theta, \hat{g}_n) = \sum_t \log[\hat{g}_n(t)/f_\theta(t)]\hat{g}_n(t)$$

over $\Theta$.

Now, let $g_\theta$ be a data-modified model density such that the minimization of the Kullback-Leibler divergence between $g_\theta$ and $\hat{g}_n$ generates the MHD estimator of the parameter based on $f_\theta$. One can see that this can be achieved if for all $t$,

$$\log[\hat{g}_n(t)/g_\theta(t)]\hat{g}_n(t) = 2(f_\theta^{1/2}(t) - \hat{g}_n^{1/2}(t))^2.$$ 

Solving the above equation yields

$$g_\theta(t) = \hat{g}_n(t) \exp\{-2[1 - \sqrt{f_\theta(t)/\hat{g}_n(t)}]^2\}.$$ 

However, unlike the Hellinger distance, there is no contribution to the Kullback-Leibler divergence from the empty cells, the terms with $\hat{g}_n(t) = 0$. Thus, the real relation should be

\begin{equation}
\sum_t \log[\hat{g}_n(t)/g_\theta(t)]\hat{g}_n(t) = 2 \left[ \sum_{\hat{g}_n(t) \neq 0} (f_\theta^{1/2}(t) - \hat{g}_n^{1/2}(t))^2 \right],
\end{equation}

whereas

\begin{equation}
2u_n^2(f_\theta, \hat{g}_n) = 2 \sum_t (f_\theta^{1/2}(t) - \hat{g}_n^{1/2}(t))^2
\end{equation}

\begin{equation}
= 2 \left[ \sum_{\hat{g}_n(t) \neq 0} (f_\theta^{1/2}(t) - \hat{g}_n^{1/2}(t))^2 + \sum_{\hat{g}_n(t) = 0} f_\theta(t) \right].
\end{equation}

By comparing Equations (1.7) and (1.8), one can see that minimizing $2u_n^2(f_\theta, \hat{g}_n)$ corresponds to minimizing

$$\sum_t \log[\hat{g}_n(t)/g_\theta(t)]\hat{g}_n(t) + 2 \sum_{\hat{g}_n(t) = 0} f_\theta(t)$$

$$= KL(g_\theta, \hat{g}_n) + 2 \sum_{\hat{g}_n(t) = 0} f_\theta(t),$$
which can be thought of as a penalized log-likelihood with the term \( 2 \sum \hat{g}_n(t) = 0 \) \( f_\theta(t) \) being a penalty applied to the Kullback-Leibler divergence using the data-modified model density, \( g_\theta \). Note that the penalty applied here is unusual in that it depends on the data, which is not the case in other applications of penalty functions.

To adjust the amount of the penalty, one multiplies the penalty by a weight, \( h \). Using different values of \( h \) generates the class of MPHD estimators which minimize twice the squared penalized Hellinger distance defined as

\[
2u^2_{\text{ph},h}(f_\theta; \hat{g}_n) = 2 \left[ \sum_{\hat{g}_n(t) \neq 0} (f_\theta^{1/2}(t) - \hat{g}_n^{1/2}(t))^2 + h \sum_{\hat{g}_n(t) = 0} f_\theta(t) \right]
\]

\[
= 2 \left[ \sum_t (f_\theta^{1/2}(t) - \hat{g}_n^{1/2}(t))^2 - (1 - h) \sum_{\hat{g}_n(t) = 0} f_\theta(t) \right]
\]

\[
= 2 \left[ 2 - 2 \sum_t f_\theta^{1/2}(t) \hat{g}_n^{1/2}(t) - (1 - h) \left( 1 - \sum_{\hat{g}_n(t) \neq 0} f_\theta(t) \right) \right],
\]

or, equivalently, maximize

\[
2 \sum_t f_\theta^{1/2}(t) \hat{g}_n^{1/2}(t) - (1 - h) \sum_{\hat{g}_n(t) \neq 0} f_\theta(t).
\]

One can obtain the ordinary MHD estimator by setting \( h \) to 1 since \( u_{\text{ph},1}(f_\theta; \hat{g}_n) = u_{\text{ph}}(f_\theta; \hat{g}_n) \).

Lindsay (1994) observed small sample inefficiency of the MHD estimator compared to the ML estimator at the model. His results suggest that this relative poor performance of the MHD estimator in small samples may be due to the large weight that the Hellinger distance puts on the inliers, cells of the sample space with fewer data than expected under the model. Lindsay studied the asymptotic and robustness properties of a subclass of MD estimators called the minimum disparity estimators, including the MHD estimator, through a characterizing function \( A(\delta(t)) \), named the residual adjustment function, which determines how the procedure treats standardized residuals of the form \( \delta(t) = \hat{g}_n(t) / f_\theta(t) - 1, \delta(t) \in [-1, \infty) \). A large positive value of \( \delta(t) \) represents an outlying observation while a negative value of \( \delta(t) \) near -1 represents an inlier. An empty cell has \( \delta(t) = -1 \). Table 3 of Lindsay
(1994) shows that, while the MHD estimator is robust against the presence of outliers, inliers appear to cause larger biases in the MHD estimator compared to the ML estimator.

The idea of the MPHD estimation is to improve the finite sample efficiency of the MHD estimator by reducing the weight $h$ on empty cells. Two questions arise.

1. What value of $h$ leads to the optimal MPHD estimator (corresponding to the smallest mean square error [MSE]) in the class?

2. How are the asymptotic and robustness properties of the MHD estimator affected by the penalty process?

Harris and Basu (1994) proposed that the MPHD estimator with $h = 0.5$ may be the best estimator in the class because in this case the estimating equation of MPHD estimation puts the same weight on the empty cells as the estimating equation of ML estimation. To understand this, a discussion of Lindsay's residual adjustment function is necessary. The residual adjustment function of minimum disparity estimation is very similar to the $\psi$ function of M-estimation. Many density-based distances can have their estimating equations written in the form

$$\sum_t A(\delta(t)) \nabla f_\theta(t) = 0,$$

where $\nabla$ represents the gradient with respect to $\theta$. The residual adjustment function $A(\delta(t))$ may be properly standardized so that $A(0) = 0$ and $A'(0) = 1$ without changing its estimating properties. Here are some examples of interest. First, the estimating equation for ML estimation is of the form

$$-\nabla \sum_t \log[\hat{g}_n(t)/f_\theta(t)]\hat{g}_n(t) = \sum_t [\hat{g}_n(t)/f_\theta(t)]\nabla f_\theta(t) = 0,$$

which can be written as

$$\sum_t \left(\frac{\hat{g}_n(t)}{f_\theta(t)} - 1\right) \nabla f_\theta(t) = \sum_t \delta(t)\nabla f_\theta(t) = 0.$$
because \( \sum_t \nabla f_\theta(t) = 0 \) for all \( \theta \). Therefore, \( A(\delta) = \delta \) in this case. Similarly, the estimating equation for MHD estimation is

\[
- \nabla^2 \sum_t (f_\theta^{1/2}(t) - \hat{g}_{n/2}(t))^2 = \sum_t 2(\sqrt{\delta(t) + 1} - 1) \cdot \nabla f_\theta(t) = 0
\]

so that \( A(\delta) = 2(\sqrt{\delta + 1} - 1) \). The multiplier 2 plays the role of standardization of the function \( A(\delta) \), otherwise \( A'(0) \) would have been 1/2 rather than the desired 1; this is the reason why twice the squared Hellinger distance is considered rather than the squared Hellinger distance itself. One can see that outliers, which relate to large positive values of \( \delta \), are downweighted in the MHD estimation procedure according to its residual adjustment function \( A(\delta) \). Finally, the estimating equation for MPH\( n \) estimation is

\[
- \nabla^2 \left[ \sum_{\hat{g}_n(t) \neq 0} (f_\theta^{1/2}(t) - \hat{g}_{n/2}(t))^2 + h \sum_{\hat{g}_n(t) = 0} f_\theta(t) \right] = \sum_{\delta(t) \neq -1} 2(\sqrt{\delta(t) + 1} - 1) \cdot \nabla f_\theta(t) - 2h \sum_{\delta(t) = -1} \nabla f_\theta(t) = 0.
\]

Thus,

(1.11) \hspace{1cm} A(\delta) = \begin{cases} 2(\sqrt{\delta + 1} - 1) & \text{if } \delta \neq -1 \\ -2h & \text{if } \delta = -1. \end{cases}

For the Kullback-Leibler divergence, \( A(\delta) \) is a linear function of \( \delta \). The properties of the other minimum disparity estimators are often analyzed by how far their \( A(\delta) \) departs from linearity. See Lindsay (1994) for more details.

One can now observe the impact of the empty cells on the MHD estimation and the MPH\( n \) estimation by comparing the values of their adjustment functions to that of the ML estimation at \( \delta = -1 \). \( A(-1) \) equals 1 for Kullback-Leibler divergence, while it equals -2 for the ordinary Hellinger distance. According to Equation (1.11), the penalized Hellinger distance with \( h = 0.5 \) has \( A(-1) = -1 \) implying that the process treats the empty cells the same way as the ML estimation does.

Finally, consider the asymptotic and robustness properties of the MPH\( n \) estimators. Note that the impact of the inliers vanishes asymptotically because, in large samples, the \( t \) values with \( \hat{g}_n(t) = 0 \) are those which have very small probabilities under the model. The
members of the penalized Hellinger distances differ from the ordinary Hellinger distance only at empty cells; therefore, as the effect of the empty cells becomes asymptotically insignificant at the model, all the MPHD estimators inherit the asymptotic properties of the ordinary MHD estimator. On the subject of robustness, Lindsay showed that the robustness of the MHD estimator depends on the quantity $A'(0)$, called the *estimation curvature*. This quantity is not affected when altering the weight of the empty cells, so the MPHD estimators also share the robustness properties of the MHD estimator.

1.5 Hellinger Deviance Testing

As before, suppose that $Y = (Y_1, Y_2, \ldots, Y_n)^T$ is an iid sample from a distribution with true probability distribution function $g$, let $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$ be the model family of distributions, and suppose that it is desired to test the hypothesis $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta \setminus \Theta_0$, where $\Theta_0 \subset \Theta$. Then, the classical log-likelihood ratio test statistic $\lambda$ is given by

$$\lambda(\hat{\theta}_0; y) = 2[l(\hat{\theta}) - l(\hat{\theta}_0)],$$

which, asymptotically, has a $\chi^2_p$ distribution under the null model. Here, $l(\theta) = \sum_{i=1}^n \log f_\theta(y_i)$, $\hat{\theta}$ and $\hat{\theta}_0$ are the maximizers of $l(\theta)$ over $\Theta$ and $\Theta_0$, respectively, and $p$ is the number of independent constraints imposed by the null hypothesis. This commonly used likelihood ratio testing procedure yields uniformly most powerful tests in simple situations but it is not robust against small perturbations of the underlying model.

In Section 1.3, the log-likelihood based on data was shown to have a distance interpretation and the maximizer $\hat{\theta}_{\text{mle}}$ of $\frac{1}{n} l(\theta)$ was shown to be nearly (exactly for discrete models) the minimizer of the Kullback-Leibler divergence $KL(f_\theta, \hat{g}_n)$. These results imply that the difference of two log-likelihoods in the LR test statistic can be seen as the difference of two distances. To explore the idea further, first focus on the discrete models. The LR test statistic may now be reformulated as follows.

$$\lambda(\hat{\theta}_0; y) = 2[l(\hat{\theta}) - l(\hat{\theta}_0)]$$
\[= 2 \sum_{i=1}^{n} [\log f_{\hat{\theta}}(y_i) - \log f_{\theta_0}(y_i)]\]

\[= 2n \cdot \frac{1}{n} \sum_{i=1}^{n} [\log f_{\hat{\theta}}(y_i) - \log f_{\theta_0}(y_i)]\]

\[= 2n \sum_{i} [\log f_{\hat{\theta}}(t) - \log f_{\theta_0}(t)] \hat{g}_n(t)\]

\[= 2n \sum_{i} [\log (f_{\hat{\theta}}(t)/\hat{g}_n(t)) - \log (f_{\theta_0}(t)/\hat{g}_n(t))] \hat{g}_n(t)\]

\[= 2n \sum_{i} [\log (\hat{g}_n(t)/f_{\theta_0}(t)) - \log (\hat{g}_n(t)/f_{\hat{\theta}}(t))] \hat{g}_n(t)\]

\[= 2n[KL(f_{\theta_0}, \hat{g}_n) - KL(f_{\hat{\theta}}, \hat{g}_n)].\]

(1.12)

Therefore, to construct a robust analogue of the LR test, it seems reasonable to replace the likelihood-based Kullback-Leibler distance by a distance measure that yields more robust estimates, such as the Hellinger distance.

Simpson (1989) introduced the Hellinger deviance test—a MHD analogue of the LR test as a robust and efficient inference approach for iid data. As with the LR test, the idea of the HDV test is to measure how much further the data are from the null model than from the unconstrained model. But instead of measuring the distance with maximized log-likelihoods, the HDV test is based on minimized Hellinger distances. Using the fact that \(KL(f_{\theta}, \hat{g}_n)\) and \(2u_n^2(f_{\theta}, \hat{g}_n)\) are asymptotically equivalent at the model (see Section 1.3) and letting \(\rho(f_{\theta}, \hat{g}_n) = \sum_{i} f_{\theta}^{1/2}(t) \hat{g}_n^{1/2}(t)\), the Hellinger deviance test statistic \(d_n\) is defined as

\[d_n(\hat{\theta}_0; y) := 2n[2u_n^2(f_{\hat{\theta}_0}, \hat{g}_n) - 2u_n^2(f_{\hat{\theta}}, \hat{g}_n)]\]

\[= 2n[4\rho(f_{\hat{\theta}}, \hat{g}_n) - 4\rho(f_{\hat{\theta}_0}, \hat{g}_n)]\] by Equations (1.1) and (1.3)

\[= 8n[\rho(f_{\hat{\theta}}, \hat{g}_n) - \rho(f_{\hat{\theta}_0}, \hat{g}_n)],\]

where \(\hat{\theta}\) and \(\hat{\theta}_0\) are the maximizers of \(\rho(f_{\theta}, \hat{g}_n)\) over \(\Theta\) and \(\Theta_0\), respectively.

As with the LR test, the HDV test may be used to decide between two competitive, nested models. Note that the HDV test statistic shares the property of the LR test statistic that if \(\Theta_0 \subset \Theta_1 \subset \Theta\), the test of \(\Theta_0\) versus \(\Theta \setminus \Theta_0\) can be partitioned into a test of \(\Theta_0\) versus \(\Theta_1 \setminus \Theta_0\) and a test of \(\Theta_1 \setminus \Theta_0\) and a test of \(\Theta_1\) versus \(\Theta \setminus \Theta_1\). Therefore, suppose that the hypothesis of
interest is \( H_0 : \theta \in \Theta_0 \) versus \( H_1 : \theta \in \Theta_1 \setminus \Theta_0 \), where \( \Theta_0 \subset \Theta_1 \subset \Theta \). The LR test statistic is then the LR test statistic of the test \( \Theta_0 \) versus \( \Theta \setminus \Theta_0 \) minus the LR test statistic of the test \( \Theta_1 \) versus \( \Theta \setminus \Theta_1 \). It has the form

\[
\Delta \lambda(\hat{\theta}_0, \hat{\theta}_1; y) = \lambda(\hat{\theta}_0; y) - \lambda(\hat{\theta}_1; y) \\
= 2[l(\hat{\theta}_1) - l(\hat{\theta}_0)] \\
= 2n[KL(f_{\hat{\theta}_0}, \hat{g}_n) - KL(f_{\hat{\theta}_1}, \hat{g}_n)].
\]

This statistic is asymptotically \( \chi^2 \)-distributed with degrees of freedom \( q - p \), where \( p \) and \( q \) are the number of independent constraints under the null and the alternative hypotheses, respectively. Similarly the HDV test takes the form

\[
\Delta d_{ni}(\hat{\theta}_0, \hat{\theta}_1; y) = d_{ni}(\hat{\theta}_0; y) - d_{ni}(\hat{\theta}_1; y) \\
= 2n[2\sigma_{ni}^2(f_{\hat{\theta}_0}, \hat{g}_n) - 2\sigma_{ni}^2(f_{\hat{\theta}_1}, \hat{g}_n)] \\
= 8n[\rho(f_{\hat{\theta}_1}, \hat{g}_n) - \rho(f_{\hat{\theta}_0}, \hat{g}_n)].
\]

In the last two test statistics, \( \hat{\theta}_0 \) and \( \hat{\theta}_1 \) are maximizers over \( \Theta_0 \) and \( \Theta_1 \), respectively, of \( l(\theta) \) for the LR test and of \( \rho(f_{\theta}, \hat{g}_n) \) for the HDV test.

In continuous models, the test statistics for HDV tests are similarly defined. The main differences are that \( \hat{g}_n \) is a nonparametric density estimate, such as a kernel density estimate, and that \( \rho(f_{\theta}, \hat{g}_n) = \int f_{\theta}^{1/2}(t) \hat{g}_n^{1/2}(t)dt \).

Simpson (1989) showed that the HDV test is asymptotically equivalent to the LR test if the parametric model is correct. This implies that the HDV test statistic should have the same asymptotic \( \chi^2 \) distribution as the LR test statistic under any null model. Using breakdown analysis, Simpson (1989) also demonstrated that HDV tests are much more robust than LR tests. Part of this dissertation focuses on developing a testing procedure for GMHD estimation (Section 2.1) called the generalized Hellinger deviance test (Section 2.4) which extends Simpson’s HDV tests to GLMs.
1.6 Penalized Hellinger Deviance Testing for Discrete Models

Despite the asymptotic equivalence of the HDV test and the LR test at the null model, Simpson (1989) noted that the convergence of the HDV test statistic to the appropriate limiting $\chi^2$ distribution in some discrete models, such as the Poisson model, is quite slow and, therefore, it requires a very large sample size for the $\chi^2$ approximation to be useful. Basu et al. (1996) applied the penalized Hellinger distance idea (Harris and Basu, 1994) in constructing robust analogues of the LR test and generated the penalized Hellinger deviance (PHDV) tests, which have been shown to have much faster convergence rates to the LR test than the ordinary HDV test if the penalty weight $h$ is appropriately chosen.

Using the notation from Section 1.5, consider the hypothesis $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta \setminus \Theta_0$, $\Theta_0 \subset \Theta$. The penalized Hellinger deviance test statistic $d_{PH,h}$ is defined by

$$d_{PH,h}(\hat{\theta}_0; y) := 2n\left[2u_{PH,h}^2(f_{\hat{\theta}_0}, \hat{g}_n) - 2u_{PH,h}^2(f_{\tilde{\theta}_0}, \tilde{g}_n)\right]$$

$$= 8n\left[\rho(f_{\tilde{\theta}_0}, \tilde{g}_n) - \rho(f_{\hat{\theta}_0}, \hat{g}_n)\right] - 4n(1 - h) \sum_{\hat{g}_n(t) \neq 0} f_{\hat{\theta}_0}(t) - \sum_{\tilde{g}_n(t) \neq 0} f_{\tilde{\theta}_0}(t) \right]$$

by Equation (1.9),

where $\hat{\theta}$ and $\tilde{\theta}_0$ are the MPHD estimators of $\theta$ under the null and the unconstrained models, respectively.

For testing two nested models $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1 \setminus \Theta_0$, $\Theta_0 \subset \Theta_1 \subset \Theta$, the PHDV test statistic has the form

$$\Delta d_{PH,h}(\hat{\theta}_0, \tilde{\theta}_1; y) = d_{PH,h}(\hat{\theta}_0; y) - d_{PH,h}(\tilde{\theta}_1; y)$$

$$= 2n\left[2u_{PH,h}^2(f_{\tilde{\theta}_0}, \tilde{g}_n) - 2u_{PH,h}^2(f_{\hat{\theta}_1}, \hat{g}_n)\right]$$

$$= 8n\left[\rho(f_{\hat{\theta}_1}, \hat{g}_n) - \rho(f_{\tilde{\theta}_0}, \tilde{g}_n)\right] - 4n(1 - h) \sum_{\hat{g}_n(t) \neq 0} f_{\hat{\theta}_1}(t) - \sum_{\tilde{g}_n(t) \neq 0} f_{\tilde{\theta}_0}(t)$$,

where $\hat{\theta}_0$ and $\tilde{\theta}_1$ are the MPHD estimators under the null and the alternative models, respectively.

Because the empty cells have no effect asymptotically on the MPHD estimation, the above PHDV test statistics have the same asymptotic distributions as the ordinary HDV
test statistics. The results of Basu et al. (1996) show that the PHDV test using a suitable penalty weight enjoys the same robustness properties as the HDV test, but is often closer to the LR test at the model in small samples. In their empirical study, the PHDV tests using \( h = 0.5 \) provide significantly more accurate levels for finite samples when using \( \chi^2 \) critical values.

1.7 Generalized Linear Models

Generalized linear models extend the idea of a linear model and may be defined as comprising three components:

1. A probability distribution for each observation belonging to a linear exponential class, with density or distribution functions of the form

\[
    f(y_i; \theta(\mu_i), \varphi) = \exp\{[y_i \theta(\mu_i) - b(\theta(\mu_i))]/a_i(\varphi) + c(y_i; \varphi)\}
\]

(1.13)

\[
    = \exp\{[y_i \theta_i - b(\theta_i)]/a_i(\varphi) + c(y_i; \varphi)\} \quad \text{for simplicity.}
\]

The \( \mu_i \)'s are the means of the \( Y_i \)'s, \( i = 1, 2, \ldots m; \varphi \) is a scale parameter which is known for some families and not for others. Each \( a_i(\cdot) \) is a function of \( \varphi \) and is commonly of the form \( a_i(\varphi) = \varphi/v_i \), where \( v_i \) are known weights. An example of unequal weights is a normal model in which each observation is the mean of \( n_i \) independent readings. In this case \( a_i(\varphi) = \varphi/n_i = \sigma^2/n_i \) so that \( v_i = n_i \). Families of distributions written in the above linear exponential form are said to be in the canonical form with canonical parameter \( \theta \).

2. A linear predictor, \( \eta_i \). For each observation, let \( x_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip-1})^T \) be a vector of values on explanatory variables and let \( \beta = (\beta_0, \beta_1, \ldots, \beta_{p-1})^T \). Then,

\[
    \eta_i = \beta_0 + \sum_{j=1}^{p-1} x_{ij} \beta_j.
\]

3. A known, monotonic, differentiable link function \( \ell \) which relates the \( \eta_i \)'s to the \( \mu_i \)'s. Specifically

\[
    \ell(\mu_i) = \eta_i.
\]
The canonical link function, defined by \( \theta(\mu_i) = \eta_i \), helps simplify the computation of the estimates in most cases. Generally, one chooses a link function which maps the range of \( \mu \) onto the whole real line so that there are no restrictions on the values of \( \beta_j \)'s. Not all canonical link functions satisfy this restriction.

The idea of the GLMs is to specify the set of means \( \mu = (\mu_1, \mu_2, \ldots, \mu_m)^T \) of the response variable in terms of a smaller set of parameters \( \beta \) based on a set of linear combinations of the explanatory variables through a known link function \( \ell \). The class of GLMs was first introduced by Nelder and Wedderburn (1972). It unifies several traditional models involving linear combinations of parameters, such as linear regression models (\( \ell(\mu_i) = \mu_i \) and the \( y_i \)'s normal), log-linear models (\( \ell(\mu_i) = \log \mu_i \) and the \( y_i \)'s Poisson) for the analysis of count data, and three models for which the \( y_i \)'s are binomial with probability of success \( \pi_i \): probit models (\( \ell(\pi_i) = \Phi^{-1}(\pi_i) \), where \( \Phi \) is the standard Normal cumulative distribution function), logit models (\( \ell(\pi_i) = \log(\frac{\pi_i}{1-\pi_i}) \)), and complementary log-log models (\( \ell(\pi_i) = \log(-\log(1-\pi_i)) \)) for the analysis of discrete proportion data.

For GLMs, Nelder and Wedderburn (1972) proposed quasi-likelihood (QL) estimation which entails estimating \( \beta \) by optimizing the likelihood only with respect to the \( \beta_j \)'s, and treating \( \varphi \) as a nuisance parameter. Nelder and Wedderburn (1972) showed that the QL estimation in all GLMs could be carried out using a common iteratively reweighted least squares (IRWLS) algorithm, with weights and adjusted response variates depending on the iteratively upgraded estimates of \( \beta \).

In situations in which there are multiple observations \( y_{i,j} \)'s per combination of values of the explanatory variables (a cell or subpopulation), the QL estimates obtained by the IRWLS algorithm may be shown to be functions of the data only through the subpopulation means, simplifying the calculations (see Section 3.2.1 for the details).

Nelder and Wedderburn (1972) developed tests for GLMs based on QL estimation, which are called generalized likelihood ratio (GLR) tests throughout this dissertation. Nelder and Wedderburn (1972) defined a likelihood-based deviance function, which is useful as a mea-
sure of goodness-of-fit and also for comparing two nested models (described in Section 2.4).

Neither QL estimation nor GLR tests are robust procedures. This dissertation generalizes the MHD and related estimation and testing techniques to GLMs, providing robust alternatives that are also efficient.

1.8 Generalized Maximum Likelihood Estimation

Lehmann (1983, §6.6) briefly mentioned the extension of the classical ML estimation to the more general case with data coming from two or more populations. This generalized procedure is referred to as generalized maximum likelihood (GML) estimation throughout this dissertation.

For iid data $Y_1, Y_2, \ldots, Y_n$ from a distribution with probability distribution function $f_\theta$, Lehmann (1983, §6.4, Theorem 4.1) states that, under regularity conditions, there exists (with probability tending to 1 as $n \to \infty$) a consistent sequence of roots $(\hat{\theta}_n)_{n=1}^{\infty}$ of the likelihood equations satisfying

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, I_1(\theta_0)^{-1}),$$

where $\theta_0$ are the true parameters and $I_1(\theta_0)$ is the Fisher information matrix of $\theta$ evaluated at $\theta_0$. Note that $I_1(\theta) = -\mathbb{E} \left[ \frac{\partial^2 \log f_\theta(Y)}{\partial \theta \partial \theta^T} \right]$ describes the amount of information contained in a single observation and that $I(\theta) = n \times I_1(\theta)$ represents the total amount of information contained in $n$ iid observations.

Lehmann (1983, §6.6) then describes how this efficient likelihood estimation may be generalized to the case of two or more samples. Suppose that we have $m$ independent samples from $m$ different populations, labelled $i = 1, 2, \ldots, m$, such that each sample $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^T$ of size $n_i$ is iid with probability distribution function $f_{i,\theta}$, where the parameters $\theta$ of dimension $p$ are assumed to be the same for all populations. The limit situation we shall consider supposes that each $n_i \to \infty$ all at the same rate, and that $m$ remains fixed. Let $n = \sum_{i=1}^{m} n_i$ be the total sample size. We, then, consider sequences of sample sizes $\{(n_{i_1}, n_{i_2}, \ldots, n_{i_m})\}_{\nu=1}^{\infty}$ with total sample size $n_{\nu} = \sum_{i=1}^{m} n_{i_\nu}$ such that the
proportions

\[ \frac{n_{iv}}{n_v} \to \lambda_i, \text{ as } \nu \to \infty, \]

where \( \sum_{i=1}^{m} \lambda_i = 1 \) and the \( \lambda_i \)'s are all greater than 0.

Now recall from Lehmann (1983, §6.4, Theorem 4.1) that a central concept in the i.i.d. case is the amount of information \( I_1(\theta) \) contained in a single observation. To see how the generalization works, an analogous quantity is defined for the multiple population case. First, define \( I_1^{(i)}(\theta) = -\mathcal{E} \left[ \frac{\partial^2 \log f_{i,\theta}(Y_{ij})}{\partial \theta \partial \theta^t} \right] \) to be the information provided by a single observation from \( f_{i,\theta} \). Then, \( I^{(i)}(\theta) = n_i I_1^{(i)}(\theta) \) is the information contained in a random sample of size \( n_i \) from \( f_{i,\theta} \). Assuming the information contained in independent observations to be additive, the total amount of information contained in the \( n = \sum_{i=1}^{m} n_i \) independent observations is

\[ I(\theta) = \sum_{i=1}^{m} n_i I_1^{(i)}(\theta), \]

and the average information per observation is therefore

\[ \frac{1}{n} I(\theta) = \sum_{i=1}^{m} \frac{n_i}{n} I_1^{(i)}(\theta), \]

which tends to

\[ I_1(\theta) = \sum_{i=1}^{m} \lambda_i I_1^{(i)}(\theta) \text{ as } n \to \infty. \]

It will later be seen that this asymptotic average amount of information plays a central role in the non-i.i.d. case (Lehmann 1983, §6.6, Theorem 6.1). Next, let

\[ l^{(i)}(\theta) = \sum_{j=1}^{n_i} \log f_{i,\theta}(y_{ij}) \]

denote the log-likelihood of \( \theta \) based on the \( i^{th} \) sample and

\[ l(\theta) = \sum_{i=1}^{m} l^{(i)}(\theta) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \log f_{i,\theta}(y_{ij}) \]

be the log-likelihood based on all \( n \) observations. Then, the log-likelihood equations are given by

\[ l'_k(\theta) = \sum_{i=1}^{m} l'_k^{(i)}(\theta) = 0, \text{ for } k = 1, 2, \ldots, p, \]
where \( l_k^{(i)}(\theta) = \sum_{j=1}^{n_i} \frac{\partial}{\partial \theta_k} \log f_i,\theta(y_{ij}). \)

Finally, in generalization of Theorem 4.1 (Lehmann 1983, §6.4), Theorem 6.1 (Lehmann 1983, §6.6) indicates that if the regularity conditions of Theorem 4.1 hold for each \( f_i,\theta \), then the conclusions of Theorem 4.1 remain valid. That is, we have

**Theorem 1.8.1** (Lehmann 1983, §6.6, Theorem 6.1) *For each \( i = 1, 2, \ldots m \), let \( Y_{i1}, Y_{i2}, \ldots, Y_{in_i} \) be iid with probability distribution function \( f_i,\theta \) satisfying the assumptions of Theorem 4.1 (Lehmann 1983, §6.4), and suppose that all \( n = \sum_{i=1}^{m} n_i \) observations are independent. Let \( \{(n_{1\nu}, n_{2\nu}, \ldots, n_{m\nu})\}_{\nu=1}^{\infty} \) be a sequence of sample sizes with \( n_{\nu} = \sum_{i=1}^{m} n_{i\nu} \) satisfying the limit condition (1.14). Then, with probability tending to 1 as \( n \to \infty \), there exist a consistent sequence of roots \( \{\hat{\theta}_n\}_{n=1}^{\infty} \) of the log-likelihood equations (Equation (1.16)) satisfying

\[ \sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \xrightarrow{D} N_p \left( 0, [I_1(\theta_0)]^{-1} \right), \]

where \( \theta_0 \) are the true parameters and \( I_1(\theta) \) is given by Equation (1.15).

A detailed proof of Theorem 1.8.1 is given in Section 5.1.
CHAPTER 2

GENERALIZED MINIMUM PENALIZED HELLINGER DISTANCE ESTIMATION
AND GENERALIZED PENALIZED HELLINGER DEVIANCE TESTING
FOR DISCRETE GENERALIZED LINEAR MODELS

2.1 Generalized Minimum Hellinger Distance
Estimation and Its Application in GLMs

The main aim of this dissertation is to generalize MHD estimation and related methods
to GLMs. This section contains the motivation for and definition of generalized minimum
Hellinger distance estimation for non-iid data and the idea of how this procedure may be
applied to GLMs.

Suppose that we have m independent random samples from m distinct populations.
For each \( i = 1, 2, \ldots, m \), let \( Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^T \) be a random sample with probability
distribution function \( g_i \), and let \( \mathcal{F}_i = \{f_{i, \theta} : \theta \in \Theta\} \) be a specified parametric family of
distributions. Assume that all m families depend on the same parameter \( \theta \) (of dimension
\( p \)) and that each \( g_i \) is either equal to or at least close to some member \( f_{i, \theta_0} \) in \( \mathcal{F}_i \).
Furthermore, following Lehmann's extension for GML estimation (Section 1.8), we assume the
limit condition that \( m \) is fixed, each \( n_i \to \infty \), and the proportions \( \frac{n_i}{n} \to \lambda_i \) as the total
sample size \( n = \sum_{i=1}^{m} n_i \to \infty \).

Following Beran's argument for MHD estimator in Section 1.3, the GMHD estimator is
heuristically related to the GML estimator if each \( g_i \) is in fact some \( f_{i, \theta_0} \) (but possibly not
otherwise!). Assume that \( g_i = f_{i, \theta_0} \in \mathcal{F}_i \) for all \( i \) and each \( n_i \) is sufficiently large. Then,
the GML estimator of \( \theta \) should be close to \( \theta_0 \) and each \( \hat{g}_{n_i} \) should be close to \( f_{i, \theta_0} = g_i \),
where \( \hat{g}_{n_i} \) is the nonparametric estimate of \( g_i \) based on the \( i^{th} \) iid random sample. Finding
the GML estimator of $\theta$ amounts to maximizing the joint log-likelihood

$$
\sum_{i=1}^{m} \sum_{j=1}^{n_i} \log f_{i,\theta}(y_{ij}) = \sum_{i=1}^{m} \frac{1}{n_i} \sum_{j=1}^{n_i} \log f_{i,\theta}(y_{ij})
$$

$$
= \sum_{i=1}^{m} n_i \int \log f_{i,\theta}(t)d\hat{G}_{n_i}(t) \quad \text{over } \theta \in \Theta,
$$

where $\hat{G}_{n_i}$ is the empirical cdf of the $i^{th}$ population. This procedure should be nearly the same as maximizing the quantity

$$
\sum_{i=1}^{m} n_i \int \log[f_{i,\theta}(t)/\hat{g}_{n_i}(t)]\hat{g}_{n_i}(t)dt
$$

$$
= - \sum_{i=1}^{m} n_i KL(f_{i,\theta}, \hat{g}_{n_i})
$$

$$
\approx - \sum_{i=1}^{m} n_i \cdot 2u_{h}^2(f_{i,\theta}, \hat{g}_{n_i}) \quad \text{over } \theta \text{ near } \theta_0,
$$

where $\cong$ holds because it has been shown in Section 1.3 that, under the assumed family of models, $KL(f_{i,\theta}, \hat{g}_{n_i})$ is asymptotically equivalent to $2u_{h}^2(f_{i,\theta}, \hat{g}_{n_i})$. That is, finding the GML estimator $\hat{\theta}_{mle}$ of $\theta$ by maximizing the joint log-likelihood over $\theta \in \Theta$ is nearly (exactly for discrete models) equivalent to minimizing

$$
\sum_{i=1}^{m} n_i KL(f_{i,\theta}, \hat{g}_{n_i}) \quad \text{over } \theta \text{ near } \theta_0
$$

and, if $g_i \in F_i$ for all $i$, is asymptotically equivalent to minimizing

$$
2 \sum_{i=1}^{m} n_i u_{h}^2(f_{i,\theta}, \hat{g}_{n_i}) \quad \text{over } \theta \text{ near } \theta_0.
$$

Thus, we may define the GMHD estimator $\hat{\theta}_H$ of $\theta_0$ based on $m$ random samples $Y_1, Y_2, \ldots, Y_m$ by

$$
\hat{\theta}_H := \arg \min_{\theta \in \Theta} \sum_{i=1}^{m} n_i u_{h}^2(f_{i,\theta}, \hat{g}_{n_i})
$$

$$
\quad = \arg \min_{\theta \in \Theta} \sum_{i=1}^{m} n_i \|f_{i,\theta}^{1/2} - \hat{g}_{n_i}^{1/2}\|_2^2,
$$

and claim that, under the model, the GMHD estimator is asymptotically close enough to the GML estimator to share its properties (Theorem 1.8.1). Also, in generalizing MHD
estimation, one can expect that the robustness properties of the MHD estimator should carry over to the GMHD estimator. Section 5.2 extends the theoretical results of Beran’s (1977) Theorem 1, which establishes the existence, uniqueness, and consistency of the MHD estimator, and Simpson’s (1987) Theorem 2, which establishes the asymptotic normality and efficiency of the MHD estimator for count data, to the GMHD estimator for count data.

It should be mentioned that the application of the GMHD estimation procedure for continuous data is currently limited to the case requiring multiple observations per population because this procedure involves continuous nonparametric density estimation (which entails estimating a scale parameter and bandwidths). Fortunately, there is no such limitation for discrete data which is the case of interest in this dissertation. According to Simpson (1987), for count data one can simply take \( \hat{g}_{n_i} \) to be the empirical density function

\[
(2.2) \quad \hat{g}_{n_i}(t) = \frac{N_{i,t}}{n_i}, \quad t = 0, 1, \ldots,
\]

where \( N_{i,t} \) is the frequency of \( t \) observed among the \( i^{th} \) random sample. Note that if sample \( i \) has only one observation \( y_i \), then by Equation (2.2) \( \hat{g}_{n_i} \) puts probability one on the point \( y_i \); i.e., \( \hat{g}_{n_i}(y_i) = 1 \).

The GMHD estimation procedure applies directly to GLMs because each distinct vector of explanatory variables in a GLM determines a subpopulation and the parameter vector \( \beta \) is the same for all the subpopulations. Consider first the GLMs with multiple observations per subpopulation. The GMHD estimation procedure can be formulated as follows. Suppose that \( g_i \) is the unknown probability distribution function of subpopulation \( i \) determined by the \( i^{th} \) vector of explanatory variables \( x_i^T \), and \( \mathcal{F}_i = \{ f_{\mu_i(b)} : \beta \in B \} \) is the assumed underlying family of distributions for \( g_i, i = 1, 2, \ldots, m \). (Recall that \( \mu_i \) depends on \( \beta \) through a link function \( \ell \).) Then, according to Equation (2.1), the GMHD estimator \( \hat{\beta}_H \) of \( \beta \) is given by

\[
\hat{\beta}_H = \arg\min_{b \in B} \sum_{i=1}^{m} n_i \| f_{\mu_i(b)}^{1/2}(t) - \hat{g}_{n_i}^{1/2}(t) \|_2^2
\]
\[ (2.3) \quad \mathbf{\hat{\beta}}_m = \max_{\beta \in \mathcal{B}} \sum_{i=1}^{m} n_i \int f^{1/2}_{\mu_i(b)}(t) \hat{g}_{n_i}^{1/2}(t) dt, \]

where \( \hat{g}_{n_i} \) is an appropriate nonparametric density estimate of \( g_i \) based on the random sample \( Y_i \) from the \( i^{th} \) subpopulation.

There are cases for which the scale parameter \( \varphi \) in Equation (1.13) is not known and must be estimated from the data. To do that, one can simply estimate \( \varphi \) along with \( \beta \) through the GMHD estimation procedure by viewing each assumed model family as \( \mathcal{F}_i = \{ f_{\mu_i(\beta), \varphi} : \beta \in \mathcal{B}, \varphi \in \Phi \} \). That is, obtaining the estimates of \( \beta \) and \( \varphi \) according to

\[ (\mathbf{\hat{\beta}}_m^T, \hat{\varphi}_m)^T = \max_{\beta \in \mathcal{B}, \varphi \in \Phi} \sum_{i=1}^{m} n_i \int f^{1/2}_{\mu_i(b), \varphi}(t) \hat{g}_{n_i}^{1/2}(t) dt. \]

Note that all three discrete GLMs—those based on Poisson, binomial, and negative binomial distributions—have \( \varphi \) equal to 1, so no estimation for \( \varphi \) is necessary.

The methodology described above is applicable to discrete GLMs with one or multiple observations per subpopulation using the empirical density estimate defined in Equation (2.2). It may be generalized in a straightforward manner to continuous GLMs with multiple observations per subpopulation, but with one observation per subpopulation it is not clear how to choose an appropriate density estimate. The study of this dissertation is focused on the inference for discrete GLMs.

### 2.2 Generalized Minimum Hellinger Distance Estimation for Discrete GLMs

For illustrative purposes, consider a GLM for count data. According to Equations (2.2) and (2.3), the GMHD estimator \( \mathbf{\hat{\beta}}_m \) of \( \beta \) in a discrete GLM is given by

\[ \mathbf{\hat{\beta}}_m = \max_{\beta \in \mathcal{B}} \sum_{i=1}^{m} n_i \sum_{t=0}^{\infty} \left( \frac{N_{\mu_i(b)}}{n_i} \right)^{1/2} f^{1/2}_{\mu_i(b)}(t) \]

\[ = \max_{\beta \in \mathcal{B}} \sum_{i=1}^{m} n_i \sum_{\hat{g}_{n_i}(t) \neq 0} \left( \frac{N_{\mu_i(b)}}{n_i} \right)^{1/2} f^{1/2}_{\mu_i(b)}(t), \]

which, for the case with only one observation \( y_i \) per cell, can be reduced to the form of

\[ \mathbf{\hat{\beta}}_m = \max_{\beta \in \mathcal{B}} \sum_{i=1}^{m} f^{1/2}_{\mu_i(b)}(y_i), \]
Simplifying the computation.

Simpson (1987) explained why the MHD estimators should be more robust than the ML estimators by observing their estimating equations. His argument may be generalized to compare the robustness between the GML estimator and the GMHD estimator. (Note that when $\varphi = 1$ such as in Poisson and binomial GLMs, the commonly used QL estimation is a version of GML estimation.) Let $\hat{g}_{ni}(t)$ be defined by Equation (2.2). Then, in discrete GLMs, the GML estimate maximizes $\sum_{i=1}^{m} n_i \sum_{t=0}^{\infty} \log[f_{\mu_i(\beta)}(t)] \hat{g}_{ni}(t)$ and the GMHD estimate maximizes $\rho_{\beta,n}^{-1} = \sum_{i=1}^{m} n_i \sum_{t=0}^{\infty} f_{\mu_i(\beta)}^{1/2}(t) \hat{g}_{ni}^{1/2}(t)$, where $n = (n_1, n_2, \ldots, n_m)$. Denote the gradient of $\log[f_{\mu_i(\beta)}(t)]$ with respect to $\beta$ by $l_{\mu_i(\beta)}'(t)$. Then, the standardized estimating equation of GML estimation is

\begin{equation}
(2.6) \quad n^{-1} \sum_{i=1}^{m} n_i \sum_{t=0}^{\infty} \hat{g}_{ni}(t) l_{\mu_i(\beta)}'(t) = 0,
\end{equation}

and the standardized estimating equation of GMHD estimation is

\begin{equation}
(2.7) \quad \rho_{\beta,n}^{-1} \sum_{i=1}^{m} n_i \sum_{t=0}^{\infty} \hat{g}_{ni}^{1/2}(t) f_{\mu_i(\beta)}^{1/2}(t) l_{\mu_i(\beta)}'(t) = 0.
\end{equation}

Provided that each $g_i$ is truly a member of $\mathcal{F}_i$, Equations (2.6) and (2.7) agree in the limit as each $n_i \to \infty$ at the same rate, but one can clearly see that a deviation from the assumed model or an outlier has much less impact on the GMHD estimator than on the GML estimator because the expectation of $l_{\mu_i(\beta)}'(t)$ is with respect to $\rho_{\beta,n}^{-1} n_i \hat{g}_{ni}^{1/2} f_{\mu_i(\beta)}^{1/2}$ for GMHD estimation rather than $n^{-1} n_i \hat{g}_{ni}$ for GML estimation. That is, the effect of an improbably data point is downweighted in GMHD estimation, but not in GML estimation. In an example for which each $f_{\mu_i(\beta)}$ has finite Fisher information, each $f_{\mu_i(\beta)}^{1/2}(t) l_{\mu_i(\beta)}'(t) \to 0$ as $t \to \infty$, showing the superior robustness of GMHD estimation compared to GML estimation.

2.3 Generalized Minimum Penalized Hellinger Distance Estimation for Discrete GLMs

Most of the GLMs do not have very many observations per vector of explanatory variables, which often leads to small total sample sizes. One should expect, as for MHD estimation, that the estimates may be very biased when implementing GMHD estimation in
small samples. As described in Section 1.4, Harris and Basu (1994) proposed the MPHD estimation to improve the small sample performance of the ordinary MHD estimation. Their penalty idea may be extended straightforwardly to GMHD estimation. According to Equation (2.1), in the general non-iid cases, the class of the GMPHD estimators $\hat{\theta}_{PH,h}$ of $\theta_0$ may be defined by

$$\hat{\theta}_{PH,h} := \arg \min_{\theta \in \Theta} \sum_{i=1}^{m} n_i u_{PH,h}^2(f_i, \theta, \hat{g}_{n_i})$$

$$= \arg \min_{\theta \in \Theta} \sum_{i=1}^{m} n_i \left[ \sum_{\hat{g}_{n_i}(t) \neq 0} (f_{i, \theta}^{1/2}(t) - \hat{g}_{n_i}^{1/2}(t))^2 + h \sum_{\hat{g}_{n_i}(t) = 0} f_{i, \theta}(t) \right]$$

$$= \arg \max_{\theta \in \Theta} \sum_{i=1}^{m} n_i \left[ 2 \sum_{t} f_{i, \theta}^{1/2}(t) \hat{g}_{n_i}^{1/2}(t) - (1 - h) \sum_{\hat{g}_{n_i}(t) \neq 0} f_{i, \theta}(t) \right] \text{ by Equation (1.10)}$$

$$= \arg \max_{\theta \in \Theta} \sum_{i=1}^{m} n_i \left[ 2 \sum_{\hat{g}_{n_i}(t) \neq 0} f_{i, \theta}^{1/2}(t) \hat{g}_{n_i}^{1/2}(t) - (1 - h) \sum_{\hat{g}_{n_i}(t) \neq 0} f_{i, \theta}(t) \right].$$

Therefore, in GLMs for count data, the GMPHD estimator $\hat{\beta}_{PH,h}$ of $\beta$ is given by

$$\hat{\beta}_{PH,h} = \arg \max_{b \in B} \sum_{i=1}^{m} \left[ 2 \sum_{\hat{g}_{n_i}(t) \neq 0} \left( \frac{N_{it}}{n_i} \right)^{1/2} f_{\hat{\mu}_i(b)}^{1/2}(t) - (1 - h) \sum_{\hat{g}_{n_i}(t) \neq 0} f_{\hat{\mu}_i(b)}(t) \right].$$

For the case with one observation $y_i$ per subpopulation, Equation (2.8) reduces to the form of

$$\hat{\beta}_{PH,h} = \arg \max_{b \in B} \sum_{i=1}^{m} \left[ 2 f_{\hat{\mu}_i(b)}^{1/2}(y_i) - (1 - h) f_{\hat{\mu}_i(b)}(y_i) \right],$$

which is much easier to compute. Setting $h$ equal to 1, one obtains the ordinary GMHD estimator.

2.3.1 Examples

In this dissertation, Poisson GLMs and binomial GLMs are the discrete GLMs for which the GMPHD estimation procedure has been implemented.

(a) Consider a Poisson GLM where each independent random sample $Y_i$, corresponding to the $i^{th}$ vector of explanatory variables $x_i^p$, is assumed to be iid from a Poisson
distribution $\text{Poi}(\mu_i)$ with mean parameter $\mu_i$. Suppose that the log link function is used; i.e., $\eta_i = \ell(\mu_i) = \log \mu_i$. Then, according to Equation (2.8), the GMPHD estimator $\hat{\beta}_{\nu,h}$ of $\beta$ can be obtained by maximizing

$$
\sum_{i=1}^m n_i \left[ 2 \sum_{\hat{\gamma}_n(t) \neq 0} \left( \frac{N_{it}}{n_i} \right)^{1/2} \exp \left\{ \frac{1}{2} \left[ t \hat{\eta}_i - \exp(\hat{\eta}_i) \right] \right\} - (1 - h) \sum_{\hat{\gamma}_n(t) \neq 0} \exp \left( \frac{t \hat{\eta}_i - \exp(\hat{\eta}_i)}{t!} \right) \right],
$$

where $\hat{\eta}_i = \sum_{j=0}^{p-1} x_{ij} b_j$.

(b) Consider a binomial GLM for which each iid sample $Y_i$, related to $x_i^T$, is assumed to come from a binomial distribution $\mathcal{B}(N_i, \pi_i)$ with known $N_i$'s. Here, $N_i$ is the number of independent bernoulli trials and $\pi_i$ is the probability of successes, so $\mu_i = N_i \pi_i$ is the mean parameter. Suppose that a logit link function with respect to $\pi$ is used; i.e., $\eta_i = \ell(\mu_i) = \log \left( \frac{\mu_i}{N_i - \mu_i} \right) = \log(\frac{\pi_i}{1 - \pi_i})$. Then, by Equation (2.8), the GMPHD estimator $\hat{\beta}_{\nu,h}$ of $\beta$ maximizes

$$
\sum_{i=1}^m n_i \left[ 2 \sum_{\hat{\gamma}_n(t) \neq 0} \left( \frac{N_{it}}{n_i} \right)^{1/2} \left[ \frac{N_i!}{t!(N_i - t)!} \right]^{1/2} \left\{ \frac{\exp(t \hat{\eta}_i)}{[1 + \exp(\hat{\eta}_i)]^N_i} \right\}^{1/2} \right] - (1 - h) \sum_{\hat{\gamma}_n(t) \neq 0} \left[ \frac{N_i!}{t!(N_i - t)!} \right] \left\{ \frac{\exp(t \hat{\eta}_i)}{[1 + \exp(\hat{\eta}_i)]^N_i} \right\}^{1/2}.
$$

2.3.2 The Choice of the Penalty Factor

The choice of the penalty weight, $h$, turns out to be crucial in determining the efficiency of GMPHD estimation and, in particular, whether GMPHD estimation is competitive with QL estimation in GLMs when the data are not contaminated. Computer simulations described in detail in Section 4.2.2 show that for small sample sizes the “optimal” value $h = 0.5$, suggested by Harris and Basu (1994), yields GMPHD estimates that have substantially higher MSE’s than QL estimates, and than GMPHD estimates obtained by using an optimal value of $h$ as determined by simulation (see Figure 4.5). Further simulations, summarized in Figures 4.1 and 4.2, show how the optimal value of $h$ (in terms of mean squared error) depends on the sample sizes. For the smaller sample sizes it is clear that the
optimal value of \( h \) may be substantially different from 0.5, even for MPHD estimation for an iid sample from a single Poisson population (Figure 4.1). The situation is even more extreme when the data are contaminated. Figures 4.3 and 4.4 show the optimal choices of \( h \) for some Poisson GLMs as functions of sample sizes for two different contamination models. The optimal \( h \)'s vary tremendously as a function of sample size for the smaller sample sizes. This is very important because in many applications of GLMs the number of observations in each subpopulation is very small.

The size of the mean parameter in a Poisson model appears to have an effect on the estimation as well in the sense that it requires even larger sample sizes for MPHD estimates to be optimal at \( h \) near 0.5 (see Figure 4.1). Simpson (1989) noticed the influence of large Poisson means on the MHD methods. Simpson (1989) found that, for Poisson models with large values of \( \mu \), the \( \chi^2 \) approximation to the HDV test statistic was much less accurate than that for Poisson models with small values of \( \mu \), and he suggested that one may consider using HDV test based on normal approximation to the Poisson rather than that based on Poisson model itself.

Figure 4.2 illustrates the impact of the sizes of the means, \( \mu_i \)'s, on GMPHD estimation in Poisson GLMs. As with MPHD estimation, it appears that the choice of the values of \( h \) not only depends on sample sizes but may also be affected by the magnitudes of the means. It suggests that a single fixed value of \( h \) may not provide the desired optimal results for all cases and that \( h \) ought to be chosen based on data.

Taken together, the simulation work motivates the search for a better method for selecting the penalty weight. A parametric bootstrap (PBoot) algorithm for “estimating” \( h \) is described below. Essentially, the idea behind the algorithm is as follows.

First fit a GLM to the data using, say, GMHD estimation. Then, generate a number of data sets from the fitted model assuming the data really are Poisson. GMPHD estimates are computed for each data set for a number of values of \( h \), and the best value of \( h \) (in terms of mean squared error) is chosen. Finally, GMPHD estimation is carried out on the
original data set using the value of $h$ obtained from the simulation.

**Algorithm 2.1**: GMPHD estimation for discrete GLMs with $h$ chosen by a parametric bootstrapping technique

1. Find the preliminary estimate $\hat{\mu}$ of $\mu$ based on GMPHD estimation:
   
   (a) Given a data set, choose a set of values of $h$, denoted by $\{h\}$, that possibly contains the true optimal $h$, for example, $\{h\} = \{-1.0, -0.9, \ldots, 0.0, \ldots, 0.9, 1.0\}$. The largest element in $\{h\}$ is always 1.0 to cover the ordinary GMHD estimation.
   
   (b) Find the preliminary GMPHD estimate $\hat{\beta}$ of $\beta$ using $h = \text{median}[h]$, where $[h] = [\min\{h\}, 1.0]$. For example, for the $\{h\}$ given in 1(a), $[h] = [-1.0, 1.0]$ and $h = 0.0$.
   
   (c) Compute each $\hat{\mu}_i = e^{-x_i^T \hat{\beta}}$.

2. Generate $r$ bootstrap samples, assuming the estimate $\hat{\mu}$ is the true $\mu$:
   
   (a) Generate each bootstrap sample $Y^* = (Y_1^*, Y_2^*, \ldots, Y_m^*)^T$ such that each $Y_i^* = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^T$ is a random sample of size $n_i$ from $f_{\mu_i}$.

3. For each value of $h$ in $\{h\}$, derive the estimate $\hat{\mu}^*$ of $\mu$ based on GMPHD estimation for each bootstrap sample:
   
   (a) For each value of $h$, compute the GMPHD estimate $\hat{\beta}^*$ of $\beta$ for each bootstrap sample.
   
   (b) For each value of $h$, compute $\hat{\mu}^*$ according to $\hat{\beta}^*$ for each bootstrap sample such that $\hat{\mu}_i^* = e^{-x_i^T \hat{\beta}^*}$, where $\hat{\eta}_i = x_i^T \hat{\beta}$.

4. Find bootstrap optimal $\hat{h}^*$:
   
   (a) For each value of $h$, compute the MSE for each estimator $\hat{\mu}_i^*$ of $\hat{\mu}_i$, $\text{MSE}^*(\hat{\mu}_i^*)$, based on $r$ bootstrap samples, assuming $\hat{\mu}_i$ is true.
(b) For each value of $h$, compute the overall MSE, $\text{MSE}^*(\hat{\mu}^*)$, of all estimators $\hat{\mu}_i^*$'s such that $\text{MSE}^*(\hat{\mu}^*) = \sum_{i=1}^{m} \frac{\text{MSE}^*(\hat{\mu}_i^*)}{\mu_i^2}$. 

(c) Choose a value of $h$ with the smallest $\text{MSE}^*(\hat{\mu}^*)$.

5. Re-estimate $\beta$ based on the original data set, using GMPHD estimation with $h = \hat{h}^*$. 

Algorithm 2.1 provides an effective way to estimate the penalty weight for GMPHD estimation in discrete GLMs which, in practice, usually have very few observations per cell. An extensive simulation study shows that the GMPHD estimates based on this algorithm (using PBoot estimated optimal $h$'s) can be very close (in terms of mean squared error) to the true optimal estimates if the assumed model is true.

Some decisions made in the algorithm are arbitrary but the results are not sensitive to these choices. For example, the range of values of $h$ used in Step 1(a), the initial value of $h$ used in Step 1(b), the number ($r$) of bootstrap samples generated in Step 2, and the criterion used to compute the total MSE, $\text{MSE}^*(\hat{\mu}^*)$ in Step 4(b). When a data set is obtained, information, such as the approximate sizes of the $\mu_i$'s and subsample sizes, is also available. According to Figure 4.2, for the cases with large values of the $\mu_i$'s a wider range of $h$ in Step 1(a) is required. For example, ranges of $[-2.0, 1.0]$ or $[-3.0, 1.0]$ may be needed for samples with sizes less than five.

Using median[$h$] as the initial value of $h$ in Step 1(b) seems to work well for most cases. Generating $r = 25$ bootstrap samples in Step 2 already provides results that are competitive to QL results; an implementation based on $r = 100$ yields very marginal improvement. Altering the computation criterion for total MSE, $\text{MSE}^*(\hat{\mu}^*)$, in Step 4(b) may yield different results; the one suggested in Step 4(b) is scale invariant and provides very efficient GMPHD estimates for small samples.

One disadvantage of Algorithm 2.1 is the lack of robustness due to using a parametric bootstrap strategy. As mentioned previously, the optimal values of $h$ for small samples may
be quite different for contaminated and uncontaminated data. To "estimate" the optimal $h$ by bootstrapping robustly, the key is to be able to generate bootstrap samples which mimic the original sample (contaminated or not) well. The PBoot method in Algorithm 2.1 generates parametric bootstrap samples from the assumed family of distributions which are not subject to the same contamination as the original sample. That is, if a given sample is contaminated, the GMPHD estimates obtained using PBoot estimated optimal $h$'s may not be close to optimal because the PBoot part of the algorithm fails to generate representative bootstrap samples. An alternative way of choosing $h$, which is based on a nonparametric bootstrap (NPBoot) method, has been considered. In the alternate algorithm, Step 2 of algorithm 2.1 is replaced by the following procedure.

2. Generate $r$ bootstrap samples, assuming the estimate $\hat{\mu}$ is the true $\mu$:

(a) Compute the standardized residuals $\hat{\varepsilon} = (\hat{\varepsilon}_1^T, \hat{\varepsilon}_2^T, \ldots, \hat{\varepsilon}_m^T)^T$, where each $\hat{\varepsilon}_i = (\hat{\varepsilon}_{i1}, \hat{\varepsilon}_{i2}, \ldots, \hat{\varepsilon}_{in_i})^T$, and $\hat{\varepsilon}_{ij} = \frac{Y_{ij} - \hat{\mu}_i}{s_i}$, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n_i$.

Here, $s_i$ is the standard error of the $i^{th}$ subpopulation and is a function of $\hat{\mu}_i$ for discrete GLMs. For example, $s_i = \sqrt{\frac{\hat{\mu}_i}{N_i}}$ for Poisson GLMs and $s_i = \sqrt{N_i \hat{\pi}_i (1 - \hat{\pi}_i)}$, where $\hat{\pi}_i = \frac{\hat{\mu}_i}{N_i}$, for binomial GLMs.

(b) For generating each bootstrap sample, resample each $\hat{\varepsilon}_i^* = (\hat{\varepsilon}_{i1}^*, \hat{\varepsilon}_{i2}^*, \ldots, \hat{\varepsilon}_{in_i}^*)^T$ of size $n_i$ with replacement from $\hat{\varepsilon}$ to get $\hat{\varepsilon}^* = (\hat{\varepsilon}_1^T, \hat{\varepsilon}_2^T, \ldots, \hat{\varepsilon}_m^T)^T$.

(c) Generate each bootstrap sample $Y^* = (Y_1^T, Y_2^T, \ldots, Y_m^T)^T$ such that each $Y_{ij}^*$ in $Y_i^* = (Y_{i1}^*, Y_{i2}^*, \ldots, Y_{in_i}^*)^T$ is computed according to $Y_{ij}^* = \hat{\mu}_i + s_i \times \hat{\varepsilon}_{ij}^*$ rounded to the closest integer. Adjust $Y_{ij}^*$ to be in the support of the underlying model distribution. For example, for Poisson GLMs $Y_{ij}^* = \max\{0, Y_{ij}\}$ and for Binomial GLMs $Y_{ij}^* = \max\{0, Y_{ij}\}$ if $Y_{ij}^* < 0$ and $Y_{ij}^* = \min\{N_i, Y_{ij}\}$ if $Y_{ij}^* > N_i$. 

There is a trade-off between using the methods of PBoot and NPBoot for choosing $h$. The GMPHD estimates obtained by this NPBoot alternative are more robust under data contamination at the cost of being less efficient under the assumed model. Figures 4.7, 4.8, and 4.9 illustrate how well Algorithm 2.1 and its NPBoot alternative work under the assumed model and under two types of contamination, individually. As shown in Figures 4.7, the GMPHD estimates using PBoot estimated $h$'s are very close to the true optimal estimates and are much more efficient than those using NPBoot estimated $h$'s when the model is true. Figures 4.8 and 4.9, on the other hand, establish that NPBoot method gives more robust GMPHD estimates than PBoot method does when the data are contaminated, especially for small $\mu_i$'s and for Type II contamination.

For the rest of this dissertation, the PBoot method (i.e., Algorithm 2.1) is used to “estimate” optimal values of $h$ when computing GMPHD estimates. The PBoot method is chosen over NPBoot method because it yields estimates that are more competitive to QL estimates under the model and at the same time are still much more robust than the QL estimates under contamination. (The three graphs on the left side of Figure 4.8 and of Figure 4.9 show that the GMPHD estimates using PBoot-optimal $h$'s are very robust compared to the QL estimates despite being less robust than those using NPBoot-optimal $h$'s.)

Special attention is needed for the cases where GLMs have only one observation per cell. Figures 4.8 and 4.9 show that Algorithm 2.1 is least robust for such cases. This result is expected because, with only one observation per cell, there is no “check” on what distributions the data come from. One can see that even the NPBoot alternative is not very robust for such cases. More work needs to be done in the future for finding the optimal values of $h$ for GMPHD estimation. The goal is to be able to find the values of $h$ through an analytical approach, or at least to be able to figure out a strategy for generating bootstrap samples that encompasses the contamination information of the original sample.

Finally, note that, while the GMPHD estimation based on Algorithm 2.1 yields very
efficient estimates, for the cases with reasonably large subsample sizes per subpopulation (e.g., 100 or more), one may consider using GMPHD estimation with a value of \( h \) close to 0.5 to save computation time. Also note that since a negative value of \( h \) may be involved in GMPHD estimation, the penalized Hellinger distance for each subpopulation \( u_{\rho h,h}(f_{i},\theta,\hat{g}_{ni}) \) defined in Equation (1.9) may not be a distance anymore. The name of the procedure remains unchanged because of the nature of its construction.

As with the relationship between MHD estimation and MPHD estimation, GMPHD estimation differs from GMHD estimation only in its weighting on the empty cells which has no effect on the estimation asymptotically and has no influence on the outlying observations. Thus, the GMPHD estimator enjoys the same asymptotic and robustness properties of the GMHD estimator.

2.4 Generalized Hellinger Deviance Testing for Discrete GLMs

The second main topic of this dissertation is to develop a MHD-based hypothesis testing procedure for GLMs as a robust alternative to the likelihood-based analysis of deviance currently used in GLMs. This section includes a review of Nelder and Wedderburn’s (1972) GLR tests based on QL estimation and the development of GHDV tests based on GMHD estimation for testing the goodness-of-fit of a model, and for selecting between two competitive, nested models.

2.4.1 The GHDV Test for Goodness-of-Fit

First, consider the discrete GLMs with multiple observations per cell. As previously denoted, each \( Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{ini})^T, i = 1, 2, \ldots, m, \) is a random sample from \( g_i \) which is assumed to be a member of the underlying family of distributions, \( \mathcal{F}_i = \{f_{\mu_i(\beta)} : \beta \in \mathcal{B}\}. \) (Note that the scale parameter \( \varphi \) is ignored in the formulation for simplicity because \( \varphi = 1 \) for discrete GLMs.) Let \( l(\mu(\beta)) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \log f_{\mu_i(\beta)}(y_{ij}) \) be the joint log-likelihood function expressed as a function of \( \mu \) which depends on \( \beta \) through a link function \( \ell \). Then,
the generalized log-likelihood ratio test statistic $D^*$ for measuring the goodness-of-fit of a GLM is defined by

$$D^*(\tilde{\mu}(b); y_1, y_2, \ldots, y_m) = 2[l(\tilde{\mu}_{\text{full}}) - l(\tilde{\mu}(b))],$$

where $\tilde{\mu}_{\text{full}}$ corresponds to the maximum value of $l(\mu)$ achievable under a full ($m$-parameter) model, and $b$ is the maximizer of $l(\mu(\beta))$ with respect to $\beta$ over the ($p$-parameter) model under investigation. Nelder and Wedderburn (1972) called this statistic the scaled deviance.

Now, following the discussion in Section 1.5, this statistic may be reformulated in terms of Kullback-Leibler distances. Recall from Section 2.1 that the maximizer of the joint log-likelihood $\sum_{i=1}^{m} \sum_{j=1}^{n_i} \log f_{\mu(\beta)}(y_{ij})$ is exactly the minimizer of $\sum_{i=1}^{m} n_i KL(f_{\mu_i(\beta)}, \hat{g}_{ni})$ for discrete models. Mimicking the derivation in Equation (1.12), the above GLR test statistic can be reformulated as

$$D^*(\tilde{\mu}(b); y_1, y_2, \ldots, y_m) = 2 \sum_{i=1}^{m} n_i [KL(f_{\tilde{\mu}_i(b)}, \hat{g}_{ni}) - KL(f_{\tilde{\mu}_i(\text{full}), \hat{g}_{ni}})].$$

Replacing the Kullback-Leibler distances in Equation (2.9) by twice the squared Hellinger distances yields the generalized Hellinger deviance test statistic $D_{H}$ defined by

$$D_{H}(\tilde{\mu}(b); y_1, y_2, \ldots, y_m) = 2 \sum_{i=1}^{m} n_i [2u_{H}^2(f_{\tilde{\mu}_i(b)}, \hat{g}_{ni}) - 2u_{H}^2(f_{\tilde{\mu}_i(\text{full}), \hat{g}_{ni}})]$$

$$= 8 \sum_{i=1}^{m} n_i [\rho(f_{\tilde{\mu}_i(\text{full}), \hat{g}_{ni}}) - \rho(f_{\tilde{\mu}_i(b)}, \hat{g}_{ni})]$$

by Equation (1.1),

where $\rho(f_{\mu_i(\beta)}, \hat{g}_{ni}) = \int f_{\mu_i(\beta)}^{1/2}(t)\hat{g}_{ni}^{1/2}(t) \, dt$, $\tilde{\mu}_{\text{full}}$ corresponds to the maximum $\sum_{i=1}^{m} n_i \rho(f_{\mu_i}, \hat{g}_{ni})$ achievable under a full ($m$-parameter) model and $b$ is the maximizer of $\sum_{i=1}^{m} n_i \rho(f_{\mu_i(\beta)}, \hat{g}_{ni})$ with respect to $\beta$ over the ($p$-parameter) model under investigation.

Assuming the number of the explanatory variables $p$ is less than the dimension of a full model $m$, the estimation of $\mu$ under a full model must be carried out directly over $\mu$. In GLR tests, $\tilde{\mu}_{\text{full}}$ is the maximizer of $l(\mu)$ over $\mu$ and can be solved analytically. The results are that $\hat{\mu}_{i,\text{full}} = y_i$ if $n_i = 1$ and $\hat{\mu}_{i,\text{full}} = \bar{y}_i$ if $n_i > 1$. In GHDV tests, $\tilde{\mu}_{\text{full}}$ maximizes $\sum_{i=1}^{m} n_i \rho(f_{\mu_i}, \hat{g}_{ni})$ over $\mu$. If $n_i = 1$, it can be solved analytically with the result
that $\hat{\mu}_{i,\text{full}} = y_i$; otherwise it has to be solved numerically because each $\hat{\mu}_{i,\text{full}}$ is the MHD estimate of $\mu_i$ calculated from the $i^{th}$ iid subsample (subject to the constraint that $\varphi$ is the same for all subpopulations when $\varphi$ is not known).

For the case where a discrete GLM has one observation $y_i$ per cell, $\hat{\mu}_{i,\text{full}} = y_i$ for all $i$, $l(\mu) = \sum_{i=1}^{m} \log f_{\mu_i(\beta)}(y_i)$, and $\sum_{i=1}^{m} n_i \rho(f_{\mu_i(\beta)}; \hat{g}_{n_i}) = \sum_{i=1}^{m} f^{1/2}_{\mu_i(\beta)}(y_i)$, so $D^*$ and $D_H$ simplify to

$$D^*(\hat{\mu}(b); y_1, y_2, \ldots, y_m) = 2 \sum_{i=1}^{m} \log \left[ \frac{f_{\hat{y}_i}(y_i)}{f_{\hat{\mu}_i(b)}(y_i)} \right],$$

and

$$D_H(\hat{\mu}(b); y_1, y_2, \ldots, y_m) = 8 \sum_{i=1}^{m} [f^{1/2}_{\hat{y}_i}(y_i) - f^{1/2}_{\hat{\mu}_i(b)}(y_i)],$$

respectively.

### 2.4.2 The GHDV Test for Comparing Two Nested Models

Let $H_0$ denote the model under test and $H_1$ be the extended model containing additional explanatory variables. That is, the test of interest is $H_0 : \beta = \beta_0 \in B_0$ against $H_1 : \beta = \beta_1 \in B_1$, where $B_0$ and $B_1$ are of dimension $p$ and $q$, respectively, $B_0 \subset B_1$, and $p < q < m$. Then, following the idea of LR tests for nested models in Section 1.5, the GLR test statistic for comparing two nested models is the difference between two scaled deviances

$$\Delta D^*(\hat{\mu}(b_0), \hat{\mu}(b_1); y_1, y_2, \ldots, y_m)$$

$$= D^*(\hat{\mu}(b_0); y_1, y_2, \ldots, y_m) - D^*(\hat{\mu}(b_1); y_1, y_2, \ldots, y_m)$$

$$= 2[l(\hat{\mu}(b_1)) - l(\hat{\mu}(b_0))],$$

where $b_0$ and $b_1$ are maximizers of $l(\hat{\mu}(\beta))$ with respect to $\beta$ under null and alternative models, respectively. Similarly, the GHDV test statistic for nested models has the form

$$\Delta D_H(\hat{\mu}(b_0), \hat{\mu}(b_1); y_1, y_2, \ldots, y_m)$$

$$= D_H(\hat{\mu}(b_0); y_1, y_2, \ldots, y_m) - D_H(\hat{\mu}(b_1); y_1, y_2, \ldots, y_m)$$

$$= 8 \sum_{i=1}^{m} n_i [\rho(f_{\hat{\mu}_i(b_1)}; \hat{g}_{n_i}) - \rho(f_{\hat{\mu}_i(b_0)}; \hat{g}_{n_i})],$$
where $b_0$ and $b_1$ are maximizers of $\sum_{i=1}^{m} n_i\beta(f_{\mu_i(\beta)}, \hat{g}_{ni})$ with respect to $\beta$ under null and alternative models, respectively.

This section concludes with three notes concerning the validity of GLR and GHDV tests and the extension of GHDV tests to continuous GLMs.

Notes:

1. McCullagh and Nelder (1989) stated that the scaled deviance function (the GLR test statistic for goodness-of-fit) may be approximated by a $\chi^2_{(m-p)}$ distribution in some circumstances—for example, in discrete data problems where the counts are large—but in general, the $\chi^2$ approximations for the scaled deviance itself are not very good even as $m \to \infty$. Further work remains to be done on the asymptotic distribution of the scaled deviance. On the other hand, McCullagh and Nelder (1989) suggested that the $\chi^2$ approximation for the difference between scaled deviances for nested models is quite accurate. The $\chi^2$ degrees of freedom for nested models is $q - p$.

2. Concerning the relationship between HDV tests and LR tests, one might expect that the GHDV tests would be asymptotically equivalent to the GLR tests if the null model is correct. An extensive study based on simulation supports this assertion whereas, for GLMs with very few observations per cell, the convergence of the GHDV test statistics, (including both $D_H$ and $\Delta D_H$), to their limiting $\chi^2$ distributions is much too slow to be useful. Section 2.5 introduces the GPHDV tests based on the GMPHD estimators (which, with suitable penalty factors, are much more efficient than the GMHD estimators in small samples). The intention is to develop alternative robust tests that converge to the GLR tests faster than the GHDV tests and see if a $\chi^2$ approximation is applicable for small samples. In Section 2.6, a nonparametric bootstrap method is used to determine critical values, which are percentiles of the bootstrap build-up sampling distributions of the GPHDV test statistics, to further improve the accuracy of the type I errors in small samples.
3. GHDV tests may be applied to continuous GLMs with multiple observations per cell. If $\varphi$ is not known (such as in normal, gamma, and inverse Gaussian GLMs), the estimate of $\varphi$, $\hat{\varphi}$, is computed from the largest model of interest and the same $\hat{\varphi}$ is used in all subsequent models. This strategy may not be satisfactory if $\varphi$ controls the shape of the model distributions, as in the gamma distributions. This is a topic to be investigated in the future research.

2.5 Generalized Penalized Hellinger Deviance Testing for Discrete GLMs

In this section, PHDV tests are generalized to discrete GLMs. The resulting generalized penalized Hellinger deviance tests, based on the GMPHD estimators, are introduced as an improved robust alternative to the GLR tests over the ordinary GHDV tests in terms of how fast their test statistics converge to the limiting $\chi^2$ distributions.

2.5.1 The GPHDV Test for Goodness-of-Fit

Following the assumptions and notation in Section 2.4, the generalized penalized Hellinger deviance test statistic for goodness-of-fit, $D_{PH,h}$, may be obtained by replacing the Kullback-Leibler distances in Equation (2.9) by twice the squared penalized Hellinger distances. According to Equation (1.9) the statistic has the form

$$D_{PH,h}(\hat{\mu}(b); y_1, y_2, \ldots, y_m) = 2 \sum_{i=1}^{m} n_i \left[ 2u_{PH,h}(f_{\hat{\mu}_{i}(b)}, \hat{g}_{n_i}) - 2u_{PH,h}(f_{\mu_{i,full}}, \hat{g}_{n_i}) \right]$$

$$= 4 \sum_{i=1}^{m} n_i \left[ 2\rho(f_{\hat{\mu}_{i,full}}, \hat{g}_{n_i}) - 2\rho(f_{\hat{\mu}_{i}(b)}, \hat{g}_{n_i}) \right] - (1 - h) \left[ \sum_{\hat{g}_{n_i}(t) \neq 0} f_{\mu_{i,full}(t)} - \sum_{\hat{g}_{n_i}(t) \neq 0} f_{\mu_{i}(b)}(t) \right],$$

where $\hat{\mu}_{full}$ corresponds to the maximum value of $\sum_{i=1}^{m} n_i \left\{ 2\rho(f_{\mu_{i}}, \hat{g}_{n_i}) - (1 - h) \sum_{\hat{g}_{n_i}(t) \neq 0} f_{\mu_{i}}(t) \right\}$ achievable under a full $(m$-parameter) model and $b$ is the maximizer of $\sum_{i=1}^{m} n_i \left\{ 2\rho(f_{\mu_{i}(\beta)}, \hat{g}_{n_i}) - (1 - h) \sum_{\hat{g}_{n_i}(t) \neq 0} f_{\mu_{i}(\beta)}(t) \right\}$ with respect to $\beta$ over the $(p$-parameter) model under investigation.

Analogous to the GHDV test, $\hat{\mu}_{full}$ may be obtained by maximizing $\sum_{i=1}^{m} n_i \left\{ 2\rho(f_{\mu_{i}}, \hat{g}_{n_i}) - (1 - h) \sum_{\hat{g}_{n_i}(t) \neq 0} f_{\mu_{i}}(t) \right\}$ over $\mu$. However, for the GPHDV test, the estimate $\hat{\mu}_{full}$ may de-
pend on the value of $h$. For $n_i = 1$, $\hat{\mu}_{i,\text{full}} = y_i$ regardless of the value of $h$, while for $n_i > 1$, $\hat{\mu}_{i,\text{full}}$ is the MPHD estimate of $\mu_i$ based on the $i^{th}$ iid subsample, subject to the constraints that $h$ and $\varphi$ are both the same for all subpopulations, and different values of $h$ may lead to different results. Therefore, it is necessary to determine $h$ in some way.

One suitable value of $h$ for the GPHDV test for goodness-of-fit is the optimal $h$ derived according to the estimation under the full model. The same $h$ is then used in the estimation under the model of investigation. Algorithm 2.2 explains how $\hat{\mu}_{\text{full}}$ and its corresponding optimal $h$ may be obtained. The idea is very similar to that of Algorithm 2.1, but $\mu$ is estimated directly. This algorithm applies to all discrete GLMs with at least one $n_i > 1$.

Algorithm 2.2: The estimation of $\mu$ for the saturated or full model for the GPHDV test of goodness-of-fit of a discrete GLM with $h$ chosen by parametric bootstrapping

1. Find the preliminary estimate $\hat{\mu}_{\text{full}}$ of $\mu$

   (a) Given a data set, choose a set of values of $h$, $\{h\}$, that possibly contains the true optimal $h$. The largest component in $\{h\}$ is always 1.0.

   (b) Find the preliminary estimate $\hat{\mu}_{\text{full}}$ of $\mu$ such that let $\hat{\mu}_{i,\text{full}} = y_i$ if $n_i = 1$ and $\hat{\mu}_{i,\text{full}}$ be the MPHD estimate of $\mu_i$ using $h = \text{median}[h]$ based on the $i^{th}$ iid subsample $Y_i$ if $n_i > 1$, where $[h] = \lfloor \min h, 1.0 \rfloor$.

2. Generate $r$ bootstrap samples, assuming the estimate $\hat{\mu}_{\text{full}}$ is the true $\mu$

   (a) Generate each bootstrap sample $Y^*_i = (Y^*_1, Y^*_2, \ldots, Y^*_n)^T$ such that each $Y^*_i = (Y^*_{i1}, Y^*_{i2}, \ldots, Y^*_{in_i})^T$ is a random sample of size $n_i$ from $f_{\hat{\mu}_{i,\text{full}}}$.

3. For each value of $h$ in $\{h\}$, derive the estimate $\hat{\mu}_{\text{full}}^*$ of $\hat{\mu}_{\text{full}}$ for each bootstrap sample:

   (a) For each given value of $h$ and for each bootstrap sample, let $\hat{\mu}_{i,\text{full}}^* = y_i^*$ if $n_i = 1$ and $\hat{\mu}_{i,\text{full}}^*$ be the MPHD estimate of $\hat{\mu}_i$ based on the $i^{th}$
iid subsample $Y^*_i$ if $n_i > 1$.

4. Find bootstrap optimal $\hat{h}^*$:

(a) For each value of $h$, compute the MSE for each estimator $\hat{\mu}^*_{i,\text{full}}$, based on $r$ bootstrap samples, assuming $\hat{\mu}^*_{i,\text{full}}$ is true.

(b) For each value of $h$, compute the overall MSE, $\text{MSE}^*(\hat{\mu}^*_{\text{full}})$, of all estimators $\hat{\mu}^*_{i,\text{full}}$'s such that $\text{MSE}^*(\hat{\mu}^*_{\text{full}}) = \sum_{i=1}^{m} \frac{\text{MSE}^*(\hat{\mu}^*_{i,\text{full}})}{\mu^*_{i,\text{full}}}.$

(c) Choose a value of $h$ with the smallest $\text{MSE}^*(\hat{\mu}^*_{\text{full}})$.

5. Re-estimate $\mu$ based on the original data set with $h = \hat{h}^*$.

For discrete GLMs with only one observation per cell (i.e., with $n_i$'s all equal to 1), the selection of $h$ needs special attention. $\hat{\mu}^*_{\text{full}}$ in this case, are simply the data values themselves no matter what value of $h$ is given; hence, no contribution for choosing the optimal $h$ will be made from computing $\hat{\mu}^*_{\text{full}}$. In such a case, the optimal $h$ used in the test may be derived according to the model under investigation using Algorithm 2.1. The test statistic for this case is reduced to the form

$$D_{PH,h} = 4 \sum_{i=1}^{m} \{2[f_{y_i}^{1/2}(y_i) - f_{\hat{\mu}^*_{i}^{1/2}}(y_i)] - (1 - h)[f_{y_i}(y_i) - f_{\hat{\mu}^*_i}(y_i)]\},$$

where $h$ is the optimal $h$ corresponding to the estimator $b$.

2.5.2 The GPHDV Test for Comparing Two Nested Models

The GPHDV test statistic for nested models is the difference between two GPHDV test statistics for goodness-of-fit of models under null and alternative hypotheses; its form is given by

$$\Delta D_{PH,h}(\hat{\mu}(b_0), \hat{\mu}(b_1); y_1, y_2, \ldots, y_m)$$

$$= D_{PH,h}(\hat{\mu}(b_0); y_1, y_2, \ldots, y_m) - D_{PH,h}(\hat{\mu}(b_1); y_1, y_2, \ldots, y_m)$$
where $b_0$ and $b_1$ are maximizers of $\sum_{i=1}^{m} n_i \left\{ 2 \rho(f_{\mu_i}(b_1), \hat{g}_{n_i}) - 2 \rho(f_{\mu_i}(b_0), \hat{g}_{n_i}) \right\} - (1 - h) \left[ \sum_{\hat{g}_{n_i}(t) \neq 0} f_{\mu_i}(b_1)(t) - \sum_{\hat{g}_{n_i}(t) \neq 0} f_{\mu_i}(b_0)(t) \right]$, with respect to $\beta$ under null and alternative models, respectively. In this case, $h$ is derived according to Algorithm 2.1 based on the larger model. The same $h$ is used in the smaller model.

This section ends with some remarks regarding how well GPHDV tests perform and why nonparametric bootstrap GPHDV tests, which are introduced in the next section, are needed. To help illustrate the comparison among all tests in this chapter, Section 4.2.3.1 contains representative test results from our preliminary simulation. Figures 4.10 and 4.11 compare the observed levels of the GLR tests, the GHDV tests, the GPHDV tests using $h = 0.5$ and the GPHDV tests using the optimal $h$ under the null hypothesis for goodness-of-fit and for comparing two nested models, respectively. It appears that, asymptotically, all GPHDV tests (including the GHDV tests) are equivalent to the GLR tests when the null model is true whereas, for GLMs with small subsamples, results of the GPHDV tests could be very different from those of the GLR tests. The results indicate that although the penalty process significantly increases the rate of convergence of GHDV tests to their limiting $\chi^2$ distributions, the convergence is still not fast enough to give usable levels for GLMs with small subsamples. Since the $\chi^2$ approximation for GPHDV tests is not very good, nonparametric bootstrap GPHDV tests which obtain critical values according to the nonparametric bootstrap sampling distributions of the test statistics are introduced in the next section. It has been shown that such GPHDV tests give much more accurate levels than those based on the $\chi^2$ approximation.

2.6 Nonparametric Bootstrap GPHDV Testing for Discrete GLMs

This section introduces a nonparametric bootstrap method for deriving much more accurate and much more robust critical values for the GPHDV tests as an alternative to the
\( \chi^2 \) approximation. The resulting tests are named the **nonparametric bootstrap generalized penalized Hellinger deviance** (NPB-GPHDV) tests.

2.6.1 The NPB-GPHDV Test for Goodness-of-Fit

Assuming that \( H_0 : \beta = \beta_0 \in B_0 \) is the goodness-of-fit test of interest, where \( B_0 \) is of dimension \( p \). Let \( X_0 = (x_{01}, x_{02}, \ldots, x_{0m})^T \) be the \( m \times p \) design matrix or the matrix of explanatory variates of the model under investigation.

**Algorithm 2.3:** Nonparametric bootstrap GPHDV test for goodness-of-fit

1. Find the GMHD estimates and derive a suitable value of \( h \) for the test:
   Given a data set, consider two cases.
   - **Case 1:** \( n_i > 1 \) for at least one \( i \):
     - (a) Use Algorithm 2.2 to compute \( \hat{\mu}_{0i} \) and \( \hat{h}^* \).
     - (b) Using \( h = \hat{h}^* \), find the GMHD estimate \( \hat{\beta}_0 \) of \( \beta_0 \) under a \( p \)-parameter model.
   - **Case 2:** \( n_i = 1 \) for all \( i \):
     - (a) Let \( \hat{\mu}_{i,full} = y_i \) for all \( i \).
     - (b) Use Algorithm 2.1 to compute the GMHD estimate \( \hat{\beta}_0 \) of \( \beta_0 \) and \( \hat{h}^* \) under a \( p \)-parameter model.

2. Compute the GPHDV test statistic, \( D_{PH,\hat{h}} \):
   - (a) Compute \( \hat{\mu}_0 \) such that each \( \hat{\mu}_{0i} = \ell^{-1}(\hat{\eta}_{0i}) \), where \( \hat{\eta}_{0i} = x_{0i}^T \hat{\beta}_0 \).
   - (b) Compute \( D_{PH,\hat{h}} \) based on \( \hat{\mu}_{0i} \) and \( \hat{\mu}_0 \).

3. Construct a nonparametric bootstrap distribution for \( D_{PH,\hat{h}} \):
   - (a) Generate 1000 bootstrap samples, assuming the estimate \( \hat{\mu}_0 \) is true:
     - (i) Compute the standardized residuals \( \hat{e} = (\hat{e}_1^T, \hat{e}_2^T, \ldots, \hat{e}_m^T)^T \), where each \( \hat{e}_i = (\hat{e}_{i1}, \hat{e}_{i2}, \ldots, \hat{e}_{im})^T \), and \( \hat{e}_{ij} = \frac{Y_{ij} - \hat{\mu}_{0i}}{s_i} \), for \( i = 1, 2, \ldots m \) and \( j = 1, 2, \ldots n_i \). Here, \( s_i \) is the standard error of the \( i^{th} \)
subpopulation and is a function of $\dot{\mu}_0i$ for discrete GLMs. For example, $s_i = \sqrt{\mu_0i}$ for Poisson GLMs and $s_i = \sqrt{N_i \hat{\pi}_0i(1 - \hat{\pi}_0i)}$, where $\hat{\pi}_0i = \frac{\mu_0i}{N_i}$, for binomial GLMs.

(ii) For generating each bootstrap sample, resample each

$\hat{\epsilon}^*_i = (\hat{\epsilon}^*_{i1}, \hat{\epsilon}^*_{i2}, \ldots, \hat{\epsilon}^*_{im})^T$ of size $n_i$ with replacement from $\hat{\epsilon}$ to get $\hat{\epsilon}^* = (\hat{\epsilon}^*_1, \hat{\epsilon}^*_2, \ldots, \hat{\epsilon}^*_m)^T$.

(iii) Generate each bootstrap sample $Y^* = (Y^*_1, Y^*_2, \ldots, Y^*_m)^T$ such that each $Y^*_{ij}$ in $Y^*_i = (Y^*_{i1}, Y^*_{i2}, \ldots, Y^*_{im})^T$ is computed according to $Y^*_{ij} = \mu_0i + s_i \times \hat{\epsilon}^*_{ij}$ rounded to the closest integer. Adjust $Y^*_{ij}$ to be in the support of the underlying null model distribution. For example, for Poisson GLMs $Y^*_{ij} = \max\{0, Y^*_{ij}\}$ and for Binomial GLMs $Y^*_{ij} = \max\{0, Y^*_{ij}\}$ if $Y^*_{ij} < 0$ and $Y^*_{ij} = \min\{N_i, Y^*_{ij}\}$ if $Y^*_{ij} > N_i$.

(b) For each bootstrap sample, find the estimate $\hat{\mu}^*_{\text{full}}$ of $\mu_0$ and the GMPHD estimate $\hat{\beta}^*_0$ of $\beta_0$ under the full and the $p$-parameter model, respectively, using $h = h^*$.

(c) Similar to Step 2, compute the GPHDV test statistic $D^*_{PH,h^*}$ based on $\hat{\mu}^*_{\text{full}}$ and $\hat{\mu}^*$ for each one of the 1000 bootstrap samples.

4. Make a decision:

(a) Given a nominal level $\alpha$, reject the null hypothesis if $D^*_{PH,h^*}$ is in the largest $1000 \times \alpha$ values of $D^*_{PH,h^*}$; otherwise do not reject.

2.6.2 The NPB-GPHDV Test for Comparing Two Nested Models

Assuming that $H_0 : \beta = \beta_0 \in B_0$ versus $H_1 : \beta = \beta_1 \in B_1$ is the test of interest, where $B_0$ and $B_1$ are of dimension $p$ and $q$, respectively, $B_0 \subset B_1$, and $p < q < m$. Let $X_0 = (x_{01}, x_{02}, \ldots, x_{0m})^T$ and $X_1 = (x_{11}, x_{12}, \ldots, x_{1m})^T$ be the $m \times p$ and $m \times q$ design matrices or the matrices of explanatory variates of the null and the alternative models,
Algorithm 2.4: Nonparametric bootstrap GPHDV test for nested GLMs

1. Find the GMPHD estimates and derive a suitable value of \( h \) for the test:
   
   (a) Given a data set, use Algorithm 2.1 to estimate \( \beta_1 \) under a
       \( q \)-parameter model to get \( \hat{\beta}_1 \) and \( h_1^* \).

   (b) Using \( h = h_1^* \), find the GMPHD estimate \( \hat{\beta}_0 \) of \( \beta_0 \) under a \( p \)-parameter
       model.

2. Compute the GPHDV test statistic, \( \Delta D_{ph,h_1^*}^* \):
   
   (a) Compute \( \hat{\eta}_0 = X_0 \hat{\beta}_0 \) and \( \hat{\eta}_1 = X_1 \hat{\beta}_1 \).

   (b) Compute \( \hat{\mu}_0 \) and \( \hat{\mu}_1 \) such that each \( \hat{\mu}_{0i} = \ell^{-1}(\hat{\eta}_{0i}) \) and each \( \hat{\mu}_{1i} = \ell^{-1}(\hat{\eta}_{1i}) \).

   (c) Compute \( \Delta D_{ph,h_1^*}^* \) based on \( \hat{\mu}_0 \) and \( \hat{\mu}_1 \).

3. Construct a nonparametric bootstrap distribution for \( \Delta D_{ph,h_1^*}^* \):
   
   (a) Same as Step 3(a) of Algorithm 2.3.

   (b) For each bootstrap sample, find the GMPHD estimates \( \hat{\beta}_0^* \) and \( \hat{\beta}_1^* \) of
       \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) under a \( p \)-parameter and a \( q \)-parameter model,
       respectively, using \( h = h_1^* \).

   (c) Similar to Step 2, compute the GPHDV test statistic \( \Delta D_{ph,h_1^*}^* \) based
       on \( \hat{\mu}_0^* \) and \( \hat{\mu}_1^* \) for each one of the 1000 bootstrap samples.

4. Make a decision:
   
   (a) Given a nominal level \( \alpha \), reject the null hypothesis if \( \Delta D_{ph,h_1^*}^* \) is
       in the largest \( 1000 \times \alpha \) values of \( \Delta D_{ph,h_1^*}^* \); otherwise do not reject.

Similar to the study for choosing optimal \( h \) for GMPHD estimation, a parametric bootstrap alternative of the NPB-GPHDV tests is also considered in the study for robust tests. These parametric bootstrap generalized penalized Hellinger deviance (PB-GPHDV) tests can be derived by replacing Step 3(a) in Algorithm 2.3 and 2.4 by the following.
(a) Generate 1000 bootstrap samples, assuming the estimate $\hat{\mu}_0$ is true, such that, in each bootstrap sample $Y^* = (Y_1^{*T}, Y_2^{*T}, \ldots, Y_m^{*T})^T$,

$Y_i^* = (Y_{i1}^*, Y_{i2}^*, \ldots, Y_{in_i}^*)^T$ is a random sample of size $n_i$ from $f_{\mu_0}$.  

As with the GMPHD estimation, a trade-off between using the PB-GPHDV tests and the NPB-GPHDV tests is expected. Section 4.2.3.1 contains results for comparing the NPB-GPHDV tests to their parametric alternative (the PB-GPHDV tests) and to other tests concerned in this chapter under no contamination, Type I contamination, and Type II contamination. (Refer to Figures 4.12, 4.13, 4.14, 4.15, 4.16, and 4.17 for the comparison.) Both the PB-GPHDV and the NPB-GPHDV tests are shown to give much more accurate observed levels than the GPHDV tests for most cases. The PB-GPHDV tests perform as well as the GLR tests when data are not contaminated, but are not very robust under data contamination. The NPB-GPHDV tests give competitive results to those of the PB-GPHDV tests except for the case where GLMs have one observation per cell under no contamination, but are much more robust than the PB-GPHDV tests under contamination. None of the tests studied are robust against Type II contamination when testing for goodness-of-fit, while the NPB-GPHDV tests are the most robust tests for comparing two nested models under Type II contamination. For the remainder of this dissertation, the NPB-GPHDV tests are chosen over the PB-GPHDV tests.

More research needs to be done to improve the performance of the NPB-GPHDV tests for the cases with one observation per cell and for goodness-of-fit tests against Type II contamination. The poor performance of the NPB-GPHDV tests for these particular cases may once again be explained by the lack of a “check” on the real data distributions. Deriving the sampling distribution for GPHDV tests statistics theoretically is not an easy task. More effort will be invested in improving the bootstrap procedure for the GPHDV tests in the future research.
CHAPTER 3
COMPUTATION AND IMPLEMENTATION

This section contains the computational details of this dissertation including some general simulation information, methods for performing optimizations, and methods for computing robust starting values for the optimizations.

3.1 General Simulation Information

All the programs are written in Fortran 77 and extensively tested on a SUN workstation at the department of Mathematics and Statistics, Utah State University. Uniform (0,1) random numbers are generated using the NAG mathematical Fortran library. Poisson random numbers are generated follows an algorithm described in Ross (1997, §5.1, page 64, Remark) which applies the relationship between the Possion and the exponential distribution. Binomial random numbers are generated by the inverse transform method (Ross 1997, §4.3). All the simulation results presented in this dissertation are based on 1000 replications except the results for iid MPHD estimation which are based on 5000 replications.

3.2 Optimization

Computation of GMPHD estimates involves an optimization. All optimizations in this dissertation are carried out by Powell’s method (Press et al., 1989) which does not require derivatives. An alternative optimization algorithm, a general fitting algorithm for computing the GMPHD estimates in discrete GLMs, was derived and tested in our early study. This alternative algorithm, which is very similar to Nelder and Wedderburn’s (1972) iteratively reweighted least squares (IRWLS) algorithm for QL estimation, is shown to produce GMPHD estimates which have slightly larger total mean squared errors of the $\mu_i$’s than those obtained using Powell’s algorithm but it is an important tool for computing robust starting values for Powell’s method. Applying this algorithm once generates the initial values of
\( \beta \) (named the one-step GMPHD estimate) for Powell’s algorithm, which are much more robust than the one-step QL estimates obtained by performing Nelder and Wedderburn’s (1972) IRWLS algorithm once.

For computational simplicity, all Powell optimizations in this dissertation are started with one-step GMHD estimates \( (h = 1) \) so that the choice of \( h \) need not be considered when computing starting values. Using different values of \( h \) in computing starting values has negligible impact on the final estimation results. Section 3.2.1 establishes how to derive an IRWLS algorithm for GMHD estimation. A slightly more complicated derivation for GMPHD estimation can be carried out using the same idea and is available but is not included in this dissertation.

3.2.1 A General Fitting Algorithm for Computing GMHD Estimators in Discrete GLMs

Nelder and Wedderburn (1972) derived a general fitting algorithm for solving QL estimators of \( \beta \) in GLMs (for both discrete and continuous distributions). The equation at the heart of the algorithm takes the form of weighted least squares with an adjusted response variable. The iterative aspect of the algorithm is due to the fact that both the weights and the response variable adjustment depend on the estimate \( b \) of \( \beta \) from the previous iteration.

To illustrate the derivation of the IRWLS algorithm of Nelder and Wedderburn (1972), consider a GLM with one observation per cell. According to Equation (1.13), the joint log-likelihood of the observations is of the form

\[
\ell(\theta, \varphi; y) = \sum_{i=1}^{m} l(\theta_i, \varphi; y_i) = \sum_{i=1}^{m} \{[y_i \theta_i - b(\theta_i)]/a_i(\varphi) + c(y_i; \varphi)\},
\]

where \( \theta_i = \theta(\mu_i(\beta)) \) is the canonical parameter. The QL estimate of \( \beta \) is the maximizer of \( \ell \) and it is given uniquely by the solution of the estimating equations \( \frac{\partial \ell}{\partial \beta} = 0 \). In general, the equations \( \frac{\partial \ell}{\partial \beta} = 0 \) have to be solved numerically and often may be solved by the Newton-Raphson method,

\[
\begin{bmatrix}
\frac{\partial^2 \ell}{\partial \beta \partial \beta^T} \\
\end{bmatrix}_{\beta = b^{(\text{old})}}
\begin{bmatrix}
b^{(\text{new})} \\
\end{bmatrix}
= \begin{bmatrix}
\frac{\partial^2 \ell}{\partial \beta \partial \beta^T} \\
\end{bmatrix}_{\beta = b^{(\text{old})}}
\begin{bmatrix}
b^{(\text{old})} \\
\end{bmatrix}
- \begin{bmatrix}
\frac{\partial \ell}{\partial \beta} \\
\end{bmatrix}_{\beta = b^{(\text{old})}}.
\]
where \([- \frac{\partial^2 l}{\partial \beta \partial \beta^T}\] is the Hessian matrix of \(l\) and \([- \frac{\partial l}{\partial \beta}\] is the gradient of \(l\). The IRWLS for QL estimation is derived using an alternative procedure, called Fisher’s method of scoring, which replaces \([- \frac{\partial^2 l}{\partial \beta \partial \beta^T}\] in the Newton-Raphson method by the Fisher information matrix of \(\beta\), \(I(\beta) = -\mathcal{E}[\frac{\partial^2 l}{\partial \beta \partial \beta^T}]\). The resulting iterative equation is

\[
I(b^{(\text{old})}) b^{(\text{new})} = I(b^{(\text{old})}) b^{(\text{old})} + \left[\frac{\partial l}{\partial \beta}\right]_{\beta = b^{(\text{old})}},
\]

where \(I(b^{(\text{old})}) = I(\beta)\) evaluated at \(b^{(\text{old})}\). After some tedious algebra (see Nelder and Wedderburn [1972] for details), this iterative equation can be reduced to

\[
(3.1) \quad X^T W^{(\text{old})} X b^{(\text{new})} = X^T W^{(\text{old})} z^{(\text{old})}.
\]

Here, \(X = (x_1, x_2, \ldots, x_m)^T\) is a \(m \times p\) design matrix or covariate matrix and \(z^{(\text{old})}\) is a \(m \times 1\) vector of adjusted dependent variates with each component

\[
z_i = \eta_i + (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i}
\]

evaluated at \(b^{(\text{old})}\). \(W\) is a \(m \times m\) diagonal weight matrix with each diagonal element

\[
w_i = \begin{cases} n_i \left( \frac{1}{b''(\theta_i)} \frac{\partial \eta_i}{\partial \mu_i} \right)^2 & \text{if } a_i(\varphi) = a(\varphi) \\ n_i \left( \frac{1}{b''(\theta_i)} \frac{\partial \eta_i}{\partial \mu_i} \right)^2 & \text{if } a_i(\varphi) = \varphi, \end{cases}
\]

where \(b''(\theta_i) = \frac{\partial^2 b(\theta_i)}{\partial \theta_i^2}\). \(W^{(\text{old})}\) is \(W\) evaluated at \(b^{(\text{old})}\).

When there is more than one observation per cell,

\[
l(\theta, \varphi; y_1, y_2, \ldots, y_m) = \sum_{i=1}^{m} l(\theta_i, \varphi; y_i) = \sum_{i=1}^{m} \{n_i [\bar{y}_i \theta_i - b(\theta_i)] / a_i(\varphi) + \sum_{j=1}^{n_i} c(y_{i,j}; \varphi)\}
\]

but the same algorithm applies only with \(z_i\) and \(w_i\) slightly modified

\[
z_i = \eta_i + (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i}
\]

\[
w_i = \begin{cases} n_i \left( \frac{1}{b''(\theta_i)} \frac{\partial \eta_i}{\partial \mu_i} \right)^2 & \text{if } a_i(\varphi) \text{ is the same for all subpopulations} \\ n_i \left( \frac{1}{b''(\theta_i)} \frac{\partial \eta_i}{\partial \mu_i} \right)^2 & \text{otherwise.}
\end{cases}
\]

In this dissertation, a similar general fitting algorithm (also called the IRWLS) is derived for GMHD estimation in discrete GLMs. Consider a discrete GLM with one observation
per cell. Denote by $Q(\theta, \varphi; y)$ (as opposed to $l(\theta, \varphi; y)$) the function to be maximized with respect to $\beta$ in GMHD estimation, parametrized in terms of the canonical parameters $\theta_i$'s, where $\theta_i = \theta(\mu_i(\beta))$. Then, according to Equation (2.5), $Q$ has the form

$$Q(\theta, \varphi; y) = \sum_{i=1}^{m} f_{\theta_i, \varphi}^{1/2}(y_i),$$

where $\varphi = 1$ for discrete GLMs. The GMHD estimate of $\beta$ is determined by solving the estimating equations $\frac{\partial Q}{\partial \beta} = 0$. To derive an IRWLS algorithm, a strategy similar to Fisher's method of scoring may be used for solving $\frac{\partial Q}{\partial \beta} = 0$. The idea is to replace the negative Hessian matrix of $Q$, $-\left[\frac{\partial^2 Q}{\partial \beta \partial \beta^T}\right]$, in the Newton-Raphson iterative equation

$$\left[\frac{\partial^2 Q}{\partial \beta \partial \beta^T}\right]_{\beta=b^{(\text{old})}} b^{(\text{new})} = \left[\frac{\partial^2 Q}{\partial \beta \partial \beta^T}\right]_{\beta=b^{(\text{old})}} b^{(\text{old})} - \left[\frac{\partial Q}{\partial \beta}\right]_{\beta=b^{(\text{old})}}$$

by $J(\beta) = -E[\frac{\partial^2 Q}{\partial \beta \partial \beta^T}]$ and then to reduce the resulting iterative equation

$$J(b^{(\text{old})}) b^{(\text{new})} = J(b^{(\text{old})}) b^{(\text{old})} + \left[\frac{\partial Q}{\partial \beta}\right]_{\beta=b^{(\text{old})}}$$

to the form of Equation (3.1). This procedure is very similar to Fisher’s method of scoring except that $J(\beta)$ is no longer the Fisher information matrix of $\beta$.

For discrete GLMs with more than one observations per cell, $Q$, by Equation (2.4), may be written as

$$Q(\theta, \varphi; y_1, y_2, \ldots, y_m) = \sum_{i=1}^{m} n_i \sum_t \left( \frac{N_{it}}{n_i} \right)^{1/2} f_{\theta_i, \varphi}^{1/2}(t),$$

where recall that each $N_{it}$ is the frequency of $t$ among $y_i$. The details of the derivation of the IRWLS algorithm are given below for this more general case.

The derivation of the IRWLS algorithm for GMHD estimation in discrete GLMs:

1. Derive $\frac{\partial Q}{\partial \beta_k}$:

$$\frac{\partial Q}{\partial \beta_k} = \sum_{i=1}^{m} \sqrt{n_i} \sum_t (N_{it})^{1/2} \left\{ \frac{\partial}{\partial \beta_k} f_{\theta_i, \varphi}^{1/2}(t) \right\}$$

$$= \sum_{i=1}^{m} \sqrt{n_i} \sum_t (N_{it})^{1/2} \left\{ \frac{\partial}{\partial \theta_i} f_{\theta_i, \varphi}^{1/2}(t) \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial n_i} \cdot \frac{\partial n_i}{\partial \beta_k} \right\}$$
\[
\begin{align*}
(3.2) \quad \frac{1}{2} \sum_{i=1}^{m} \sqrt{n_i} \sum_{t} (N_{it})^{1/2} \left\{ f_{\theta_i, \varphi}^{1/2}(t) \cdot \frac{t - \mu_i}{\text{Var}(Y_{ij})} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_{ik} \right\} \\
\text{because} \\
\frac{\partial}{\partial \theta_i} f_{\theta_i, \varphi}^{1/2}(t) = \frac{1}{2} \left[ \frac{\partial}{\partial \theta_i} l(\theta_i, \varphi; t) \right] f_{\theta_i, \varphi}^{1/2}(t) \\
= \frac{1}{2} \left[ \frac{t - \mu_i}{a_i(\varphi)} \right] f_{\theta_i, \varphi}^{1/2}(t), \\
\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{\frac{\partial \mu_i}{\partial \theta_i}} = \frac{1}{b''(\theta_i)}, \\
\text{and} \\
\frac{\partial \eta_i}{\partial \beta_k} = x_{ik}.
\end{align*}
\]

(Recall that \( b'(\theta_i) = \mu_i \) and \( a_i(\varphi) b''(\theta_i) = \text{Var}(Y_{ij}) \) by the properties of the standard exponential family.)

(ii) Derive \( \frac{\partial^2 Q}{\partial \beta_k \partial \beta_\ell} \):

\[
\begin{align*}
\frac{\partial^2 Q}{\partial \beta_k \partial \beta_\ell} & = \frac{1}{2} \sum_{i=1}^{m} \sqrt{n_i} \sum_{t} (N_{it})^{1/2} \left\{ \frac{\partial}{\partial \beta_\ell} \left[ f_{\theta_i, \varphi}^{1/2}(t) \cdot \frac{t - \mu_i}{\text{Var}(Y_{ij})} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_{ik} \right] \right\} \\
& = \frac{1}{2} \sum_{i=1}^{m} \sqrt{n_i} \sum_{t} (N_{it})^{1/2} \left\{ \left[ \frac{\partial}{\partial \beta_\ell} f_{\theta_i, \varphi}^{1/2}(t) \right] \cdot \frac{t - \mu_i}{\text{Var}(Y_{ij})} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_{ik} \\
& \quad + f_{\theta_i, \varphi}^{1/2}(t) \cdot \left[ \frac{\partial}{\partial \beta_\ell} \left( \frac{t - \mu_i}{\text{Var}(Y_{ij})} \right) \right] \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_{ik} \\
& \quad + f_{\theta_i, \varphi}^{1/2}(t) \cdot \frac{t - \mu_i}{\text{Var}(Y_{ij})} \cdot \left[ \frac{\partial}{\partial \beta_\ell} \left( \frac{\partial \mu_i}{\partial \eta_i} \right) \right] \cdot x_{ik} \right\} \\
& = \frac{1}{2} \sum_{i=1}^{m} \sqrt{n_i} \sum_{t} (N_{it})^{1/2} \left\{ \frac{1}{2} f_{\theta_i, \varphi}^{1/2}(t) \cdot \left[ \frac{t - \mu_i}{\text{Var}(Y_{ij})} \right]^2 \cdot \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \cdot x_{ik} \cdot x_{i\ell} \\
& \quad - f_{\theta_i, \varphi}^{1/2}(t) \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \cdot x_{ik} \cdot x_{i\ell} \\
& \quad + f_{\theta_i, \varphi}^{1/2}(t) \cdot (t - \mu_i) \cdot \frac{1}{a_i(\varphi)} \cdot \frac{\partial^2 \theta_i}{\partial \beta_\ell^2} \cdot \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \cdot x_{ik} \cdot x_{i\ell} \\
& \quad + f_{\theta_i, \varphi}^{1/2}(t) \cdot \frac{t - \mu_i}{\text{Var}(Y_{ij})} \cdot \frac{\partial^2 \mu_i}{\partial \eta_i^2} \cdot x_{ik} \cdot x_{i\ell} \right\} \\
\text{(3.3)}
\end{align*}
\]
because

$$
\frac{\partial}{\partial \beta_t} f_{\theta_1, \varphi}(t) = \frac{1}{2} f_{\theta_1, \varphi}(t) \cdot \frac{t - \mu_i}{\text{Var}(Y_{ij})} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_{it},
$$

$$
\frac{\partial}{\partial \beta_t} (t - \mu_i) = -\frac{\partial}{\partial \beta_t} \mu_i = -\frac{\partial}{\partial \beta_t} \eta_i \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_{it},
$$

\[
\frac{\partial}{\partial \beta_t} \left( \frac{1}{\text{Var}(Y_{ij})} \right) = \frac{\partial}{\partial \beta_t} \left( \frac{1}{a_i(\varphi) b''(\theta_i)} \right) = \frac{1}{a_i(\varphi)} \left[ \frac{\partial}{\partial \beta_t} \left( \frac{\partial \theta_i}{\partial \mu_i} \right) \right] \\
= \frac{1}{a_i(\varphi)} \left[ \frac{\partial \eta_i}{\partial \beta_t} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \right] \\
= \frac{1}{a_i(\varphi)} \left[ \frac{\partial^2 \theta_i}{\partial \mu_i^2} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_{it} \right],
\]

and

\[
\frac{\partial}{\partial \beta_t} \left( \frac{\partial \mu_i}{\partial \eta_i} \right) = \frac{\partial}{\partial \eta_i} \left( \frac{\partial \mu_i}{\partial \eta_i} \right) = \frac{\partial^2 \mu_i}{\partial \eta_i^2} \cdot x_{it}.
\]

(iii) Derive \( E \left[ -\frac{\partial^2 Q}{\partial \beta_k \partial \beta_t} \right] \):

Taking the negative expected value of every term in Equation (3.3) and rearranging,

\[
[J(\beta)]_{kl} := E \left[ -\frac{\partial^2 Q}{\partial \beta_k \partial \beta_t} \right] \\
= \frac{1}{2} \sum_{i=1}^{m} \sqrt{n_i} \sum_{t} E \left[ (N_{it})^{1/2} \right] \left\{ f_{\theta_1, \varphi}(t) \cdot \frac{1}{\text{Var}(Y_{ij})} \cdot \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \cdot x_{ik} \cdot x_{it} \right. \\
- f_{\theta_1, \varphi}(t) \cdot (t - \mu_i) \cdot \frac{1}{a_i(\varphi)} \cdot \frac{\partial^2 \theta_i}{\partial \mu_i^2} \cdot \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \cdot x_{ik} \cdot x_{it} \\
- f_{\theta_1, \varphi}(t) \cdot \frac{t - \mu_i}{\text{Var}(Y_{ij})} \cdot \frac{\partial^2 \mu_i}{\partial \eta_i^2} \cdot x_{ik} \cdot x_{it} \\
\left. - \frac{1}{2} f_{\theta_1, \varphi}(t) \cdot \left[ \frac{t - \mu_i}{\text{Var}(Y_{ij})} \right] \cdot \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \cdot x_{ik} \cdot x_{it} \right\},
\]

where

\[
E \left[ (N_{it})^{1/2} \right] = \sum_{n_{it}=0}^{n_i} (n_{it})^{1/2} \left( \binom{n_i}{n_{it}} \right) \left[ f_{\theta_1, \varphi}(t) \right]^{n_{it}} \left[ 1 - f_{\theta_1, \varphi}(t) \right]^{n_i - n_{it}}.
\]

Note that when each value of \( t \) is assigned, only \( (N_{it})^{1/2} \) is random so only its expected value needs to be computed. \( N_{it} \) depends on data but \( E \left[ (N_{it})^{1/2} \right] \) does not.
(iv) Derive a formula for Equation (3.1):

First, assume that $a_i(\varphi) = a(\varphi)$ for all $i$.

(a) Multiplying Equation (3.4) through by $2 \cdot a(\varphi)$ and pulling out the factor $\frac{1}{b'(\theta_i)} \cdot (\frac{\partial \mu_i}{\partial \mu_i})^2 \cdot x_{ik} \cdot x_{it}$, one can get

\[
2 \cdot a(\varphi) \cdot [J(\beta)]_{kt} = \sum_{i=1}^{m} \sqrt{n_i} \left\{ \sum_{t} E \left[ (N_{it})^{1/2} \right] \left[ f_{\theta_i,\varphi}^{1/2}(t) \right. \\
- f_{\theta_i,\varphi}^{1/2}(t) \cdot (t - \mu_i) \cdot b''(\theta_i) \cdot \frac{\partial^2 \theta_i}{\partial \mu_i^2} \\
- f_{\theta_i,\varphi}^{1/2}(t) \cdot (t - \mu_i) \cdot \frac{\partial^2 \mu_i}{\partial \eta_i^2} \cdot \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 \\
- \frac{1}{2} f_{\theta_i,\varphi}^{1/2}(t) \cdot (t - \mu_i)^2 \cdot \frac{1}{\text{Var}(Y_{ij})} \right] \cdot \frac{1}{b''(\theta_i)} \cdot \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \cdot x_{ik} \cdot x_{it}. \\
\]

Define $\rho_i = \rho(\theta_i, \eta_i, \mu_i)$ by

\[
\rho_i := \left\{ \sum_{t} E \left[ (N_{it})^{1/2} \right] \left[ f_{\theta_i,\varphi}^{1/2}(t) \right. \\
- f_{\theta_i,\varphi}^{1/2}(t) \cdot (t - \mu_i) \cdot b''(\theta_i) \cdot \frac{\partial^2 \theta_i}{\partial \mu_i^2} \\
- f_{\theta_i,\varphi}^{1/2}(t) \cdot (t - \mu_i) \cdot \frac{\partial^2 \mu_i}{\partial \eta_i^2} \cdot \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 \\
- \frac{1}{2} f_{\theta_i,\varphi}^{1/2}(t) \cdot (t - \mu_i)^2 \cdot \frac{1}{\text{Var}(Y_{ij})} \right] \right\}. \\
\]

which is no longer a function of data $y_i$. Let

\[
w_i = \sqrt{n_i} \cdot \rho_i \cdot \frac{1}{b''(\theta_i)} \cdot \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2
\]

be the diagonal element of the $m \times m$ diagonal weight matrix $W$. Then,

\[
2 \cdot a(\varphi) \cdot [J(\beta)] = X^\top W X.
\]

(b) Multiplying Equation (3.2) through by $2 \cdot a(\varphi)$ and reformulating, one has

\[
2 \cdot a(\varphi) \cdot \left[ \frac{\partial Q}{\partial \beta_k} \right] = \sum_{i=1}^{m} \sqrt{n_i} \sum_{t} (N_{it})^{1/2} \left\{ f_{\theta_i,\varphi}^{1/2}(t) \cdot \frac{t - \mu_i}{b''(\theta_i)} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_{ik} \right\}
\]
\[
\begin{align*}
&= \sum_{i=1}^{m} \left\{ \sqrt{n_i} \cdot \rho_i \cdot \frac{1}{b''(\theta_i)} \cdot \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \right\} \sum_{t} \left( N_{it} \right)^{1/2} \cdot f_{\theta_i, \varphi}^{1/2}(t) \cdot \frac{t - \mu_i}{\rho_i} \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ik} \\
&= \sum_{i=1}^{m} w_i \sum_{t} \left( N_{it} \right)^{1/2} \cdot f_{\theta_i, \varphi}^{1/2}(t) \cdot \frac{t - \mu_i}{\rho_i} \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ik} \\
&= \sum_{i=1}^{m} w_i \sum_{\tilde{g}_{n_i}(t) \neq 0} \left( N_{it} \right)^{1/2} \cdot f_{\theta_i, \varphi}^{1/2}(t) \cdot \frac{t - \mu_i}{\rho_i} \cdot \frac{\partial \eta_i}{\partial \mu_i} \cdot x_{ik}.
\end{align*}
\]

Then,
\[
2 \cdot a(\varphi) \cdot \left[ \frac{\partial Q}{\partial \beta} \right] = X^T W \xi,
\]

where
\[
\xi = \begin{bmatrix}
\xi_1 \\
\xi_2 \\
\vdots \\
\xi_m
\end{bmatrix} = \begin{bmatrix}
\sum_{\tilde{g}_{n_1}(t) \neq 0} \frac{\left( N_{1t} \right)^{1/2} \cdot f_{\theta_1, \varphi}^{1/2}(t) \cdot \left( t - \mu_1 \right)}{\rho_1} \cdot \frac{\partial \eta_1}{\partial \mu_1} \\
\sum_{\tilde{g}_{n_2}(t) \neq 0} \frac{\left( N_{2t} \right)^{1/2} \cdot f_{\theta_2, \varphi}^{1/2}(t) \cdot \left( t - \mu_2 \right)}{\rho_2} \cdot \frac{\partial \eta_2}{\partial \mu_2} \\
\vdots \\
\sum_{\tilde{g}_{n_m}(t) \neq 0} \frac{\left( N_{mt} \right)^{1/2} \cdot f_{\theta_m, \varphi}^{1/2}(t) \cdot \left( t - \mu_m \right)}{\rho_m} \cdot \frac{\partial \eta_m}{\partial \mu_m}
\end{bmatrix}
\]

(c) Thus, one can get the IRWLS formula such that

\[
2 \cdot a(\varphi) \cdot \left[ J(\beta^{(\text{old})}) \right] \beta^{(\text{new})} \approx 2 \cdot a(\varphi) \cdot \left[ J(\beta^{(\text{old})}) \right] \beta^{(\text{old})} + 2 \cdot a(\varphi) \cdot \left[ \frac{\partial Q}{\partial \beta} \right]_{\beta = b^{(\text{old})}}
\]

\[
\Leftrightarrow X^T W^{(\text{old})} X \beta^{(\text{new})} \approx X^T W^{(\text{old})} X \beta^{(\text{old})} + X^T W^{(\text{old})} \xi^{(\text{old})}
\]

\[
\Leftrightarrow X^T W^{(\text{old})} X \beta^{(\text{new})} \approx X^T W^{(\text{old})} \xi^{(\text{old})},
\]

where
\[
z_i = \eta_i + \xi_i = \eta_i + \sum_{\tilde{g}_{n_i}(t) \neq 0} \frac{\left( N_{it} \right)^{1/2} \cdot f_{\theta_i, \varphi}^{1/2}(t) \cdot \left( t - \mu_i \right)}{\rho_i} \cdot \frac{\partial \eta_i}{\partial \mu_i}.
\]

Finally, consider the case where $a_i(\varphi) = \frac{\varphi}{\nu_i}$ varies from subpopulation to subpopulation. Very little change needs to be made. All that has to be done is to substitute $2 \cdot \varphi \cdot [J(\beta)]$ and $2 \cdot \varphi \cdot \left[ \frac{\partial Q}{\partial \beta} \right]$ for $2 \cdot a(\varphi) \cdot [J(\beta)]$ and $2 \cdot a(\varphi) \cdot \left[ \frac{\partial Q}{\partial \beta} \right]$, respectively, in the derivation. Then, one gets

\[
w_i = \sqrt{n_i} \cdot \rho_i \cdot \frac{\nu_i}{b''(\theta_i)} \cdot \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2.
\]
Notes:

1. Although \( Q(\theta, \varphi; y_1, y_2, \ldots, y_m) = \sum_{i=1}^{m} n_i \sum_t \left( \frac{N_{it}}{n_i} \right)^{1/2} f_{\theta, \varphi}(t) \) can be reduced to \( \sum_{i=1}^{m} n_i \sum_{g_{n_i}(t) \neq 0} \left( \frac{N_{it}}{n_i} \right)^{1/2} f_{\theta, \varphi}(t) \) from the beginning of the derivation, using the latter makes the derivation more complicated because the index of the second summation depends on data.

2. It is not difficult to evaluate \( \rho_i \) because

\[
E \left[ (N_{it})^{1/2} \cdot f_{\theta, \varphi}^{1/2}(t) \right] \rightarrow 0
\]

\[
E \left[ (N_{it})^{1/2} \cdot f_{\theta, \varphi}^{1/2}(t) \cdot (t - \mu_i) \right] \rightarrow 0
\]

\[
E \left[ (N_{it})^{1/2} \cdot f_{\theta, \varphi}^{1/2}(t) \cdot (t - \mu_i)^2 \right] \rightarrow 0
\]

as \( t \to \infty \).

3. When the canonical link is used (i.e., \( \eta = \theta \)), the middle two terms of \( \rho_i \) cancel out because \( b''(\theta_i) \cdot \frac{\partial^2 \theta_i}{\partial \eta_i^2} = -\frac{\partial^2 \mu_i}{\partial \eta_i^2} \cdot \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 \). Therefore, \( \rho_i \) is simplified to

\[
\rho_i = \left\{ \sum_t E \left[ (N_{it})^{1/2} \cdot f_{\theta, \varphi}^{1/2}(t) \cdot \left[ 1 - \frac{(t - \mu_i)^2}{2 \cdot \text{Var}(Y_{ij})} \right] \right] \right\}
\]

which is very easy to compute. For both Poisson and binomial GLMs studied in this dissertation, canonical links are used, which makes the IRWLS algorithm very easy to apply.

4. For discrete GLMs with one observation per cell,

\[
E \left[ (N_{it})^{1/2} \right] = \sum_{n_{it}=0}^{1} \left( n_{it} \right)^{1/2} \cdot \left( \frac{1}{n_{it}} \right) \cdot \left[ f_{\theta, \varphi}(t) \right]^{n_{it}} \cdot \left[ 1 - f_{\theta, \varphi}(t) \right]^{1-n_{it}} = f_{\theta, \varphi}(t),
\]

so

\[
\rho_i = \left\{ \sum_t f_{\theta, \varphi}^{3/2}(t) \cdot \left(-f_{\theta, \varphi}^{3/2}(t) \cdot (t - \mu_i) \cdot b''(\theta_i) \cdot \frac{\partial^2 \theta_i}{\partial \mu_i^2}\right) \right\}
\]
which can be reduced to

\[-f_{\theta_i, \psi}^{3/2}(t) \cdot (t - \mu_i) \cdot \frac{\partial^2 \mu_i}{\partial \eta_i^2} \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 \]

\[-\frac{1}{2} f_{\theta_i, \psi}^{3/2}(t) \cdot (t - \mu_i)^2 \cdot \frac{1}{\text{Var}(Y_i)} \]

which can be reduced to

\[\rho_i = \sum_t f_{\theta_i, \psi}^{3/2}(t) \left[ 1 - \frac{(t - \mu_i)^2}{2 \cdot \text{Var}(Y_i)} \right] \]

when the canonical link is used. Also,

\[w_i = \rho_i \cdot \frac{1}{b''(\theta_i)} \cdot \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \]

and \(N_{it}\) is nonzero and equal to 1 only when \(t = y_i\) so

\[z_i = \eta_i + \left[ f_{\theta_i, \psi}^{1/2}(y_i) \cdot (y_i - \mu_i) \cdot \frac{\partial \eta_i}{\partial \mu_i} \right]. \]

The IRWLS algorithm derived above for GMHD estimation is used throughout this dissertation to compute robust starting values of \(\beta\) for Powell’s method in computing GMPHD estimates.

### 3.2.2 Robust Starting Values for GMPHDE in Discrete GLMs

McCullagh and Nelder (1989, page 41) suggest a simple starting procedure to get the IRWLS iteration under way. It consists of using the data themselves as the first estimate of \(\mu^{(\text{old})}\) and from this deriving \(W^{(\text{old})}\) and \(z^{(\text{old})}\). Then, the first estimate \(b^{(\text{new})}\) of \(\beta\) (named the one-step QL estimate of \(\beta\)) may be computed. Some adjustments may be required to the data to prevent any \(\mu_i\) taking a value that would cause inappropriate evaluation of the link function. For example, in Poisson GLMs, \(\mu_i\) cannot be equal to 0 and in binomial GLMs, \(\mu_i = N_i \cdot \pi_i\) cannot be exactly equal to 0 or \(N_i\) because \(\log(0)\) and \(\log\left(\frac{1}{1-0}\right)\) or \(\log\left(\frac{1}{1-1}\right)\) are undefined.

In our simulation, for computing QL estimates for Poisson GLMs, \(\mu_i = y_i = 0\) is replaced by \(\mu_i = \frac{1}{1\cdot1}\). For binomial GLMs, \(\pi_i = \frac{0}{N_i} = 0\) is replaced by \(\pi_i = \frac{0.5}{10+N_i}\) and \(\pi_i = \frac{N_i}{N_i} = 1\).
is replaced by $\pi_i = \frac{9.5 + N_i}{10 + N_i}$. These are Bayes estimates of $\mu_i$ and $\pi_i$ using Gamma and beta conjugate prior distributions (with arbitrary hyper-parameters) for $\mu_i$ and $\pi_i$, respectively. When there is more than one observation per subpopulation, $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}$ is used as an initial estimate of $\mu_i$.

The same strategy applies to generating the first GMHD estimate $b^{(\text{new})}$ of $\beta$. Using the data themselves (to be adjusted if necessary) as the first estimate of $\mu^{(\text{old})}$ and from it deriving $W^{(\text{old})}$ and $z^{(\text{old})}$, the first GMHD estimate of $\beta$ (named the one-step GMHD estimate of $\beta$) can be obtained by implementing the IRWLS algorithm described in Section 3.2.1 once. This one-step GMHD estimate of $\beta$, which is a much more robust starting estimate of $\beta$ than McCullagh and Nelder’s one-step QL estimate, is then used to start the Powell optimization process.

Since the IRWLS algorithm for GMPHD estimation does not perform as well as Powell’s method, this algorithm is now only used to compute starting values for Powell’s method. However, finding a general fitting algorithm as an alternative to the generic optimization Powell’s method is still a goal. Therefore, in our future research, more time will be devoted to improving the performance of the IRWLS algorithm for GMPHD estimation.
4.1 Introduction

This chapter contains the results of extensive simulation to evaluate the performance of GMPHD estimates and the NPB-GPHDV tests for Poisson and binomial GLMs. The results for GMPHD estimation and for NPB-GPHDV tests have two main parts: the results obtained by using the true optimal $h$ (derived by simulation based on 1000 replications), and results obtained by using the PBoot-estimated optimal $h$'s. The purpose of including the results using the true optimal $h$ is to demonstrate the best performance of these estimation and testing procedures and show the difference of using PBoot optimal $h$'s.

Some preliminary simulation results for estimation and testing in Poisson GLMs are also given. These early results help one understand how our methods were developed and why they are chosen over other alternatives. For example, GMPHD estimates obtained using the NPBoot estimated optimal $h$'s, and levels of PB-GPHDV tests (using the true optimal $h$) are included in the results for Poisson GLMs.

Simulation results for binomial GLMs are also used to demonstrate the utility of the GMPHD estimation and the NPB-GPHDV testing procedures. The results suggest that our methods perform as well for the binomial GLMs as they do for the Poisson GLMs.

The rest of this chapter is organized as follows. Section 4.2 contains the technical details of the simulations and results for Poisson GLMs. A detailed description of the Poisson models (without and with contamination) used in the simulation for estimation and testing is given in Section 4.2.1. Estimation results are presented in Section 4.2.2, while testing results are included in Section 4.2.3, in which Section 4.2.3.1 has the levels of NPB-GPHDV tests obtained by using the true optimal $h$ and Section 4.2.3.2 has those obtained by using PBoot optimal $h$'s. A description of the model and simulation results for binomial GLMs
are presented in Section 4.3. The model is specified in Section 4.3.1; estimation results are given in Section 4.3.2; and testing results are given in Section 4.3.3. Section 4.3.3 comprises two subsections. The first contains simulation results for NPB-GPHDV tests using the true optimal $h$ and the second contains simulation results for the tests using PBoot optimal $h$'s.

Discussion of power in testing situations is complicated because of the interaction between the level and power of a test. If level is not controlled, one can always make the power arbitrarily large. What happens in contaminated data situations for GLMs is that the level of the GLR test is typically quite different from (usually much larger than) the nominal level, but the power is very high. We had trouble finding a model that would allow us to control the level of the GLR test, even under contamination, so that we could evaluate the power of the test properly. Consequently, no simulation results for the power of NPB-GPHDV tests are presented.
4.2 Poisson GLMs

4.2.1 Simulation Parameters

Three Poisson GLMs, representing cases with different sizes of the Poisson means, were designated for the simulation. Each of the three models uses the same design matrix

\[
X = \begin{pmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & 1 & -1 \\
1 & 1 & -1 & 1 \\
1 & 1 & -1 & -1 \\
1 & -1 & 1 & 1 \\
1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1 \\
1 & -1 & -1 & -1 \\
1 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 \\
1 & -1 & -1 & 0 \\
1 & -1 & 0 & -1 \\
1 & -1 & 0 & 0 \\
1 & 0 & -1 & -1 \\
1 & 0 & -1 & 0 \\
1 & 0 & 0 & -1
\end{pmatrix}
\]

Hence, the number of subpopulations, \( m \), used in the simulations is 21. The \( \beta \) values for three models are (0.9, 0.2, 0.35, 0.15)\(^T\), (2.37771, 0.07510, 0.07002, 0.04341)\(^T\), and (3.84921, -0.04330, -0.05530, -0.07004)\(^T\) yielding the averages of the \( \mu_i \)'s, 2.65155, 10.83681, and 47.15180, individually.

Two types of contamination are considered in this dissertation. Type I contamination is a contamination across all subpopulations, while Type II contamination randomly contaminates some subpopulations completely and leaves the rest of the subpopulations uncontaminated.

To be more specific, for Type I contamination, each observation \( Y_{ij} \) of the \( i^{th} \) subpopu-
The contamination comes from the mixture model

\[ f_{\mu_i(\beta),\alpha_i,d_i}(y_{ij}) = (1 - \alpha_i)f_{\mu_i(\beta)}(y_{ij}) + \alpha_i \Delta z_i(y_{ij}), \]

where \( \alpha_i \in [0, 1) \) is the contamination factor of the \( i^{th} \) subpopulation, and the second component \( \Delta z_i \) is a point mass contamination which puts probability 1 on the point \( z_i = \mu_i + d_i\sigma_i \) rounded to the closest integer in the support of the model distribution, a point that is approximately \( d_i \) standard deviations larger than the \( i^{th} \) mean. Here, \( \sigma_i \) is \( \sqrt{\mu_i} \) for Poisson GLMs and is \( \sqrt{N_i\pi_i(1 - \pi_i)} \) for binomial GLMs. In this dissertation, it is assumed that \( \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0.1 \) and that \( d_1 = d_2 = \ldots = d_m = 5 \). The total contamination factor \( \alpha = \sum_{i=1}^{m} \frac{\alpha_i n_i}{n} \) is therefore 0.1.

For Type II contamination, observations \( Y_{i1}, Y_{i2}, \ldots, Y_{in_i} \) of the \( i^{th} \) subpopulation have the joint probability distribution function

\[ f_{\mu_i(\beta),k,d_i}(y_{i1}, y_{i2}, \ldots, y_{in_i}) = \frac{m-k}{m} \left[ \prod_{j=1}^{n_i} f_{\mu_i(\beta)}(y_{ij}) \right] + \frac{k}{m} \left[ \prod_{j=1}^{n_i} \Delta z_i(y_{ij}) \right], \]

where \( k \in \{0, 1, 2, \ldots, m\} \), \( \frac{k}{m} \) is the probability of each subpopulation to be completely contaminated by the point mass \( z_i \) (defined above) and \( \frac{m-k}{m} \) is that to be completely uncontaminated. It is easy to verify that the total contamination factor is \( \frac{k}{m} \). In this dissertation \( k = 2 \), so the total contamination factor is \( \frac{2}{21} \approx 0.1 \).

In the simulation for testing, the Poisson GLMs used for null models, either no contamination or contamination models, are the same ones used for estimation described above. For comparing two nested models, the test of interest is \( H_0 : \beta_4 = 0 \) versus \( H_1 : \beta_4 \neq 0 \). The design matrices used are \( X_0 \) which is the \( X \) given in Equation (4.1), and \( X_1 \) which is
given by

\[
X_1 = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & -1 & 0 \\
1 & 1 & -1 & 1 & -1 \\
1 & 1 & -1 & -1 & 1 \\
1 & -1 & 1 & 1 & 0 \\
1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & -1 & 0 \\
1 & 1 & 1 & 0 & -1 \\
1 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & -1 \\
1 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & -1 \\
1 & -1 & -1 & 0 & 1 \\
1 & -1 & 0 & -1 & 0 \\
1 & -1 & 0 & 0 & -1 \\
1 & 0 & -1 & -1 & 1 \\
1 & 0 & -1 & 0 & 0 \\
1 & 0 & 0 & -1 & -1 
\end{pmatrix}
\]

(4.4)

4.2.2 GMPHD Estimation Results

This section contains the estimation results for Poisson GLMs including those of our early work in developing GMPHD estimation, especially for the study of choosing appropriate penalty weights, $h$.

In Figure 4.1, MPHD estimation is implemented for iid Poisson models with mean values of 2, 10, and 50 for samples of sizes from 2 to 1000. All results for iid models are derived according to 5000 replications. Optimal values of $h$, which yield estimates that have the smallest MSE, are plotted against sample sizes in each graph. The three graphs on the right present the complete results for all sample sizes in the study while the three graphs on the left emphasize the behavior of the estimators for small samples. It appears that the optimal MPHD estimates are obtained at values of $h$ that are much smaller than 0.5 when sample sizes are very small and at values of $h$ around 0.5 only when the sample sizes are reasonably large. It also shows that the larger the Poisson mean, the smaller (more negative) the value
Figure 4.1. Plots of penalty weights $h$ where the optimal MPHD estimators occur (according to MSE) vs. sample sizes $n$ for iid Poisson models.
of \( h \) is needed for MPHD estimators to be optimal at very small samples and the larger the sample size is needed for optimal value of \( h \) to approach 0.5.

Figure 4.2 is analogous to Figure 4.1 for GMPHD estimation in Poisson GLMs with values of \( \mu_i \)'s around 2, 10, and 50, for equal subsample sizes of 1, 2, \ldots , 1000, in which the optimal \( h \) is associated with the smallest total MSE of the \( \hat{\mu}_i \)'s, \( \text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2} \).

All simulations for GLMs are based on 1000 replications. As with MPHD estimation, it shows that the optimal values of \( h \) depend not only on sample sizes but also on the magnitudes of the means.

Figure 4.3 and 4.4 illustrate the behavior of the optimal values of \( h \) for GMPHD estimation under two different types of contamination models for Poisson GLMs with small subsamples of sizes from 1 to 20. One can observe from Figures 4.3 and 4.4 that the pattern of the optimal values of \( h \) changes significantly from that in Figure 4.2 because of the contamination. Type II contamination, especially, causes instability in values of optimal \( h \). In comparing Figures 4.3 and 4.4 to the three graphs on the left side of Figure 4.2, an encouraging finding is that the optimal values of \( h \) under contamination for very small samples seem to be less negative than those under the model. This finding makes Step 1(a) in Algorithm 2.1 robust and applicable under contamination because the optimal values of \( h \) under contamination could usually be well covered by the set of values of \( h \), \( \{h\} \), used for uncontaminated data.

Figure 4.5 and Figure 4.6 compare the adjusted total MSEs, \( n_c \cdot \text{MSE}(\hat{\mu}) \), among the QL estimates (solid line), the optimal GMPHD estimates (dotted line), the GMPHD estimates obtained using \( h = 0.5 \) (short dashed line), and the ordinary GMHD estimates (long dashed line) for the same GLMs used in Figure 4.2, where \( n_c \) denotes the common sample size of all subpopulations. Figure 4.6 establishes the asymptotic behavior of these estimators. It reveals the asymptotic equivalency of the GMPHD estimators and the QL estimators under the model. Figure 4.5 zooms in on the results for the cases with small subsample sizes (1 to 20). It shows how well the penalty aspect of GMPHD estimation improves the small
Figure 4.2. Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$) vs. common sample sizes $n_c$ for Poisson GLMs. (No contamination)
Optimal $h$ for GMPHD Estimation
Type I Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure 4.3. Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i}$) vs. common sample sizes $n_c$ for Poisson GLMs. (Type I contamination)
Figure 4.4. Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$) vs. common sample sizes $n_c$ for Poisson GLMs. (Type II contamination)
Figure 4.5. Plots of adjusted total MSEs, $n_c \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$, of the QL estimates, the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, and the ordinary GMHD estimates vs. the common sample sizes $n_c$ (of 1 to 20) for Poisson GLMs. (No contamination)
Estimation Results
No Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure 4.6. Plots of adjusted total MSEs, \( n_c \sum_{i=1}^{m} \frac{MSE(\hat{\mu}_i)}{\mu_i^2} \), of the QL estimates, the optimal GMPHD estimates, the GMPHD estimates using \( h = 0.5 \), and the ordinary GMHD estimates vs. the common sample sizes \( n_c \) (of 1 to 1000) for Poisson GLMs. (No contamination)
sample performance of the GMHD estimators, and that the estimates associated with the fixed penalty factor \( h = 0.5 \) may not always be optimal in the class. It also shows that for small Poisson means and for very small subsamples (of size 5 or less, for example), the optimal GMPHD estimates can be very competitive to QL estimates, while for large Poisson means, the optimal GMPHD estimators in the class seem to be less efficient than the QL estimators.

In Figures 4.7, 4.8, and 4.9, the GMPHD estimates obtained using PBoot estimated optimal values of \( h \) (octagon) and those obtained using NPBoot estimated optimal values of \( h \) (filled diamond) are compared to the true optimal GMPHD estimates (dotted line). The results of QL estimation (solid line), of GMPHD estimation obtained using \( h = 0.5 \) (short dashed line), and of ordinary GMHD estimation obtained using \( h = 1.0 \) (long dashed line) are also given for comparison purposes. Both the total MSEs and the adjusted total MSEs, \( n_c \times \text{MSE}(\hat{\mu}) \), of these estimates are compared over common subsample sizes of 1 to 20 under no contamination (Figures 4.7), Type I contamination (Figure 4.8), and Type II contamination (Figure 4.9). The QL estimation results are excluded from the three graphs on the right side of Figure 4.8 and of Figure 4.9 in order to see the differences among the GMPHD estimates more clearly. Figure 4.7 shows that the GMPHD estimates using PBoot \( h \)'s are very close to the GMPHD estimates using true optimal \( h \) values and are much more efficient than those using NPBoot \( h \)'s when data are not contaminated. The graphs on the left side of Figures 4.8 and of 4.9 indicate that the PBoot method still yields much more robust estimates compared to QL estimation but it is less robust than the NPBoot method, as shown in the right side of the graphs. One can also see that neither the PBoot method nor the NPBoot method performs very well for the case where GLMs have only one observation per cell when data are contaminated.
Figure 4.7. Plots of adjusted total MSEs, \( n_c \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2} \), of the QL estimates, the optimal GMPHD estimates, the GMPHD estimates using \( h = 0.5 \), the ordinary GMHD estimates, the GMPHD estimates using PBoot optimal \( h \), and the GMPHD estimates using NPBoot optimal \( h \) vs. the common sample sizes \( n_c \) for Poisson GLMs. (No contamination)
Estimation Results
Type I Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure 4.8. Plots of total MSEs, $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$, and the adjusted total MSEs, $n_c \cdot \text{MSE}(\hat{\mu})$, of the QL estimates (for plots on the left only), the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, the GMPHD estimates using PBoot optimal $h$, and the GMPHD estimates using NPBBoot optimal $h$ vs. the common sample sizes $n_c$ for Poisson GLMs. (Type I contamination)
Estimation Results
Type II Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure 4.9. Plots of total MSEs, $MSE(\hat{\mu}) = \sum_{i=1}^{n} \frac{MSE(\hat{\mu}_i)}{\mu_i^2}$, and the adjusted total MSEs, $n_c*MSE(\hat{\mu})$, of the QL estimates (for plots on the left only), the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, the GMPHD estimates using PBoot optimal $h$, and the GMPHD estimates using NPBoot optimal $h$ vs. the common sample sizes $n_c$ for Poisson GLMs. (Type II contamination)
4.2.3 NPB-GPHDV Test Results

4.2.3.1 Levels Obtained by Using the True Optimal $h$

The first part of Section 4.2.3 includes results for comparing the performance of all tests introduced in Chapter 2, in which levels of GPHDV, PB-GPHDV, and NPB-GPHDV tests are computed using the true optimal $h$. All tests are compared for Poisson GLMs with different sizes of $\mu_i$'s and with no contamination, Type I contamination, and Type II contamination, to illustrate the need for NPB-GPHDV tests.

The comparison is focused on the convergence of the observed levels of these tests to the 0.1 nominal level. Each observed level is based on a simulation of 1000 replications and is determined by the proportion of test statistics exceeding the upper 10% critical value of either an appropriate limiting $\chi^2$ distribution or a bootstrap build-up distribution. Assuming binomial rejection frequencies, the standard deviation of an estimated rejection proportion, $\hat{p}$, may be computed as \[ \left[ \frac{n(1-p)}{1000} \right]^{1/2} \] which can be no greater than \[ \left[ \frac{0.5 \times 0.5}{1000} \right]^{1/2} = 0.0158. \] For example, with nominal level 0.1, the standard deviation for estimating $p = 0.1$ is approximately 0.009.

Results for testing goodness-of-fit are given in Figures 4.10, 4.12, 4.14, and 4.16 while Figures 4.11, 4.13, 4.15, and 4.17 are results for comparing two nested models. Each figure contains three graphs for Poisson models with $\mu_i$'s approximately equal to 2, 10, and 50.

In Figures 4.10 and 4.11, observed levels of the GLR tests (solid line), the GPHDV tests using optimal $h$ (dotted line), the GPHDV tests using $h = 0.5$ (short dashed line), and the ordinary GHDV tests (long dashed line) are compared under the null model over common subsample sizes of 1 to 100 for testing goodness-of-fit and for comparing two nested models, individually. Note that, as described in Sections 2.5.1 and 2.5.2, the optimal $h$ used for GPHDV tests is obtained either according to the full model for goodness-of-fit or the larger model for comparing two nested models. It shows, as expected, that the penalty process does improve the rate of convergence of the observed levels resulted from the ordinary GHDV tests to the nominal level. For both testing for goodness-of-fit and for
Tests for Goodness-of-fit (Nominal Level: 0.1)
No Contamination
Poisson Means Approximately Equal to 2

Figure 4.10. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, and the ordinary GHDV tests vs. common sample sizes $n_c$ (of 1 to 100) for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)

No Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure 4.11. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, and the ordinary GHDV tests vs. common sample sizes $n_c$ (of 1 to 100) for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.1)
Tests for Goodness-of-fit (Nominal Level: 0.1)

No Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure 4.12. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)
No Contamination
Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure 4.13. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.1)
Tests for Goodness-of-fit (Nominal Level: 0.1)
Type I Contamination
Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure 4.14. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type I contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)

Type I Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure 4.15. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type I contamination; Nominal level 0.1)
Tests for Goodness-of-fit (Nominal Level: 0.1)
Type II Contamination
Poisson Means Approximately Equal to 2

Figure 4.16. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type II contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)
Type II Contamination
Poisson Means Approximately Equal to 2

Figure 4.17. Plots of observed levels of the GLR tests, the GPHDV tests using optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, the PB-GPHDV tests using optimal \( h \), and the NPB-GPHDV tests using optimal \( h \) vs. common sample sizes \( n_c \) for comparing two nested Poisson GLMs. (Type II contamination; Nominal level 0.1)
nested models, GPHDV tests appear to give observed levels that are much closer to those of the GLR tests, compared to those of the GHDV tests. However, one can see that, for cases with very few observations per subpopulation as in most GLMs, the regular GPHDV tests still do not converge fast enough to the GLR tests for the $\chi^2$ approximation to be usable. The GPHDV tests using optimal $h$, surprisingly, do not further improve the rate of convergence of the GPHDV tests using $h = 0.5$ to the GLR tests. Thus, for the GPHDV tests to be useful in small samples, one must not rely on the critical values based on the limiting $\chi^2$ distributions of the test statistics.

As with the GMPHD estimation, the sizes of the $\mu_i$'s appear to influence the GPHDV tests in the sense that larger sample sizes are needed for the observed levels to converge to the nominal level when the values of the $\mu_i$'s are larger.

In Figure 4.10, one may observe that the GLR test for goodness-of-fit gives less accurate levels for the case where a GLM has one observation per cell than those for other cases when $\mu_i$'s are approximately equal to 2. Recall from Note 1 of Section 2.4 that the GLR test for goodness-of-fit is more reliable in discrete data problems where the counts are large. This behavior is illustrated in Figure 4.10. As the sizes of the $\mu_i$'s increase, the GLR test gives more and more accurate results at $n_i = 1$.

In addition to the tests compared in Figures 4.10 and 4.11, the PB-GPHDV tests using optimal $h$ (octagons) and the NPB-GPHDV tests using optimal $h$ (filled diamonds) are also compared under no contamination (Figures 4.12 and 4.13), under the Type I contamination model (Figures 4.14 and 4.15), and under the Type II contamination model (Figures 4.16 and 4.17), but only for small common subsample sizes of 1 to 20.

When the null model is true, the superiority of the PB-GPHDV and the NPB-GPHDV tests over the GPHDV tests (including the GHDV tests) is clearly illustrated in Figures 4.12 and 4.13. Both PB-GPHDV and NPB-GPHDV tests give much more accurate observed levels than the GPHDV tests. One can see that the PB-GPHDV tests perform as well as the GLR tests, even for very small subsamples. Results of the NPB-GPHDV tests are
very competitive to those of the PB-GPHDV tests for most cases, but they seem to be less satisfying for the case of GLMs with only one observation per cell.

Both PB-GPHDV tests and NPB-GPHDV tests have one major advantage over the regular GPHDV tests and the ordinary GHDV tests in that their performance is not affected by the sizes of the $\mu_i$’s. As the sizes of the $\mu_i$’s increase, the PB-GPHDV tests and the NPB-GPHDV tests both give consistent results, while the GPHDV tests and the GHDV tests not only perform worse in small samples but also have slower rates of convergence to the nominal level. This behavior of the bootstrap GPHDV tests is very impressive considering that the GMPHD estimates on which the tests are based are influenced by the sizes of the $\mu_i$’s (see Figure 4.5).

When data have been contaminated according to the Type I contamination model, Figures 4.14 and 4.15 show the superior robustness of the NPB-GPHDV tests over all other tests studied in this chapter, including the PB-GPHDV tests. For different sizes of $\mu_i$’s and for both testing for goodness-of-fit and for comparing two nested models, the NPB-GPHDV tests consistently give very accurate levels for most cases except maybe for the case with subsample size of 1, which is a problem case for the NPB-GPHDV tests even when data are not contaminated.

Tests results for Type II contamination are contained in Figures 4.16 and 4.17. For testing goodness-of-fit, none of the tests introduced (including the NPB-GPHDV tests) are robust. Comparatively, the NPB-GPHDV tests seem give reasonably robust results for tests of comparing two nested models.

More tests results for nominal levels 0.05 and 0.01 are included in Appendix A. For nominal level 0.05, the results of the parametric bootstrap GPHDV tests are as accurate as those for nominal level 0.1. The tail behavior of the tests, where we have so little data to rely on, is displayed by using nominal level 0.01. The NPB-GPHDV tests perform convincingly well even for such extreme nominal level.

In conclusion of this section, the following statements are made. Despite the fact that
GHDV tests and GPHDV tests are asymptotically equivalent to the GLR tests at the null model, the $\chi^2$ approximation is not very good when they are applied to GLMs with very small subsample sizes. The PB-GPHDV tests provide very accurate significance levels not only asymptotically but also for very small samples, but are less robust than the NPB-GPHDV tests. Except for the case with common subsample size of 1, the NPB-GPHDV tests perform very well under the null model and are very robust against Type I contamination. For Type II contamination, the NPB-GPHDV tests are reasonably robust for comparing two competitive nested models while they too fail to produce robust results for goodness-of-fit. More research needs to be done for developing tests that are robust for GLMs with one observation per sub-population and for different types of contamination or model misspecification.

4.2.3.2 Levels Obtained by Using the PBoot Optimal $h$

In this second part of Section 4.2.3, the levels for GPHDV and NPB-GPHDV tests are computed using PBoot optimal $h$'s. The results for these simulations are graphically represented in Figures 4.18, 4.19, 4.20, 4.21, and 4.22, which are entirely analogous to Figures 4.12, 4.13, 4.14, 4.15, and 4.17 from the first part of Section 4.2.3. The results for goodness-of-fit under Type II contamination are not included in this part because NPB-GPHDV test fails for this case even when the true optimal $h$'s are used.

Because the PBoot-estimated optimal $h$'s are not robust, it is interesting to compare the results of the two parts and see how much difference is made by using PBoot optimal $h$'s. One can see that most of the results are very similar to those in Section 4.2.3.1 except for three cases.

The first two cases are the tests of goodness-of-fit for the Poisson models with means approximately equal to 10 and 50 under Type I contamination. By comparing Figure 4.20 to Figure 4.14, one can see that the levels of NPB-GPHDV test using PBoot optimal $h$'s converge slightly more slowly to the nominal level than those using the true optimal $h$. 
Tests for Goodness-of-fit (Nominal Level: 0.1)

No Contamination

Poisson Means Approximately Equal to 2

![Graph showing observed levels of GLR tests, GPHDV tests using PBoot optimal h, GPHDV tests using h = 0.5, ordinary GHDV tests, and NPB-GPHDV tests using PBoot optimal h vs. common sample sizes n_c for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)  
No Contamination  
Poisson Means Approximately Equal to 2

Figure 4.19. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.1)
Tests for Goodness-of-fit (Nominal Level: 0.1)

Type I Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure 4.20. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type I contamination; Nominal level 0.1)
Figure 4.21. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type I contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)

Type II Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure 4.22. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type II contamination; Nominal level 0.1)
The third case is the test for nested models for the Poisson model with means approximately equal to 2 under Type II contamination. For this case, the levels of the NPB-GPHDV test are less accurate than those using the true optimal $h$ in the sense that the NPB-GPHDV test using PBoot optimal $h$'s tends to reject the null hypothesis more often than those using the true optimal $h$. (See Figure 4.22 and Figure 4.17 for the comparison.) That is, for this particular case, the critical values computed by nonparametric bootstrap using PBoot optimal $h$'s are lower than those obtained using the true optimal $h$'s and certainly lower than the true critical values of the GPHDV test statistics.

Additional simulation results for nominal levels 0.05 and 0.01 are included in Appendix B.

4.3 Binomial GLMs

4.3.1 Simulation Parameters

The estimation and testing methods introduced in Chapter 2 are also applied to the binomial GLMs. Five binomial GLMs, covering some interesting cases, are considered in the simulation.

One question of interest is the impact of three different types of binomial GLMs: one, named the N model, in which the normal approximation is valid or nearly valid (largish $N_i$'s, moderate $\pi_i$'s), one, named the P model, in which the Poisson approximation is valid or nearly valid (largish $N_i$'s, smallish $\pi_i$'s), and one, named the B model, in which neither approximation is valid (smallish $N_i$'s, moderate $\pi_i$'s). Our study includes one N model, one P model, and three B models (denoted by B1, B2, and B3).

Another issue is the effect of the size of the $\mu_i$'s. To be comparable to the simulation for Poisson GLMs, B1, P, and N models were designed so that the values of $\mu_i$'s (where each $\mu_i = N_i \times \pi_i$) are approximately 2, 10, and 50. These three models all have equal $N_i$'s.

The third issue of interest is the influence of equal $N_i$'s versus (very) unequal $N_i$'s. Both B2 and B3 models were designated to have the same values of $\pi_i$'s, but B2 has equal $N_i$'s
and B3 has very unequal \( N_i \)'s, where the values of \( N_i \)'s were chosen so that B2 model has values of \( \mu_i \)'s about 10 and B3 model (with very different values of \( \mu_i \)'s) has the average of the \( \mu_i \)'s about 10. Doing such, besides studying the impact of the equal \( N_i \)'s versus unequal \( N_i \)'s, one also gets two more B models to observe and may compare them to the P model which also has \( \mu_i \)'s approximately 10.

Here are the details of the five binomial GLMs. Same design matrices used for Poisson GLMs are used for binomial GLMs. That is, for estimation \( X \) is given in Equation (4.1); for testing, \( X_0 \) and \( X_1 \) are given in Equations (4.1) and (4.4). Therefore, \( m = 21 \).

For B1 model, \( \beta = (0.12232, -0.05748, 0.33354, -0.15993)^T \) and \( N_i = 5 \) for all \( i \) giving the average of the \( \pi_i \)'s 0.52994 and the average of the \( \mu_i \)'s 2.64970.

For P model, \( \beta = (-2.89985, -0.07681, 0.18521, 0.05790)^T \) and \( N_i = 200 \) for all \( i \) giving the average of the \( \pi_i \)'s 0.05277 and the average of the \( \mu_i \)'s 10.55433.

For N model, \( \beta = (-0.13836, 0.17437, 0.04283, 0.04771)^T \) and \( N_i = 100 \) for all \( i \) giving the average of the \( \pi_i \)'s 0.46569 and the average of the \( \mu_i \)'s 46.56903.

For B2 model, \( \beta = (0.9, 0.2, 0.35, 0.15)^T \) and \( N_i = 15 \) for all \( i \) giving the average of the \( \pi_i \)'s 0.70457 and the average of the \( \mu_i \)'s 10.56853.

For B3 model, \( \beta = (0.9, 0.2, 0.35, 0.15)^T \) and \( (N_1, N_2, \ldots, N_m) = (24, 11, 18, 24, 24, 11, 12, 20, 16, 18, 3, 19, 19, 2, 11, 13, 12, 8, 33, 5, 12) \) giving the average of the \( \pi_i \)'s 0.70457 and the average of the \( \mu_i \)'s 10.56237.

4.3.2 GMPHD Estimation Results

This section gives extensive estimation results for binomial GLMs. They are presented in Figures 4.23, 4.24, 4.25, 4.26, 4.27, 4.28, 4.29, 4.30, 4.31, 4.32, 4.33, and 4.34. Because these results are essentially analogous to the estimation results for Poisson GLMs in Section 4.2.2, the detailed description of the figures in this section is omitted for simplicity whereas a few interesting findings are discussed.

The performance of the GMPHD estimation for P and N models is very much like that
Figure 4.23. Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$) vs. common sample sizes $n_c$ for binomial GLMs. (B1, P, and N models; No contamination)
Figure 4.24. Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$) vs. common sample sizes $n_c$ for binomial GLMs. (B2 and B3 models; No contamination)
Figure 4.25. Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\mu) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$) vs. common sample sizes $n_c$ for binomial GLMs. (B1, P, and N models; Type I contamination)
Optimal h for GMPHD Estimation
Type I Contamination
Binomial Means Approximately Equal to 10
B2 Model (N_i's are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (N_i's are Different)

Figure 4.26. Plots of penalty weights h where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}^{(i,h)})}{\hat{\mu}^{(i,h)}}$) vs. common sample sizes $n_c$ for binomial GLMs. (B2 and B3 models; Type I contamination)
Figure 4.27. Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$) vs. common sample sizes $n_c$ for binomial GLMs. (B1, P, and N models; Type II contamination)
Figure 4.28. Plots of penalty weights $h$ where the optimal GMPHD estimators occur (according to $\text{MSE}(\hat{\mu}) = \sum \text{MSE}(\hat{\mu}_i)$) vs. common sample sizes $n_c$ for binomial GLMs. (B2 and B3 models; Type II contamination)
Figure 4.29. Plots of adjusted total MSEs, \( n_c \times \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2} \), of the QL estimates, the optimal GMPHD estimates, the GMPHD estimates using \( h = 0.5 \), the ordinary GMHD estimates, and the GMPHD estimates using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for binomial GLMs. (B1, P, and N models; No contamination)
Figure 4.30. Plots of adjusted total MSEs, $n_e \ast \sum_{i=1}^{n} \frac{MSE(\hat{\mu}_i)}{\mu_i^2}$, of the QL estimates, the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, and the GMPHD estimates using PBoot optimal $h$ vs. common sample sizes $n_e$ for binomial GLMs. (B2 and B3 models; No contamination)
Estimation Results
Type I Contamination

Figure 4.31. Plots of total MSEs, $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\hat{\mu}_i}$, and the adjusted total MSEs, $n_c \times \text{MSE}(\hat{\mu})$, of the QL estimates (for plots on the left only), the optimal GMHPD estimates, the GMHPD estimates using $h = 0.5$, the ordinary GMHD estimates, and the GMHPD estimates using PBoot optimal $h$ vs. common sample sizes $n_c$ for binomial GLMs. (B1, P, and N models; Type I contamination)
Figure 4.32. Plots of total MSEs, $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$, and the adjusted total MSEs, $n_e \text{MSE}(\hat{\mu})$, of the QL estimates (for plots on the left only), the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, and the GMPHD estimates using PBoot optimal $h$ vs. common sample sizes $n_e$ for binomial GLMs. (B2 and B3 models; Type I contamination)
Figure 4.33. Plots of total MSEs, $\text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\mu_i^2}$, and the adjusted total MSEs, $n_c \times \text{MSE}(\hat{\mu})$, of the QL estimates (for plots on the left only), the optimal GMPHD estimates, the GMPHD estimates using $h = 0.5$, the ordinary GMHD estimates, and the GMPHD estimates using PBoot optimal $h$ vs. common sample sizes $n_c$ for binomial GLMs. (B1, P, and N models; Type II contamination)
Estimation Results
Type II Contamination

Figure 4.34. Plots of total MSEs, \( \text{MSE}(\hat{\mu}) = \sum_{i=1}^{m} \frac{\text{MSE}(\hat{\mu}_i)}{\nu_i^2} \), and the adjusted total MSEs, \( n_c \times \text{MSE}(\hat{\mu}) \), of the QL estimates (for plots on the left only), the optimal GMPHD estimates, the GMHD estimates using \( h = 0.5 \), the ordinary GMHD estimates, and the GMPHD estimates using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for binomial GLMs. (B2 and B3 models; Type II contamination)
for Poisson GLMs with means around 10 and 50, respectively. Nonetheless, the results for
the three B models under two types of contamination suggest that the method of GMPHD
estimation is not robust! Figures 4.31, 4.32, 4.33, and 4.34 show that, under contamination,
the results for the B1 model are very different from those for Poisson model with means
around 2 and the results for B2 and B3 models are nothing like those for Poisson GLM
with means around 10. This finding implies that, in addition to the sample size and the size
of the $\mu_i$'s, there is another factor that affects the GMPHD estimation in binomial GLMs.
Due to the way we compute the contaminated data, $z_i$'s, this factor is the size of the $N_i$'s.

Recall that the point mass $z_i$ is set to be the nearest integer in $[0, N_i]$ of the value
$N_i \cdot \pi_i + 5 \cdot \sqrt{N_i \cdot \pi_i \cdot (1 - \pi_i)}$ in our simulation for binomial GLMs. Because B models
usually have small $N_i$'s and moderate $\pi_i$'s, the value of $z_i$ is often truncated to $N_i$, which is
usually between 2 and 3 (not 5) standard deviations larger than the $\mu_i = N_i \cdot \pi_i$. Such con­
taminated data are usually the worst kind of contamination for a typical Hellinger method.
Section 5.3.2 explains this concept in detail by means of the $\alpha$-influence function. The
$\alpha$-influence curve of B1 model in Figure 5.2 and those of B2 and B3 models in Figure 5.3
illustrate the idea clearly.

For the issue of equal $N_i$ versus unequal $N_i$'s, comparing the results of B2 model and of
B3 model in Figures 4.24, 4.26, 4.28, 4.30, 4.32, and 4.34, there is no significant difference
between the results of the two models.

4.3.3 NPB-GPHDV Test Results

4.3.3.1 Levels Obtained by Using the True Optimal $h$

The first part of Section 4.3.3 contains extensive simulation results, which are analogous
to the results for Poisson GLMs in Section 4.2.3.1, for binomial GLMs. In these simulations,
levels of GPHDV test and NPB-GPHDV test are computed using the true optimal $h$.
Figures 4.35, 4.36, 4.37, 4.38, 4.39, 4.40, 4.41, 4.42, 4.43, 4.44, 4.45, and 4.46 contains
these results. Again, for simplicity, we omit the detailed description of the figures. For a
Tests for Goodness-of-fit (Nominal Level: 0.1)
No Contamination
Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure 4.35. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.1)
Figure 4.36. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)

No Contamination

Binomial Means Approximately Equal to 2

B1 Model

Binomial Means Approximately Equal to 10

P Model

Binomial Means Approximately Equal to 50

N Model

Figure 4.37. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)
No Contamination
Binomial Means Approximately Equal to 10
B2 Model (N's are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (N's are Different)

Figure 4.38. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$,
the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests
using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs.
(B2 and B3 models; No contamination; Nominal level 0.1)
Tests for Goodness-of-fit (Nominal Level: 0.1)
Type I Contamination
Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure 4.39. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.1)
Tests for Goodness-of-fit (Nominal Level: 0.1)
Type I Contamination
Binomial Means Approximately Equal to 10
B2 Model (N's are all Equal)

Figure 4.40. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)
Type I Contamination

Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure 4.41. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)
Type I Contamination

Binomial Means Approximately Equal to 10
B2 Model (Ni’s are all Equal)

Figure 4.42. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.1)
Tests for Goodness-of-fit (Nominal Level: 0.1)
Type II Contamination

Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure 4.43. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.1)
Figure 4.44. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.1)
Figure 4.45. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.1)
Figure 4.46. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.1)
description of similar figures, see Section 4.2.3.

In general, the simulation results suggest that the tests behave in a similar way for binomial GLMs and for Poisson GLMs. For example, (1) the one observation per cell case is still problematic, (2) no tests for goodness-of-fit are robust against Type II contamination (Figures 4.43 and 4.44), and (3) except for the \( n_i = 1 \) cases, NPB-GPHDV tests for goodness-of-fit and for nested model are very robust against Type I contamination (Figures 4.39, 4.40, 4.41, and 4.42). There are two exceptions. First, for the test for goodness-of-fit of binomial B1 model under no contamination, the levels of NPB-GPHDV test are not very accurate. (See Figure 4.35.) It appears that the true critical values of the GPHDV test statistics are lower than the critical values computed by the nonparametric bootstrap. Second, in contrast to the results for Poisson GLMs, the test for comparing nested models does not always work well for binomial models under Type II contamination. Figures 4.45 and 4.46 show that no tests are robust for binomial B models against Type II contamination. Obviously, the lack of robustness of GMPHD estimation for B models is carried over to NPB-GPHDV tests.

Additional simulation results for nominal levels 0.05 and 0.01 are included in the Appendix C.

4.3.3.2 Levels Obtained by Using the PBoot Optimal \( h \)

The simulations that are reported in the second part of Section 4.3.3 are exactly analogous to those in Section 4.3.3.1 except that the levels of the GPHDV and the NPB-GPHDV tests were computed using PBoot optimal \( h \)'s. These simulation results are given in Figures 4.47, 4.48, 4.49, 4.50, 4.51, 4.52, 4.53, 4.54, 4.55, and 4.56. The results for goodness-of-fit under Type II contamination are again excluded because even when using the true optimal \( h \), the NPB-GPHDV test results are unsatisfactory.

The simulation results suggest that NPB-GPHDV tests using PBoot optimal \( h \)'s have levels almost identical to those obtained using the true optimal \( h \). Hence, by using PBoot
Tests for Goodness-of-fit (Nominal Level: 0.1)
No Contamination
Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure 4.47. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.1)
Tests for Goodness-of-fit (Nominal Level: 0.1)
No Contamination
Binomial Means Approximately Equal to 10
B2 Model (N's are all Equal)

Figure 4.48. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)  
No Contamination  
Binomial Means Approximately Equal to 2  
B1 Model

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Sample Sizes (n1=n2=...=nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLR</td>
<td>5, 10, 15, 20</td>
</tr>
<tr>
<td>GPHDV</td>
<td>5, 10, 15, 20</td>
</tr>
<tr>
<td>NPB-GPHDV</td>
<td>5, 10, 15, 20</td>
</tr>
</tbody>
</table>

Figure 4.49. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for comparing two nested binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)
No Contamination
Binomial Means Approximately Equal to 10
B2 Model (Ni’s are all Equal)

B3 Model (Ni’s are Different)

Figure 4.50. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.1)
Tests for Goodness-of-fit (Nominal Level: 0.1)
Type I Contamination
Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure 4.51. Plots of observed levels of the GLR tests, the GPHDV tests using PBootstrap optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBootstrap optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.1)
Tests for Goodness-of-fit (Nominal Level: 0.1)
Type I Contamination

Binomial Means Approximately Equal to 10
B2 Model (N's are all Equal)

Figure 4.52. Plots of observed levels of the GLR tests, the GPHDV tests using PBootstrap optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBootstrap optimal $h$ vs. common sample sizes $n_e$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.1)
Figure 4.53. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)
Type I Contamination

Binomial Means Approximately Equal to 10
B2 Model (Nis are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (Nis are Different)

Figure 4.54. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for comparing two nested binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)
Type II Contamination
Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure 4.55. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.1)
Tests for Nested Models (Nominal Level: 0.1)
Type II Contamination
Binomial Means Approximately Equal to 10
B2 Model (Ni’s are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (Ni’s are Different)

Figure 4.56. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.1)
optimal $h$'s for NPB-GPHDV test, one loses almost no information.

There is only one case for which results using PBoot optimal $h$'s are less accurate than those using the true optimal $h$. By comparing Figures 4.51 and 4.39, one can see that for the test of goodness-of-fit for $P$ model under Type I contamination, the levels of NPB-GPHDV test using PBoot optimal $h$'s converge more slowly to the nominal level than those obtained using the true optimal $h$. This outcome is actually very similar to one of the results for Poisson GLMs with means approximately equal to 10, as described in Section 4.2.3.2.

Additional simulation results for nominal levels 0.05 and 0.01 are included in the Appendix D.
CHAPTER 5

PROPERTIES OF GMPHD ESTIMATORS FOR DISCRETE GLMS

The aim of this chapter is to explore the asymptotic efficiency and robustness properties of the GMPHD estimators in GLMs. Because the influence of the penalty—which is a function of empty cells—vanishes as the total sample size tends to infinity, the GMPHD estimator is asymptotically equivalent to the GMHD estimator and hence shares the asymptotic properties of the latter. Thus, it suffices to focus the study on the GMHD estimator. Because the proofs of results for GMHD estimation have similar structures to the analogous results for GML estimation, the asymptotic properties of GML estimators are also reviewed in this chapter. Robustness properties of the GMPHD estimator for GLMs are examined via theoretical and empirical $\alpha$-influence function analysis. Once again, because the theoretical $\alpha$-influence curve is an asymptotic version of the finite sample empirical alpha-influence curve, the former only needs to be considered for the GMHD estimator.

This chapter is organized as follows. Sections 5.1 and 5.2 establish asymptotic properties of the GML estimator and the GMHD estimator, respectively, for the general multi-sample case (which includes GLMs). The application of the results to estimation in GLMs is discussed at the end of Section 5.2. Section 5.3 contains the robustness results of the GMPHD estimator in the GLM setting.

5.1 Asymptotic Properties of GML Estimators

This section contains proofs of the existence, uniqueness, consistency, and asymptotic normality of the GML estimator of the multiple population case. These results are stated without proof in §6.6 of Lehmann (1983). In Section 5.1.1 the existence and consistency of the GML estimators are established, while the asymptotic normality and efficiency of the GML estimators are shown in Section 5.1.2. The results are valid for both the continuous
and the discrete distributions.

For convenience, all the definitions of quantities used in the proofs are given first.

**Notation:**

1. \( l^{(ij)}(\theta) = \log f_{i, \theta}(Y_{ij}) \) is the log-likelihood of \( \theta \) for a single (say the \( j^{th} \)) observation of the \( i^{th} \) i.i.d. sample.

2. \( l^{(i)}(\theta) = \sum_{j=1}^{n_i} l^{(ij)}(\theta) \) is the total log-likelihood of \( \theta \) from the \( i^{th} \) sample.

3. \( l(\theta) = \sum_{i=1}^{m} l^{(i)}(\theta) \) is the total log-likelihood of \( \theta \) from all \( m \), independent samples.

4. \( l^{(ij)r}(\theta) = \frac{\partial}{\partial \theta_k} l^{(ij)}(\theta) \)

5. \( l^{(i)r}(\theta) = \frac{\partial}{\partial \theta_k} l^{(i)}(\theta) = \sum_{j=1}^{n_i} l^{(ij)r}(\theta) \) by additivity of differentiation

6. \( l^{(i)}(\theta) = \sum_{k=1}^{m} l^{(i)}(\theta) \)

7. \( l^{(ij)r}(\theta) = \frac{\partial}{\partial \theta_k} l^{(ij)}(\theta) \) is the \( p \times 1 \) vector of partial derivatives \( l^{(ij)r}(\theta) \).

8. \( l^{(i)}(\theta) = \sum_{j=1}^{n_i} l^{(i)r}(\theta) \) is the \( p \times 1 \) vector of partial derivatives \( l^{(i)}(\theta) \).

9. \( l'(\theta) = \sum_{i=1}^{m} l^{(i)}(\theta) \) is the \( p \times 1 \) vector of partial derivatives \( l'(\theta) \).

10. \( l^{(ij)nr}(\theta) = \frac{\partial^2}{\partial \theta_k \partial \theta_{\ell}} l^{(ij)}(\theta) \)

11. \( l^{(i)nr}(\theta) = \sum_{j=1}^{n_i} l^{(ij)nr}(\theta) \)

12. \( l^{(i)}(\theta) = \sum_{k=1}^{m} l^{(i)nr}(\theta) \)

13. \( l^{(ij)n}(\theta) = \frac{\partial^2}{\partial \theta_k \partial \theta_{\ell}} l^{(ij)}(\theta) \) is the \( p \times p \) matrix of partial derivatives \( l^{(ij)n}(\theta) \).

14. \( l^{(i)n}(\theta) = \sum_{j=1}^{n_i} l^{(ij)n}(\theta) \) is the \( p \times p \) matrix of partial derivatives \( l^{(i)n}(\theta) \).

15. \( l''(\theta) = \sum_{i=1}^{m} l^{(i)n}(\theta) \) is the \( p \times p \) matrix of partial derivatives \( l''(\theta) \).

16. \( l^{(ij)nr}(\theta) = \frac{\partial^3}{\partial \theta_k \partial \theta_{\ell} \partial \theta_s} l^{(ij)}(\theta) \)

17. \( l^{(i)nr}(\theta) = \sum_{j=1}^{n_i} l^{(ij)nr}(\theta) \)
18. $l_{k\ell}^{(m)}(\theta) = \sum_{i=1}^{m} l_{k\ell}^{(i)(m)}(\theta)$

19. $[\mathcal{I}_1^{(i)}(\theta)]_{k\ell} = E \left[ l_{k\ell}^{(ij)(i)(\theta)} \right] = -E \left[ l_{k\ell}^{(ij)(i)(\theta)} \right]$ is the information about $\theta$ contained in a single observation in the $i^{th}$ i.i.d. sample.

20. $I_1^{(i)}(\theta) = E \left[ l_{k\ell}^{(ij)(i)(\theta)} \cdot l_{k\ell}^{(ij)(i)(\theta)^T} \right] = -E \left[ l_{k\ell}^{(ij)(i)(\theta)} \right]$ is the information about $\theta$ contained in the $i^{th}$ i.i.d. sample.

21. $I^{(i)}(\theta) = n_i I_1^{(i)}(\theta)$ is the total information about $\theta$ contained in the $i^{th}$ i.i.d. sample.

22. $I(\theta) = \sum_{i=1}^{m} I^{(i)}(\theta)$ is the total information about $\theta$ contained in all $m$ independent samples.

23. $I_1(\theta) = \sum_{i=1}^{m} \lambda_i I_1^{(i)}(\theta)$ is the limit of the average information per observation $\frac{1}{n} I(\theta)$, where $\lambda_i$ is defined by Equation (1.14).

Note that all the log-likelihood functions in Notation 1, \ldots, 18 depend on the observations, $y_{ij}$'s, but $Y_{ij}$ is omitted from the notation for simplicity.

Next, we state the regularity conditions and assumptions that need to be satisfied by each of the probability distribution functions $f_{i,\theta}$.

**Regularity Conditions:**

(R1) The distributions $f_{i,\theta}$ of the observations are identifiable. That is, $f_{i,\theta_1} = f_{i,\theta_2}$ implies $\theta_1 = \theta_2$.

(R2) The distributions $f_{i,\theta}$ have common support. That is, the set $\{ y : f_{i,\theta}(y) > 0 \}$ is independent of $\theta$.

(R3) The observations $Y_{i1}, Y_{i2}, \ldots, Y_{im_i}$ are i.i.d. with common density function $f_{i,\theta}(y_{i,j})$ with respect to a $\sigma$-finite measure $\mu$.

(R4) The parameter space $\Theta$ contains an open subset $\omega$ of which the true parameter, $\theta_0$, is an interior point.

**Assumptions:**
(A1) For all \( \theta \in \omega \) and for almost all \( y \), the density function \( f_{i, \theta}(y) \) admits all third derivatives \( \frac{\partial^3 f_{i, \theta}(y)}{\partial \theta_i \partial \theta_j \partial \theta_k} \).

(A2) The integral \( \int f_{i, \theta}(y)dy \) or the summation \( \sum_y f_{i, \theta}(y) \) may be twice differentiated under the integral or the summation sign so that the first and second log-density derivatives of \( f_{i, \theta} \) satisfy the equations

\[
E_\theta \left[ l^{(ij)}_{k}(\theta) \right] = 0 \quad \text{for } k = 1, 2, \ldots, p,
\]

and hence

\[
\frac{\partial}{\partial \theta_i} \left[ E_\theta \left[ l^{(ij)}_{k}(\theta) \right] \right] = \frac{\partial}{\partial \theta_i} \left[ l^{(ij)}_{k}(\theta) \right] = 0 \quad \text{for } k, \ell = 1, 2, \ldots, p.
\]

Equivalently,

\[
E_\theta \left[ l^{(ij)'}(\theta) \right] = 0,
\]

and

\[
I^{(i)}_{k}(\theta) = E_\theta \left[ l^{(ij)'}(\theta) \cdot l^{(ij)'}(\theta) \right] = -E_\theta \left[ l^{(ij)''}(\theta) \right] \quad \text{for } k, \ell = 1, 2, \ldots, p.
\]

(A3) All \( I^{(i)}_{k}(\theta) \) are finite and that \( I^{(i)}_{k}(\theta) \) is positive definite for all \( \theta \in \omega \) which implies that \( I^{(i)}_{k}(\theta) \) is non-singular and that the \( p \) statistics \( I^{(i)}_{1}(\theta), \ldots, I^{(i)}_{p}(\theta) \) are independent with probability 1.

(A4) There exist functions \( M^{(i)}_{k\ell s} \) such that

\[
\left| l^{(ij)''}_{k\ell s}(\theta) \right| \leq M^{(i)}_{k\ell s}(y_{ij}) \quad \text{for all } \theta \in \omega,
\]

where

\[
m^{(i)}_{k\ell s} = E_{\theta_0} \left[ M^{(i)}_{k\ell s}(Y_{ij}) \right] < \infty \quad \text{for all } k, \ell, s = 1, 2, \ldots, p.
\]

5.1.1 Uniqueness, Existence, and Consistency

The existence, uniqueness, and consistency of the GML estimator are established in Theorem 5.1.1.
Theorem 5.1.1 Suppose each \( f_{i, \theta} \) satisfies Regularity Conditions (R1)-(R4) and Assumptions (A1)-(A4) and suppose sample sizes \( n_i \)'s satisfy the limit condition (1.14), then with probability tending to 1 as \( n \to \infty \), the log-likelihood equations (Equation (1.16)) have a sequence of roots \( \{ \hat{\theta}_n \}_{n=1}^{\infty} \) such that

\[
\hat{\theta}_n \xrightarrow{p} \theta_0 \text{ componentwise.}
\]

Outline of the proof: To prove this result, consider the behavior of the log-likelihood \( l(\theta) \) on the sphere \( Q_r \) with center at the true point \( \theta_0 \) and radius \( r \). First, it is shown that for any sufficiently small \( r \),

\[
P_{\theta_0} \left( l(\theta) < l(\theta_0) \right) \longrightarrow 1 \text{ as } n \to \infty
\]

at all points \( \theta \) on the surface of \( Q_r \), and hence \( l(\theta) \) has a local maximum in the interior of \( Q_r \) with probability tending to 1 as \( n \to \infty \). It then follows that for any small enough \( r > 0 \), with probability tending to 1 as \( n \to \infty \), the log-likelihood equations have a sequence of roots \( \{ \hat{\theta}_n \}_{n=1}^{\infty} = \{ \hat{\theta}_n(r) \}_{n=1}^{\infty} \) within \( Q_r \) such that

\[
P_{\theta_0} \left( \left| \hat{\theta}_{nk} - \theta_{0k} \right| < r \right) \longrightarrow 1 \text{ for all } k = 1, 2, \ldots, p.
\]

Finally, proof is completed by showing that such a sequence can be determined without depending on \( r \).

In preparation for establishing Theorem 5.1.1, four results ("tools") used in the proof are given.

Tools:

\[(T1) \ \frac{1}{n} l'_k(\theta_0) \xrightarrow{p} 0.\]

Proof: By the law of large numbers and by Assumption (A2) Equation (5.1),

\[
\frac{1}{n} l'_k(\theta_0) = \sum_{i=1}^{m} \left( \frac{n_i}{n} \right) \left( \frac{1}{n_i} \sum_{j=1}^{n_i} l'_{k(ij)}(\theta_0) \right) \xrightarrow{p} \sum_{i=1}^{m} \lambda_i E_{\theta_0} \left[ l'_{k(ij)}(\theta_0) \right] = 0,
\]

where \( l'_k(\theta_0) \) and \( l'_{k(ij)}(\theta_0) \) are \( l'_k(\theta) \) and \( l'_{k(ij)}(\theta) \) evaluated at \( \theta_0 \).
Proof: By the law of large numbers and by Assumption (A2) Equation (5.2),
\[
-\frac{1}{n} l_{kl}(\theta_0) \rightarrow \sum_{i=1}^{m} \left( \frac{n_i}{n} \right) \left( -\frac{1}{n_i} \sum_{j=1}^{n_i} l_{kl}^{(ij)}(\theta_0) \right) \rightarrow \sum_{i=1}^{m} \lambda_i E_{\theta_0} \left[ -l_{kl}^{(ij)}(\theta_0) \right] = \sum_{i=1}^{m} \lambda_i \left[ l_{kl}^{(ij)}(\theta_0) \right]_{kl} = \left[ l_{kl}^{(ij)}(\theta_0) \right]_{kl},
\]
where \( l_{kl}^{(ij)}(\theta_0), l_{kl}^{(ij)}(\theta_0), [I_1(\theta_0)]_{kl}, \) and \( [I_1^{(ij)}(\theta_0)]_{kl} \) are \( l_{kl}^{(ij)}(\theta), l_{kl}^{(ij)}(\theta), [I_1(\theta)]_{kl}, \) and \( [I_1^{(ij)}(\theta)]_{kl} \) evaluated at \( \theta_0 \).

(T3) \( \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} M_{kl}^{(ij)}(y_{ij}) \rightarrow \sum_{i=1}^{m} \lambda_i m_{kl}^{(ij)} \).

Proof: By the law of large numbers and by Equation (5.6) in Assumption (A4),
\[
\frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} M_{kl}^{(ij)}(y_{ij}) \rightarrow \sum_{i=1}^{m} \left( \frac{n_i}{n} \right) \left( \frac{1}{n_i} \sum_{j=1}^{n_i} M_{kl}^{(ij)}(y_{ij}) \right) \rightarrow \sum_{i=1}^{m} \lambda_i E_{\theta_0} \left[ M_{kl}^{(ij)}(y_{ij}) \right] = \sum_{i=1}^{m} \lambda_i m_{kl}^{(ij)} < \infty.
\]

(T4) Suppose that \( A \) is a \( p \times p \) symmetric matrix with eigenvalues \( \alpha_1, \alpha_2, \ldots, \alpha_p \) and corresponding unit eigenvectors \( v_1, v_2, \ldots, v_p \). Then, a quadratic form \( x^T A x \) can be expressed in terms of coordinates \( x^* \) with respect to the basis \( \{v_1, v_2, \ldots, v_p\} \) such that

\[
x^T A x = x^* D x^* = \sum_{k=1}^{p} \alpha_k x_k^2, \quad \text{and}
\]

\[
\sum_{k=1}^{p} x_k^2 = \sum_{k=1}^{p} x_k^2,
\]

where \( D = \text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_p) \).

Proof: Since \( A \) is symmetric, it is orthogonally diagonalizable. That is, there exists an orthogonal matrix \( B \) and a diagonal matrix \( D \) such that

\[
A = B D B^{-1} = B D B^T \quad \text{and}
\]

\[
D = B^{-1} A B = B^T A B.
\]
where $D = \text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_p)$ and $B = (v_1, v_2, \ldots, v_p)$. Now, substitute $x$ in the quadratic form $x^TAx$ by $Bx^*$; it yields

$$x^TAx = (Bx^*)^TA(Bx^*) = x^*(B^TAB)x^*$$

$$= x^*^TDx^* = \sum_{k=1}^{p} \alpha_k x_k^2,$$

and that

$$\sum_{k=1}^{p} x_k^2 = x^*^x = (B^{-1}x)^T(B^{-1}x) = x^TBB^{-1}x$$

$$= x^Tx = \sum_{k=1}^{p} x_k^2.$$

\[\square\]

**Proof of Theorem 5.1.1:** The proof starts by expanding the log-likelihood $l(\theta)$ about the true parameter value $\theta_0$ and dividing through by $n$ to get

$$\frac{1}{n}l(\theta) - \frac{1}{n}l(\theta_0) = \frac{1}{n} \sum_{k=1}^{p} l_k'(\theta_0)(\theta_k - \theta_{0k})$$

$$+ \frac{1}{2n} \sum_{k=1}^{p} \sum_{\ell=1}^{p} l_k''(\theta_0)(\theta_k - \theta_{0k})(\theta_\ell - \theta_{0\ell})$$

$$+ \frac{1}{6n} \sum_{k=1}^{p} \sum_{\ell=1}^{p} \sum_{s=1}^{p} l_k''''(\theta^*) (\theta_k - \theta_{0k})(\theta_\ell - \theta_{0\ell})(\theta_s - \theta_{0s})$$

$$= S_1 + S_2 + S_3,$$

where $\theta^*$ is on the line segment joining $\theta$ to $\theta_0$. By Inequality (5.5) in Assumption (A4),

$$|l^{(i)}_{k\ell s}(\theta^*)| \leq M^{(i)}_{k\ell s}(y_{ij}),$$

$S_3$ may be rewritten as

$$\frac{1}{6n} \sum_{k=1}^{p} \sum_{\ell=1}^{p} \sum_{s=1}^{p} (\theta_k - \theta_{0k})(\theta_\ell - \theta_{0\ell})(\theta_s - \theta_{0s}) \sum_{i=1}^{m} \sum_{j=1}^{n_1} \gamma^{(i)}_{k\ell s}(y_{ij}) M^{(i)}_{k\ell s}(y_{ij}),$$

where

$$0 \leq |\gamma^{(i)}_{k\ell s}(y_{ij})| \leq 1 \quad \text{for all } i, j, k, \ell, \text{ and } s.$$

This will be the $S_3$ to which the rest of proof refers.
Now, it is desired to show that
\[ P_{\theta_0} \left( \frac{1}{n} l(\theta) - \frac{1}{n} l(\theta_0) < 0 \right) \rightarrow 1 \quad \text{as } n \to \infty \text{ for all } \theta \text{ on } Q_r \]
if \( r \) is sufficiently small. To do this, observe the difference \( \frac{1}{n} l(\theta) - \frac{1}{n} l(\theta_0) \) for \( \theta \) on \( Q_r \) through its three terms \( S_1, S_2, \) and \( S_3 \). First see that, for every \( \theta \) on the surface of \( Q_r \),
\[ |S_1| \leq \frac{1}{n} \sum_{k=1}^{p} |l_k'(\theta_0)||\theta_k - \theta_0k| = \frac{r}{n} \sum_{k=1}^{p} |l_k'(\theta_0)|. \]
Then, for any given \( r \), it follows from Tool (T1) that
\[ P_{\theta_0} \left( \frac{1}{n} l_k'(\theta_0) < r^2 \right) \rightarrow 1 \quad \text{as } n \to \infty, \]
which implies that
\[ P_{\theta_0} \left( \frac{r}{n} \sum_{k=1}^{p} |l_k'(\theta_0)| < pr^3 \right) \rightarrow 1 \quad \text{as } n \to \infty, \]
and hence that
\[ (5.7) \quad P_{\theta_0} \left( |S_1| < pr^3 \right) \rightarrow 1 \quad \text{as } n \to \infty. \]

Next, consider
\[ 2S_2 = \frac{1}{n} \sum_{k=1}^{p} \sum_{\ell=1}^{p} l_k''(\theta_0)(\theta_k - \theta_0k)(\theta_\ell - \theta_0\ell) \]
\[ = \sum_{k=1}^{p} \sum_{\ell=1}^{p} \left\{ -[I_1(\theta_0)]_{k\ell}(\theta_k - \theta_0k)(\theta_\ell - \theta_0\ell) \right\} \]
\[ + \sum_{k=1}^{p} \sum_{\ell=1}^{p} \left\{ \frac{1}{n} l_k''(\theta_0) - [-I_1(\theta_0)]_{k\ell} \right\} (\theta_k - \theta_0k)(\theta_\ell - \theta_0\ell) \]
\[ = S_{21} + S_{22}. \]
\[ S_{21} = -(\theta - \theta_0)^T I_1(\theta_0)(\theta - \theta_0), \]
which is a negative quadratic form in the variables \((\theta_k - \theta_0k)\). Since \( I_1(\theta_0) \) is symmetric and by Assumption (A3) positive definite, therefore, Tool (T4) indicates that \( S_{21} \) can be orthogonally diagonalized into diagonal form
\[ -\sum_{k=1}^{p} \alpha_k \xi_k^2 \] with \( Q_r \) becoming
\[ \sum_{k=1}^{p} \xi_k^2 = \sum_{k=1}^{p} (\theta_k - \theta_0k)^2 = r^2, \]
where \( \alpha_1, \alpha_2, \ldots, \alpha_p (> 0) \) are eigenvalues of \( I_1(\theta) \). Let \( \alpha_{\min} = \min \{ \alpha_1, \alpha_2, \ldots, \alpha_p \} \). Then,
\[ S_{21} = -\sum_{k=1}^{p} \alpha_k \xi_k^2 \leq -\alpha_{\min} \sum_{k=1}^{p} \xi_k^2 = -\alpha_{\min} r^2. \]
Now observe that, for every \( \theta \) on \( Q_r \),

\[
|S_{22}| \leq \sum_{k=1}^{p} \sum_{\ell=1}^{p} \left| \frac{1}{n} l_{k\ell}(\theta) - [-1(\theta)]_{k\ell} \right| |\theta_k - \theta_{0k}| |\theta_\ell - \theta_{0\ell}|
\]

\[
= r^2 \sum_{k=1}^{p} \sum_{\ell=1}^{p} \left| \frac{1}{n} l_{k\ell}(\theta) - [-1(\theta)]_{k\ell} \right| ,
\]

and, according to Tool (T2), for any \( r \),

\[
P_{\theta_0} \left( \left| \frac{1}{n} l_{k\ell}(\theta) - [-1(\theta)]_{k\ell} \right| < r \right) \rightarrow 1 \text{ as } n \rightarrow \infty.
\]

It follows from an argument analogous to that for \( S_1 \) that

\[
P_{\theta_0} \left( |S_{22}| \leq p^2 r^3 \right) \rightarrow 1 \text{ as } n \rightarrow \infty.
\]

Therefore, with probability tending to 1 as \( n \rightarrow \infty \),

\[
(5.8) \quad S_2 = \frac{1}{2} (S_{21} + S_{22}) \leq \frac{1}{2} (S_{21} + |S_{22}|) \leq -\frac{1}{2} \alpha_{\min} r^2 + \frac{1}{2} p^2 r^3.
\]

Finally, consider \( S_3 \). For every \( \theta \) on \( Q_r \),

\[
|S_3| \leq \frac{1}{6} \sum_{k=1}^{p} \sum_{\ell=1}^{p} \sum_{s=1}^{p} |\theta_k - \theta_{0k}| |\theta_\ell - \theta_{0\ell}| |\theta_s - \theta_{0s}| \left| \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \gamma_{k\ell s}(y_{ij}) M_{k\ell s}^{(i)}(y_{ij}) \right|
\]

\[
= \frac{r^3}{6} \sum_{k=1}^{p} \sum_{\ell=1}^{p} \sum_{s=1}^{p} \left| \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \gamma_{k\ell s}(y_{ij}) M_{k\ell s}^{(i)}(y_{ij}) \right|
\]

\[
\leq \frac{r^3}{6} \sum_{k=1}^{p} \sum_{\ell=1}^{p} \sum_{s=1}^{p} \left| \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} M_{k\ell s}^{(i)}(y_{ij}) \right|
\]

because \( |\gamma_{k\ell s}^{(i)}(y_{ij})| \leq 1 \) and \( M_{k\ell s}^{(i)}(y_{ij}) > 0 \). Now, using Tool (T3), we get

\[
P_{\theta_0} \left( \left| \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} M_{k\ell s}^{(i)}(y_{ij}) \right| < 2 \sum_{i=1}^{m} \lambda_i m_{k\ell s}^{(i)} \right) \rightarrow 1 \text{ as } n \rightarrow \infty,
\]

which implies that

\[
(5.9) \quad P_{\theta_0} \left( |S_3| \leq \frac{r^3}{3} \sum_{k=1}^{p} \sum_{\ell=1}^{p} \sum_{s=1}^{p} \left( \sum_{i=1}^{m} \lambda_i m_{k\ell s}^{(i)} \right) \right) \rightarrow 1 \text{ as } n \rightarrow \infty.
\]
Combining the results from inequalities (5.7), (5.8), and (5.9), one sees that, with probability tending to 1 as \( n \to \infty \),

\[
\max\{S_1 + S_2 + S_3\} \leq \max\{|S_1| + S_2 + |S_3|\}
\leq pn^3 - \frac{1}{2} \alpha_{\min} r^2 + \frac{1}{2} p^2 r^3 + \frac{r^3}{3} \sum_{k=1}^{p} \sum_{\ell=1}^{p} \sum_{s=1}^{p} \left( \sum_{i=1}^{m} \lambda_{i} m_{i\ell s}^{(i)} \right)
\]

\[
= -\frac{1}{2} \alpha_{\min} r^2 + \left[ p + \frac{p^2}{2} + b \right] r^3,
\]

(where \( b = \frac{1}{3} \sum_{k=1}^{p} \sum_{\ell=1}^{p} \sum_{s=1}^{p} (\sum_{i=1}^{m} \lambda_{i} m_{i\ell s}^{(i)}) \)), which is negative when

\[
-\frac{\alpha_{\min}}{2} + \left[ p + \frac{p^2}{2} + b \right] r < 0,
\]

i.e., when

\[
r < \frac{\alpha_{\min}}{p^2 + 2p + 2b}.
\]

This result says that with \( r \) sufficiently small,

\[
P_{\theta_0} (l(\theta) - l(\theta_0) < 0) \to 1 \text{ as } n \to \infty
\]

at all points \( \theta \) on \( Q_r \), and hence that there exists a consistent sequence of solutions

\[
\{\hat{\theta}_n\}_{n=1}^{\infty} = \{\hat{\theta}_n(r)\}_{n=1}^{\infty} \text{ within } Q_r \text{ of the likelihood equations.}
\]

It remains to show that such a sequence can be independent of \( r \). By the continuity of \( l(\theta) \), the limit of a sequence of roots of likelihood equations is also a root. Let \( \hat{\theta}_{n}^{\ast} \) be the root closest to \( \theta_0 \). Then clearly

\[
P_{\theta_0} \left( |\hat{\theta}_{nk}^{\ast} - \theta_{0k}| < r \right) \to 1 \text{ for all } k = 1, 2, \ldots, p,
\]

and this completes the proof of Theorem 5.1.1.

\begin{remark}
A consistent root of the log-likelihood equations may not be unique, but if it is unique, then it is the GML estimator.
\end{remark}

5.1.2 Asymptotic Normality and Efficiency

Theorem 5.1.2 establishes the asymptotic normality and asymptotic efficiency (under the true model) of the GML estimator.
Theorem 5.1.2 If regularity conditions (R1)-(R4) and Assumptions (A1)-(A4) hold for each $f_{i,t}$, and the limit condition (1.14) is satisfied, then any consistent sequence of roots $\left\{ \hat{\theta}_n \right\}_{n=1}^{\infty}$ of the log-likelihood equations (Equation (1.16)) is asymptotically normal. That is,

$$\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \xrightarrow{D} N_p \left( 0, [I_1(\theta_0)]^{-1} \right).$$

This implies asymptotic efficiency of the estimators $\hat{\theta}_{nk}$ in the sense that

$$\sqrt{n} \left( \hat{\theta}_{nk} - \theta_{0k} \right) \xrightarrow{D} N_1 \left( 0, [I_1(\theta_0)]^{-1}_{kk} \right),$$

where $[I_1(\theta_0)]^{-1}_{kk}$ is the $(k,k)^{th}$ element of $[I_1(\theta_0)]^{-1}$.

Outline of the proof: All that needs to be shown is that for any consistent root $\hat{\theta}_n$ of the log-likelihood equations, $\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right)$ has the same limit distribution as that of $[I_1(\theta_0)]^{-1} \left( \frac{1}{\sqrt{n}} l'(\theta_0) \right)$ and that $[I_1(\theta_0)]^{-1} \left( \frac{1}{\sqrt{n}} l'(\theta_0) \right) \xrightarrow{D} N_p \left( 0, [I_1(\theta_0)]^{-1} \right)$.

Three additional tools are needed for establishing Theorem 5.1.2.

Tools:

(T5)

\begin{align*}
\sqrt{n} l'_k(\theta_0) & \xrightarrow{D} N_1 \left( 0, [I_1(\theta_0)]_{kk} \right) \quad \text{for all } k = 1, 2, \ldots, p, \text{ and} \\
\sqrt{n} l'(\theta_0) & \xrightarrow{D} N_p \left( 0, I_1(\theta_0) \right). \quad (5.11)
\end{align*}

Proof: By the multivariate central limit theorem, Slutsky’s theorem, and Equations (5.1) and (5.2) in Assumption (A2),

$$\sqrt{n} l'_k(\theta_0) = \sum_{i=1}^{m} \sqrt{n_i} \cdot \sqrt{n_i} \left( \frac{1}{n_i} \sum_{j=1}^{n_i} l^{(ij)}_k(\theta_0) \right) \xrightarrow{D} N_1 \left( 0, \sum_{i=1}^{m} \lambda_i \mathbf{E}_{\theta_0} \left[ l^{(ij)}_k(\theta_0) \cdot l^{(ij)}_k(\theta_0) \right] \right)$$

$$= N_1 \left( 0, \sum_{i=1}^{m} \lambda_i \left[ I_1^{(ij)}(\theta_0) \right]_{kk} \right)$$

$$= N_1 \left( 0, [I_1(\theta_0)]_{kk} \right)$$
Similarly, using Equations (5.3) and (5.4) instead of Equations (5.1) and (5.2),

\[
\frac{1}{\sqrt{n}} l'(\theta_0) = \sum_{i=1}^{m} \sqrt{\frac{n_i}{n}} \cdot \sqrt{n_i} \left( \frac{1}{n_i} \sum_{j=1}^{n_i} l^{ij}(\theta_0) \right)
\]

\[
\overset{D}{\rightarrow} N_p\left(0, \sum_{i=1}^{m} \lambda_i E_{\theta_0}\left[l^{ij}(\theta_0) \cdot l^{ij}(\theta_0)^T\right]\right)
\]

\[
= N_p\left(0, \sum_{i=1}^{m} \lambda_i I_i(\theta_0)\right)
\]

\[
= N_p\left(0, I_1(\theta_0)\right)
\]

\(\square\)

(T6) \(\left| \frac{1}{n} l_{k\ell s}^m(\theta) \right|\) is bounded asymptotically for all \(k, \ell, s = 1, 2, \ldots, p\) and for all \(\theta \in \omega\).

Proof: By Assumption (A4) Inequality (5.5) and result from Tool (T3),

\[
\left| \frac{1}{n} l_{k\ell s}^m(\theta) \right| = \left| \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} l_{k\ell s}^{ij}(\theta) \right| \leq \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left| l_{k\ell s}^{ij}(\theta) \right| \leq \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} M_{k\ell s}^{ij}(y_{ij})
\]

\[\overset{p}{\rightarrow} \sum_{i=1}^{m} \lambda_i m_{k\ell s}^{(i)} < \infty\]

\(\square\)

(T7) (Lehmann 1983, §6.4, Lemma 4.1) Let \(T_n = (T_{n1}, T_{n2}, \ldots, T_{np})^T\) and let \(\{T_n\}_{n=1}^{\infty}\) be a sequence of random vectors tending weakly (in distribution) to \(T = (T_1, T_2, \ldots, T_p)^T\) and suppose that for each fixed \(k\) and \(\ell\), the \((k, \ell)^{th}\) elements \(\{A_n\}_{k\ell}^{\infty}\) of matrices \(\{A_n\}_{n=1}^{\infty}\) is a sequence of random variables tending in probability to constants \([A]_{k\ell}\) for which the matrix \(A\) is nonsingular. Let \(B = A^{-1}\). Then if the distribution of \(T_n\) has a density with respect to Lebesgue measure over \(R^p\), the solutions \(X_n = (X_{n1}, X_{n2}, \ldots, X_{np})^T\) of a system of \(p\) random linear equations in \(p\) unknowns

\[T_{nk} = \sum_{\ell=1}^{p} [A_n]_{k\ell} X_{n\ell} \quad \text{for all } k = 1, 2, \ldots, p\]

have the same limit distribution as those of

\[X_{nk} = \sum_{\ell=1}^{p} [B]_{k\ell} T_{n\ell} \quad \text{for all } k = 1, 2, \ldots, p\]
and tend in probability to the solutions $X = (X_1, X_2, \ldots, X_p)^T$ of

$$T_k = \sum_{\ell=1}^{p} [A]_{k\ell} X_\ell \quad \text{for all } k = 1, 2, \ldots, p$$

(5.14)

given by

$$X_k = \sum_{\ell=1}^{p} [B]_{k\ell} T_\ell \quad \text{for all } k = 1, 2, \ldots, p.$$  

(5.15)


Remark: This result is still true if the distribution of $T_n$ has a density with respect to counting measure over $\mathbb{R}^p$.

Proof of Theorem 5.1.2: Expand $l_k'(\theta)$ about $\theta_0$ to obtain

$$l_k'(\theta) = l_k'(\theta_0) + \sum_{\ell=1}^{p} l_k''(\theta_0)(\theta_\ell - \theta_{0\ell}) + \frac{1}{2} \sum_{\ell=1}^{p} \sum_{s=1}^{p} l_k'''(\theta^*)(\theta_\ell - \theta_{0\ell})(\theta_s - \theta_{0s}),$$

where $\theta^*$ is a point on the line segment connecting $\theta$ and $\theta_0$. In this expansion, replace $\theta$ by a consistent root of the log-likelihood equations $\hat{\theta}_n$, which exists with probability tending to 1 as $n \to \infty$ by the first part of the theorem, so that $l_k'(\hat{\theta}_n) = 0$ for all $k = 1, 2, \ldots, p$.

Then, multiplying the expansion (now with $\hat{\theta}_n$) through by $\frac{1}{\sqrt{n}}$ and rearrange to get

$$\frac{1}{\sqrt{n}} l_k'(\theta_0) = \sqrt{n} \sum_{\ell=1}^{p} (\hat{\theta}_{nk} - \theta_{0k}) \left\{ -\frac{1}{n} l_k''(\theta_0) - \frac{1}{2n} \sum_{s=1}^{p} (\hat{\theta}_{ns} - \theta_{0s}) l_k'''(\theta^*) \right\} \quad \text{for } k = 1, 2, \ldots, p.$$

This is a system of $p$ random linear equations, which needs to be solved for $p$ differences $(\hat{\theta}_{nk} - \theta_{0k})$ to obtain the limiting distributions of $\sqrt{n}(\hat{\theta}_{nk} - \theta_{0k})$. One can see that this system of equations has the form (5.12) in Tool (T7) with

$$T_{nk} = \frac{1}{\sqrt{n}} l_k'(\theta_0),$$

$$[A_n]_{k\ell} = -\frac{1}{n} l_k''(\theta_0) - \frac{1}{2n} \sum_{s=1}^{p} (\hat{\theta}_{ns} - \theta_{0s}) l_k'''(\theta^*), \quad \text{and}$$

$$X_{n\ell} = \sqrt{n}(\hat{\theta}_{n\ell} - \theta_{0\ell}).$$

According to Tool (T5) Equations (5.10) and (5.11),

$$T_{nk} \xrightarrow{D} T_k = N_1(0, [I_1(\theta_0)]_{kk}) \quad \text{for } k = 1, 2, \ldots, p, \quad \text{and}$$

$$T_n \xrightarrow{D} T = N_p(0, I_1(\theta_0)),$$
while

\[ [A_n]_{k\ell} \xrightarrow{p} [A]_{k\ell} = [I_1(\theta_0)]_{k\ell} \]

because, by Tool (T2),

\[-\frac{1}{n} l_{k\ell}''(\theta_0) \xrightarrow{p} [I_1(\theta_0)]_{k\ell} \]

and, by Tool (T6) and by the consistency of \( \hat{\theta}_n \),

\[-\frac{1}{2n} \sum_{s=1}^{p} (\hat{\theta}_{ns} - \theta_{0s}) l_{k\ell s}'''(\theta^*) \xrightarrow{p} 0.\]

Let \( B = A^{-1} = I_1(\theta_0) \). Then, by Tool (T7) Equations (5.13), (5.14), and (5.15), the limit distributions of the \( \sqrt{n}(\hat{\theta}_{nk} - \theta_{0k}) \)'s are the limit distributions of the solutions \( X_{nk} \)'s of the equations

\[ X_{nk} = \sum_{\ell=1}^{p} [I_1(\theta_0)]_{k\ell}^{-1} \left( \frac{1}{\sqrt{n}} l_{k\ell}'(\theta_0) \right) \]

and tend in probability to the distributions of the solutions \( X_k \)'s of the equations

\[ T_k = \sum_{\ell=1}^{p} [I_1(\theta_0)]_{k\ell} X_{k\ell} \]

given by

\[ X_k = \sum_{\ell=1}^{p} [I_1(\theta_0)]_{k\ell}^{-1} T_{k\ell}. \]

Hence, \( \sqrt{n}(\hat{\theta}_n - \theta_0) \) has the same limit distribution as that of \( [I_1(\theta_0)]^{-1} \left( \frac{1}{\sqrt{n}} l'(\theta_0) \right) \), which is the distribution of \( X \), i.e., the distribution of \( [I(\theta_0)]^{-1} T \), and which is \( N_p \left( 0, [I_1(\theta_0)]^{-1} \right) \).

This completes the proof of asymptotic normality and efficiency of \( \hat{\theta}_n \).

5.2 Asymptotic Properties of GMHD Estimators

This section presents the asymptotic properties of the GMHD estimator. In Section 5.2.1, the existence, uniqueness, and consistency of the GMHD estimator are established. This is an extension of Beran's (1977) Theorem 1 for MHD estimator. These results are applicable to both discrete and continuous families of distributions. In Section 5.2.2, the GMHD estimator is shown to be asymptotically normal and to be asymptotically efficient.
under the assumed model for the case of count data. This is a generalization of Simpson’s (1987) Theorem 2 for the MHD estimator.

As with the GML estimator in Section 5.1, the results in this section are established for the cases where the parameters $\theta$ are not necessarily regression parameters. The corresponding results of GML and GMHD estimators in GLM setting are then discussed at the end of this section.

In preparation for developments in this section, it is necessary to define the GMHD functional. Recall from Section 2.1 the definition of the GMHD estimator. Suppose that there are $m$ random samples $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^T$ for $i = 1, 2, \ldots, m$ from $m$ different populations with distribution density functions $g_1, g_2, \ldots, g_m$ with respect to a $\sigma$-finite measure $\mu$. Let each $F_i = \{f_i, \theta : \theta \in \Theta\}$ be the set of density functions of the assumed parametric family of distributions for $g_i$ and assume that parameter $\theta$ of dimension $p$ is the common parameter of all $m$ families. Furthermore, as with GML estimation, the limit condition supposes that $m$ is fixed and each individual sample size $n_i \to \infty$ all at the same rate such that the proportions $\frac{n_i}{\sum n_i} \to \lambda_i$ as the total sample size $n = \sum_{i=1}^m n_i \to \infty$. Then, according to Equation (2.1), the aim of GMHD estimation is to identify the element $\theta_0 \in \Theta$ that minimizes $\sum_{i=1}^m \lambda_i \|f_i, \theta - g_i\|_2$. That is, it is desired to find the GMHD functional $T(g) = T(g_1, g_2, \ldots, g_m) = \theta_0 \in \Theta$ defined by

$$T(g) = T(g_n) = T(\hat{g}_{n1}, \hat{g}_{n2}, \ldots, \hat{g}_{nm})$$

This definition is equivalent to the definition given in Equation (2.1).
Next, consider how the results for the MHD estimator may be extended to the GMHD estimator. Analogous to the discussion of how ML results are extended to GML estimation in Section 1.8, the central concept in the proofs for MHD estimation in the i.i.d. case (Beran 1977 and Simpson 1987) is the theoretical squared Hellinger distance

\[ H(\theta; g) = u_n^2(f_{\theta}, g) = \left\| f_{\theta}^{1/2} - g^{1/2} \right\|_2^2 \]

provided by any random sample from \( g \). To generalize the proofs for MHD estimation to GMHD estimation, one must recognize what the corresponding key element is for the non-i.i.d. case. First, denote the theoretical squared Hellinger distance provided by any random sample from \( g_i \) to be

\[ H(\theta; g_i) = u_n^2(f_{i,\theta}, g_i) = \left\| f_{i,\theta}^{1/2} - g_i^{1/2} \right\|_2^2 \]

and assume that the squared Hellinger distance provided by independent samples is additive, then the total theoretical squared Hellinger distance provided by \( m \) random samples is

\[ H(\theta; g) = H(\theta; g_1, g_2, \ldots, g_m) = \sum_{i=1}^{m} n_i H(\theta; g_i) = \sum_{i=1}^{m} n_i \left\| f_{i,\theta}^{1/2} - g_i^{1/2} \right\|_2^2, \]

and hence the average theoretical squared Hellinger distance provided by any one of the \( m \) random samples is

\[ \frac{1}{n} H(\theta; g) = \frac{1}{n} H(\theta; g_1, g_2, \ldots, g_m) = \sum_{i=1}^{m} \frac{n_i}{n} H(\theta; g_i) = \sum_{i=1}^{m} \frac{n_i}{n} \left\| f_{i,\theta}^{1/2} - g_i^{1/2} \right\|_2^2, \]

which tends to

\[ \sum_{i=1}^{m} \lambda_i \left\| f_{i,\theta}^{1/2} - g_i^{1/2} \right\|_2^2 \quad \text{as} \quad n \to \infty. \]

We shall see below that this theoretical squared Hellinger distance per random sample (among all \( m \) samples) \( \frac{1}{n} H(\theta; g) \) plays a very important role in establishing the asymptotic properties of the GMHD estimator.

Again, for convenience, the notation used in this section is given first.

**Notation:**

1. \( \Psi = (\Psi, \Psi, \ldots, \Psi)^T \), where \( \Psi \) is the set of all densities with respect to Lebesgue measure on the real line.
2. \( \mathcal{F} = (\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_m)^T \), where each \( \mathcal{F}_i = \{ f_{i,\theta} : \theta \in \Theta \} \).

3. \( f_\theta = (f_{1,\theta}, f_{2,\theta}, \ldots, f_{m,\theta})^T \), where each \( f_{i,\theta} \) is a member of \( \mathcal{F}_i \).

4. \( \mathcal{G} = (\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_m)^T \), where each \( \mathcal{G}_i \) is the set of densities of some class of distributions, which may or may not be \( \mathcal{F}_i \).

5. \( g = (g_1, g_2, \ldots, g_m)^T \), where each \( g_i \) is a member of \( \mathcal{G}_i \).

6. \( g \in \Psi \) means \( g_i \in \Psi \) for all \( i = 1, 2, \ldots m \).

7. \( g \in \mathcal{G} \) means \( g_i \in \mathcal{G}_i \) for all \( i = 1, 2, \ldots m \).

8. \( g = f_{\theta_0} \in \mathcal{F} \) means \( g_i = f_{i,\theta_0} \in \mathcal{F}_i \) for \( i = 1, 2, \ldots m \).

9. \( n = (n_1, n_2, \ldots, n_m)^T \)

10. \( \hat{g}_n = (\hat{g}_{n_1}, \hat{g}_{n_2}, \ldots, \hat{g}_{n_m})^T \)

11. \( \hat{g}_n \overset{p}{\rightarrow} g \) means \( \hat{g}_{n_i} \overset{p}{\rightarrow} g_i \) for all \( i = 1, 2, \ldots m \).

12. \( \hat{\theta}_n \overset{p}{\rightarrow} \theta \) means \( \hat{\theta}_{n,k} \overset{p}{\rightarrow} \theta_k \) for all \( k = 1, 2, \ldots, p \).

13. \( T(\hat{g}_n) \overset{p}{\rightarrow} T(g) \) means \( [T(\hat{g}_n)]_k \overset{p}{\rightarrow} [T(g)]_k \) for all \( k = 1, 2, \ldots, p \).

14. \( s_{i,\theta} = f_{i,\theta}^{1/2} \)

15. \( \hat{s}_{i,\theta}^{(k)} = \frac{\partial}{\partial \theta_k} s_{i,\theta} \)

16. \( \check{s}_{i,\theta}^{(k)} = \frac{\partial^2}{\partial \theta_k \partial \theta_l} s_{i,\theta} \)

17. \( \check{s}_{i,\theta} = \frac{\partial}{\partial \theta} s_{i,\theta} \) is the \( p \times 1 \) vector of partial derivatives \( \check{s}_{i,\theta}^{(k)} \).

18. \( \check{s}_{i,\theta} = \frac{\partial^2}{\partial \theta^2} s_{i,\theta} \) is the \( p \times p \) matrix of partial derivatives \( \check{s}_{i,\theta}^{(kf)} \).

19. \( H(\theta; g_i) = v_n^2(f_{i,\theta}, g_i) = \left\| s_{i,\theta} - g_i^{1/2} \right\|_2^2 = 2 - 2 \int s_{i,\theta} \cdot g_i^{1/2} \, du \)

20. \( \dot{H}(\theta; g_i) = -2 \int \dot{s}_{i,\theta} \cdot g_i^{1/2} \, du \) is a \( p \times 1 \) vector.
21. \( \tilde{H}(\theta; g_i) = -2 \int \tilde{s}_{i,\theta} \cdot g_i^{1/2} du \) is a \( p \times p \) matrix.

22. \( H(\theta; g) = \sum_{i=1}^{m} n_i H(\theta; g_i) = \sum_{i=1}^{m} n_i \left\| s_{i,\theta} - g_i^{1/2} \right\|_2^2 = \sum_{i=1}^{m} n_i \left[ 2 - 2 \int \tilde{s}_{i,\theta} \cdot g_i^{1/2} du \right] \)

23. \( \tilde{H}(\theta; g) = \sum_{i=1}^{m} n_i \tilde{H}(\theta; g_i) = \sum_{i=1}^{m} n_i \left[ -2 \int \tilde{s}_{i,\theta} \cdot g_i^{1/2} du \right] \) is a \( p \times 1 \) vector.

24. \( \tilde{H}(\theta; g) = \sum_{i=1}^{m} n_i \tilde{H}(\theta; g_i) = \sum_{i=1}^{m} n_i \left[ -2 \int \tilde{s}_{i,\theta} \cdot g_i^{1/2} du \right] \) is a \( p \times p \) matrix.

25. \( H(\theta; \hat{g}_{n_i}) = \nu_i^2(f_{i,\theta}, \hat{g}_{n_i}) = \left\| \tilde{s}_{i,\theta} - \hat{g}_{n_i}^{1/2} \right\|_2^2 = 2 - 2 \int s_{i,\theta} \cdot \hat{g}_{n_i}^{1/2} du \)

26. \( \hat{H}(\theta; \hat{g}_{n_i}) = -2 \int \tilde{s}_{i,\theta} \cdot \hat{g}_{n_i}^{1/2} du \) is a \( p \times 1 \) vector.

27. \( \hat{H}(\theta; \hat{g}_{n_i}) = -2 \int \tilde{s}_{i,\theta} \cdot \hat{g}_{n_i}^{1/2} du \) is a \( p \times p \) matrix.

28. \( H(\theta; \hat{g}_{n_i}) = \sum_{i=1}^{m} n_i H(\theta; \hat{g}_{n_i}) = \sum_{i=1}^{m} n_i \left\| s_{i,\theta} - \hat{g}_{n_i}^{1/2} \right\|_2^2 = \sum_{i=1}^{m} n_i \left[ 2 - 2 \int s_{i,\theta} \cdot \hat{g}_{n_i}^{1/2} du \right] \)

29. \( \hat{H}(\theta; \hat{g}_{n_i}) = \sum_{i=1}^{m} n_i \hat{H}(\theta; \hat{g}_{n_i}) = \sum_{i=1}^{m} n_i \left[ -2 \int \tilde{s}_{i,\theta} \cdot \hat{g}_{n_i}^{1/2} du \right] \) is a \( p \times 1 \) vector.

30. \( \hat{H}(\theta; \hat{g}_{n_i}) = \sum_{i=1}^{m} n_i \hat{H}(\theta; \hat{g}_{n_i}) = \sum_{i=1}^{m} n_i \left[ -2 \int \tilde{s}_{i,\theta} \cdot \hat{g}_{n_i}^{1/2} du \right] \) is a \( p \times p \) matrix.

31. \( l^{(i)}(\theta; y) = \log f_{i,\theta}(y) \)

Note that \( l^{(i)}(\theta; Y_{ij}) = l^{(ij)}(\theta) \) (Section 5.1, Notation 1.)

32. \( l^{(i)'}(\theta; y) = \frac{\partial}{\partial \theta} l^{(i)}(\theta; y) \) is a \( p \times 1 \) vector.

Note that \( l^{(i)''}(\theta; Y_{ij}) = l^{(ij)''}(\theta) \) (Section 5.1, Notation 7.)

33. \( l^{(i)'''}(\theta; y) = \frac{\partial^2}{\partial \theta \partial \theta^T} l^{(i)}(\theta; y) \) is a \( p \times p \) matrix.

Note that \( l^{(i)'''}(\theta; Y_{ij}) = l^{(ij)'''}(\theta) \) (Section 5.1, Notation 13.)

Note that for discrete cases, the integral signs in Notation 20, \ldots, 30 are replaced by summation signs because the integration is with respect to counting measure.

5.2.1 Uniqueness, Existence, and Consistency

This subsection concerns the existence, uniqueness, and consistency of the GMHD estimator. The results are applicable to both discrete and continuous families of distributions.

The following three assumptions are used in this subsection.
Assumptions:

(A1) \( \Theta \) is a compact subset of \( \mathbb{R}^p \).

(A2) For each \( f_{i, \theta}, \theta \neq \theta_0 \) implies \( f_{i, \theta_1} \neq f_{i, \theta_2} \) on a set of positive Lebesgue measure.

(A3) Each \( f_{i, \theta}(y) \) is continuous in \( \theta \) for almost every \( y \).

To establish the existence, uniqueness, and consistency of the GMHD estimator (Theorem 5.2.3), the following two lemmas are central.

**Lemma 5.2.1** If Assumption (A3) holds, then \( \frac{1}{n} H(\theta; g) \) is continuous in \( \theta \) for every \( g \in \Psi \).

**Proof:** Let \( \{\theta_n\}_{n=1}^{\infty} \) be any sequence of vectors of values in \( \Theta \) satisfying \( \theta_n \to \theta \in \Theta \). Then,

\[
\lim_{n \to \infty} \frac{1}{n} H(\theta_n; g) = \frac{1}{n} H(\theta; g)
\]

by the pointwise continuity assumption of the \( f_{i, \theta} \)'s and Scheffé's theorem.

**Remark:** By the pointwise continuity assumption, for every \( \{\theta_n\}_{n=1}^{\infty} \in \Theta \) satisfying \( \theta_n \to \theta \), \( f_{i, \theta_n} \to f_{i, \theta} \), which ensures the pointwise convergence of the \( f_{i, \theta} \)'s. Then, Scheffé's
Theorem states that pointwise convergence of densities implies convergence of densities in $L^1$ norm which, in turn, implies convergence of densities in the Hellinger norm. Therefore, to apply this result, all it needs is the pointwise continuity of the assumed model densities.

**Lemma 5.2.2** Suppose that $f(x)$ and $g(x)$ are two functions with a common domain $D$ and that $D$ is compact. Then,

$$\left| \inf_{x \in D} f(x) - \inf_{x \in D} g(x) \right| \leq \sup_{x \in D} |f(x) - g(x)|. $$

**Proof:** Write $\inf_x \equiv \inf_{x \in D}$ and $\sup_x \equiv \sup_{x \in D}$ for convenience. Then,

$$f(x) - g(x) \leq |f(x) - g(x)|$$

$$\rightarrow f(x) \leq g(x) + |f(x) - g(x)|$$

$$\rightarrow \inf_x f(x) \leq g(x) + |f(x) - g(x)|$$

(5.17) $$\rightarrow \inf_x f(x) \leq g(x) + \sup_x |f(x) - g(x)|. $$

Since inequality (5.17) is true for all values of $x \in D$, it implies that the inequality

$$\inf_x f(x) \leq \inf_x g(x) + \sup_x |f(x) - g(x)|$$

is also true, i.e.,

$$\inf_x f(x) - \inf_x g(x) \leq \sup_x |f(x) - g(x)|. $$

Similarly, one can get

$$\inf_x g(x) - \inf_x f(x) \leq \sup_x |f(x) - g(x)|. $$

Hence,

$$\left| \inf_x f(x) - \inf_x g(x) \right| \leq \sup_x |f(x) - g(x)|. $$

**Theorem 5.2.3** Suppose that assumptions (A1)-(A3) hold. Then,

(i) For every $g \in \Phi$, there exists $T(g) \in \Theta$ satisfying Equation (5.16).
(ii) \( T(f_0) = \theta \) uniquely for every \( \theta \in \Theta \).

(iii) If \( T(g) \) is unique, the functional \( T \) is continuous at \( g \) in the generalized Hellinger topology.

Proof:

(i) Existence. By the pointwise continuity assumption of \( f_{i,\theta} \)'s and the compactness assumption of \( \Theta \), Lemma 5.2.1 implies that for any \( g \in \Psi \), \( \frac{1}{n} H(\theta; g) \) is continuous in \( \theta \) and achieves a minimum \( T(g) \in \Theta \).

(ii) Uniqueness. This result is saying that \( T(g) = \theta_0 \) is unique if every \( g_i \) is really a member \( f_{i,\theta_0} \) of \( F_i \). To show this, observe Equation (5.16) assuming \( g_i = f_{i,\theta} \in F_i \), which becomes

\[
\sum_{i=1}^{m} \lambda_i \left\| f_{i,T(f_\theta)}^{1/2} - f_{i,T(f_\theta)}^{1/2} \right\|_2^2 = \min_{\theta \in \Theta} \sum_{i=1}^{m} \lambda_i \left\| f_{i,T(f_\theta)}^{1/2} - f_{i,T(f_\theta)}^{1/2} \right\|_2^2 = 0.
\]

The result then follows immediately from the identifiability assumption of the \( f_{i,\theta} \)'s.

(iii) Continuity of \( T \). Let each \( \{g_n\}_{n=1}^{\infty} \) be any sequence of densities satisfying

\[
\left\| g_{n_i}^{1/2} - g_i^{1/2} \right\|_2 \to 0, \text{ as } n_i \to \infty \text{ all at the same rate (i.e., } g_{n_i}^{1/2} \to g_i^{1/2} \text{ componentwise in } L^2. \)
\]

Denote \( T(g_n) \) be any one of the possible vectors of values of \( \theta \) that minimizes \( H(\theta; g_n) \). Then, for any \( \theta \in \Theta \),

\[
\left| \frac{1}{n} H(\theta; g_n) - \frac{1}{n} H(\theta; g) \right| \\
= 2 \left[ \sum_{i=1}^{m} \frac{n_i}{n} \int s_{i,\theta} \left[ g_{n_i}^{1/2} - g_i^{1/2} \right] du \right] \\
\leq 2 \left[ \sum_{i=1}^{m} \frac{n_i}{n} \int s_{i,\theta} \left[ g_{n_i}^{1/2} - g_i^{1/2} \right] du \right] \\
\leq 2 \sum_{i=1}^{m} \frac{n_i}{n} \sqrt{\int \left( g_{n_i}^{1/2} - g_i^{1/2} \right)^2 du} \quad \text{by the Cauchy-Schwarz inequality} \\
= 2 \sum_{i=1}^{m} \frac{n_i}{n} \left\| g_{n_i}^{1/2} - g_i^{1/2} \right\|_2 \\
\to 2 \sum_{i=1}^{m} \lambda_i \cdot 0 = 0
\]
as \( n \to \infty \). Because \( \Theta \) is compact, this implies that

\[
(5.18) \quad \sup_{\theta \in \Theta} \left| \frac{1}{n} H(\theta; g_n) - \frac{1}{n} H(\theta; g) \right| \to 0
\]
as \( n \to \infty \), which by Lemma 5.2.2 implies that

\[
\left| \inf_{\theta \in \Theta} \frac{1}{n} H(\theta; g_n) - \inf_{\theta \in \Theta} \frac{1}{n} H(\theta; g) \right| \to 0
\]
as \( n \to \infty \) or equivalently

\[
\left| \frac{1}{n} H(T(g_n); g_n) - \frac{1}{n} H(T(g); g) \right| \to 0
\]
as \( n \to \infty \). Equation (5.18) also implies that

\[
\left| \frac{1}{n} H(T(g_n); g_n) - \frac{1}{n} H(T(g_n); g) \right| \to 0
\]
as \( n \to \infty \), so one may conclude that

\[
(5.19) \quad \frac{1}{n} H(T(g_n); g) \to \frac{1}{n} H(T(g); g)
\]
as \( n \to \infty \). Now one shows \( T(g_n) \to T(g) = \theta_0 \) by contradiction. Suppose that \( T(g_n) \) does not converge to \( T(g) \). Then, compactness of \( \Theta \) ensures existence of a subsequence \( \{T(g_{n_i})\}_{i=1}^{\infty} \subset \{T(g_n)\} \) such that \( T(g_{n_i}) \to \theta_1 \neq T(g) \), which implies \( \frac{1}{n} H(T(g_{n_i}); g) \to \frac{1}{n} H(\theta_1; g) \) by the continuity of \( \frac{1}{n} H(\theta; g) \) as a function of \( \theta \) (Lemma 5.2.1). By (5.19), \( \frac{1}{n} H(\theta_1; g) = \frac{1}{n} H(T(g); g) \), which contradicts the assumed uniqueness of \( T(g) \).

Remarks:

(a) For the cases where \( \Theta \) is not compact, the results of Theorem 5.2.3 are still applicable if \( \Theta \) can be embedded in a compact space \( \tilde{\Theta} \). This embedding strategy comes from Beran (1977) for MHD estimation in the continuous case, but it also applies to GMHD estimation in the discrete case. In the discussion that follows Beran's (1977) Theorem 1, Beran gives an example to explain how the result of the theorem may be extended
to a location-scale family of continuous distributions whose parameter space is not compact but may be embedded in a compact set. Here is a similar example for MHD estimation in the discrete case. Suppose \( \mathcal{F} = \{f_\theta(y) : \theta \in \Theta\} \) is a Poisson family of distributions, where \( f_\theta(y) = \exp\{y \log \theta - \theta - \log(y!))\} \) and \( \Theta = [0, \infty) \). Let \( \theta' = \frac{\theta}{1+\theta} \) and write \( f_{\theta'}(y) = \exp\{y \log \left( \frac{\theta'}{1-\theta'} \right) - \left( \frac{\theta'}{1-\theta'} \right) - \log(y!))\}. \) Thus, the family can be represented as \( \mathcal{F} = \{f_{\theta'}(y) : \theta' \in \Theta'\} \), where \( \Theta' = [0, 1) \). Because \( \lim_{\theta' \to 1} H(\theta'; g) = \lim_{\theta' \to 0} \|f_{\theta'}^{1/2} - g^{1/2}\|_2^2 = \lim_{\theta \to \infty} \|f_{\theta}^{1/2} - g^{1/2}\|_2^2 = 2 \), therefore, \( H(\cdot ; g) \) can be extended to a function on \( \tilde{\Theta} = [0, 1] \), which is compact, and the extended function achieves a minimum in \( \tilde{\Theta} \). In fact the minimum must occur in \( \text{int}(\Theta') \) because \( 0 \leq H(\theta'; g) \leq 2 \) for every \( \theta' \in \Theta' \) and \( H(\theta'; g) = 2 \) is impossible. Consequently the conclusions of Beran's (1977) Theorem 1 remain valid for the Poisson model. For GMHD estimation, an embedding for the parameter space of GLM parameter \( \beta \) is discussed at the end of Section 5.2.

(b) Despite that the uniqueness of \( T(g) \) for all \( g \) satisfying \( g = f_{\theta_0} \in \mathcal{F} \) is ensured by the identifiability assumption of the \( f_i, \theta_i \)'s, establishing the uniqueness of \( T(g) \) for all \( g \) where each \( g_i \) is in some class \( \mathcal{G}_i \) is much more complicated and must be done on a case-by-case basis. One possible way to show the uniqueness of \( T(g) \) for all \( g \) in some \( \mathcal{G} \) is to establish that \( \tilde{H}(\theta; g) \) (if it exists) is positive definite for all \( g \in \mathcal{G} \).

(c) Because \( T(g) \) is unique when the assumed model is true, the continuity of the GMHD functional \( T \) at \( g \) is ensured under the model. Hence, if \( g = f_{\theta_0} \in \mathcal{F} \), the consistency of the functional \( T(\hat{g}_n) \overset{p}{\longrightarrow} T(g) \) depends on the consistency of the density estimators \( \hat{g}_n \overset{p}{\longrightarrow} g \) themselves.

**Corollary 5.2.4** Under the Assumptions (A1)-(A3), if \( \hat{g}_n \overset{p}{\longrightarrow} g \), then \( T(\hat{g}_n) \overset{p}{\longrightarrow} T(g) \).

**Remark:** For count data, each \( \hat{g}_{ni} \) is the empirical density estimator of \( g_i \), and therefore the condition \( \hat{g}_n \overset{p}{\longrightarrow} g \) is always true.
5.2.2 Asymptotic Normality and Efficiency

The asymptotic normality and asymptotic efficiency (under the assumed model) of the GMHD estimator is established in this subsection. Throughout this subsection, the proofs are derived assuming that the distributions of the assumed parametric families and the true distributions of the data are supported on the nonnegative integers, but the results are also applicable to any discrete distributions supported on nonnegative values.

For further developments, the following additional assumptions are necessary, in which the differentiability of the key element $\frac{1}{n}H(\theta; g)$ with respect to $\theta$ is particularly important. As in Beran (1977), one needs to impose smoothness conditions on the model $s_{i, \theta}$ to ensure that $\frac{1}{n}H(\theta; g)$ is twice differentiable with respect to $\theta$ for all $\theta \in \text{int}(\Theta)$.

Assumptions:

(A4) Each $s_{i, \theta}$ is twice differentiable with first partial derivatives $\dot{s}_{i, \theta}^{(k)}$ and second partial derivatives $\ddot{s}_{i, \theta}^{(k)}$ satisfying

(i) (Beran 1977, Lemma 1) for $k = 1, 2, \ldots, p$, $s_{i, \theta}^{(k)} \in L^2$, and $s_{i, \theta}^{(k)}$ and $\|s_{i, \theta}^{(k)}\|_2$ are continuous in $\theta$.

(ii) (Beran 1977, Lemma 2) for $k, \ell = 1, 2, \ldots, p$, $s_{i, \theta}^{(k, \ell)} \in L^2$, and $s_{i, \theta}^{(k, \ell)}$ and $\|s_{i, \theta}^{(k, \ell)}\|_2$ are continuous in $\theta$.

That is, for any specified $\theta_0 \in \Theta \subset \mathbb{R}^p$ that there exist a $p \times 1$ vector $\dot{s}_{i, \theta_0}$ with components in $L^2$ and a $p \times p$ matrix $\ddot{s}_{i, \theta_0}$ with components in $L^2$ such that for any $p \times 1$ real vector $e$ of unit euclidean length and for any scalar $\alpha$ in a neighborhood of zero,

$$s_{i, \theta_0 + \alpha e} = s_{i, \theta_0} + \alpha \cdot e^T \cdot \dot{s}_{i, \theta_0} + \alpha \cdot e^T \cdot u_\alpha$$

$$\dot{s}_{i, \theta_0 + \alpha e} = \dot{s}_{i, \theta_0} + \alpha \cdot \dddot{s}_{i, \theta_0} \cdot e + \alpha \cdot V_\alpha \cdot e,$$

where $u_\alpha$ is a $p \times 1$ vector whose components $\|u_\alpha^{(k)}\|_2 \rightarrow 0$ as $\alpha \rightarrow 0$, and $V_\alpha$ is a $p \times p$ matrix whose components $\|V_\alpha^{(k, \ell)}\|_2 \rightarrow 0$ as $\alpha \rightarrow 0$. 
Rewriting the above two equations by letting \( \theta = \theta_0 + \alpha e \), one can obtain the Taylor expansions of \( s_{i, \theta} \) about \( \theta_0 \) and of \( \dot{s}_{i, \theta} \) about \( \theta_0 \)

\[
(5.20) \quad s_{i, \theta} = s_{i, \theta_0} + \dot{s}_{i, \theta_0}(\theta - \theta_0) + u_{\alpha}(\theta - \theta_0),
\]

and

\[
(5.21) \quad \dot{s}_{i, \theta} = \dot{s}_{i, \theta_0} + \ddot{s}_{i, \theta_0}(\theta - \theta_0) + V_{\alpha}(\theta - \theta_0),
\]

where \( \alpha e = \theta - \theta_0 \); hence, each \( \| V_{\alpha}^{(k\ell)} \|_2 \to 0 \) as \( \theta - \theta_0 \to 0 \).

Note that by the Cauchy-Schwarz inequality, the above two smoothness conditions (i) and (ii) of \( \dot{s}_{i, \theta} \) ensure the twice-differentiability of \( H(\theta; g) \) with respect to \( \theta \).

(A5) \( T(g) = \theta_0 \) exists, is unique, and lies in \( \text{int} (\Theta) \subset R^p \). That is, \( \dot{H}(\theta_0; g) = 0 \) for some \( \theta_0 \in \text{int} (\Theta) \subset R^p \).

(A6) \( \ddot{H}(\theta_0; g) \) is nonsingular.

(A7) The functional \( T(\cdot) \) is continuous at \( g \) in the generalized Hellinger topology.

(A8) \( \dot{s}_{i, \theta_0} \in L^1 \) componentwise for all \( i = 1, 2, \ldots, m \).

(A9) The assumption for information matrices. (See Assumptions (A2) and (A3) in Section 5.1.)

Under the Assumption (A4), observe the following facts.

Facts:

(F1) \[
\dot{s}_{i, \theta}(y) = \frac{\partial}{\partial \theta} f_{i, \theta}^{1/2}(y) = \frac{1}{2} f_{i, \theta}^{1/2}(y) \cdot \frac{\partial \log f_{i, \theta}(y)}{\partial \theta} = \frac{1}{2} s_{i, \theta}(y) \cdot l^{(i)'}(\theta; y)
\]

(F2) \[
\ddot{s}_{i, \theta}(y) = \frac{\partial}{\partial \theta} \dot{s}_{i, \theta}(y) = \frac{1}{2} \left[ \frac{1}{2} s_{i, \theta}(y) \cdot l^{(i)'}(\theta; y) \cdot l^{(i)'}(\theta; y)^\top + s_{i, \theta}(y) \cdot \frac{\partial^2 \log f_{i, \theta}(y)}{\partial \theta \partial \theta^\top} \right]
\]
\[
= \left[ \frac{1}{4} l^{(i)'}(\theta; y) \cdot l^{(i)'}(\theta; y)^\top + \frac{1}{2} l^{(i)''}(\theta; y) \right] s_{i, \theta}(y)
\]
(F3) For any \( g \in \Phi \),

\[
\dot{H}(\theta; g_i) = -2 \int \dot{s}_{i,\theta}(y) \cdot g_i^{1/2}(y)dy
\]

\[
= - \int l^{(i)'}(\theta; y) \cdot s_{i,\theta}(y) \cdot g_i^{1/2}(y)dy \quad \text{by (F1)},
\]

and

\[
\frac{1}{n} \dot{H}(\theta; g) = \sum_{i=1}^{m} \frac{n_i}{n} \dot{H}(\theta; g_i)
\]

\[
= \sum_{i=1}^{m} \frac{n_i}{n} \left[ - \int l^{(i)'}(\theta; y) \cdot s_{i,\theta}(y) \cdot g_i^{1/2}(y)dy \right].
\]

If \( g = f_{\theta_0} \in \mathcal{F} \), then

\[
\dot{H}(\theta_0; g_i) = -\mathcal{E}_{\theta_0} \left[ l^{(i)'}(\theta_0; Y_{ij}) \right] = -\mathcal{E}_{\theta_0} \left[ l^{(ij)'}(\theta_0) \right] = 0 \quad \text{by Equation (5.3)},
\]

and

\[
\frac{1}{n} \dot{H}(\theta_0; g) = \sum_{i=1}^{m} \frac{n_i}{n} \left\{ -\mathcal{E}_{\theta_0} \left[ l^{(ij)'}(\theta_0) \right] \right\} = 0.
\]

(F4) For any \( g \in \Phi \),

\[
\ddot{H}(\theta; g_i) = -2 \int \ddot{s}_{i,\theta}(y) \cdot g_i^{1/2}(y)dy
\]

\[
= - \int \left[ \frac{1}{2} l^{(i)''}(\theta; y) \cdot l^{(i)'}(\theta; y)^T + l^{(i)''}(\theta; y) \right] s_{i,\theta}(y) \cdot g_i^{1/2}(y)dy \quad \text{by (F2)},
\]

and

\[
\frac{1}{n} \ddot{H}(\theta; g) = \sum_{i=1}^{m} \frac{n_i}{n} \ddot{H}(\theta; g_i)
\]

\[
= \sum_{i=1}^{m} \frac{n_i}{n} \left\{ - \int \left[ \frac{1}{2} l^{(i)''}(\theta; y) \cdot l^{(i)'}(\theta; y)^T + l^{(i)''}(\theta; y) \right] s_{i,\theta}(y) \cdot g_i^{1/2}(y)dy \right\}.
\]

If \( g = f_{\theta_0} \in \mathcal{F} \), then

\[
\ddot{H}(\theta; g_i) = -\frac{1}{2} \mathcal{E} \left[ l^{(i)''}(\theta; Y_{ij}) \cdot l^{(i)'}(\theta; Y_{ij})^T \right] + \mathcal{E} \left[ -l^{(i)''}(\theta; Y_{ij}) \right]
\]

\[
= -\frac{1}{2} \mathcal{E} \left[ l^{(ij)''}(\theta) \cdot l^{(ij)'}(\theta)^T \right] + \mathcal{E} \left[ -l^{(ij)''}(\theta) \right]
\]

\[
= -\frac{1}{2} 1^{(i)}(\theta) + 1^{(i)}(\theta) \quad \text{by Equation (5.4)}
\]

\[
= \frac{1}{2} 1^{(i)}(\theta),
\]
and
\[
\frac{1}{n} \hat{H}(\theta; g) = \sum_{i=1}^{m} \frac{r_i}{n} \left\{ \frac{1}{2} l_1^{(i)}(\theta) \right\} \\
\rightarrow \sum_{i=1}^{m} \lambda_i \left\{ \frac{1}{2} l_1^{(i)}(\theta) \right\} \quad \text{as } n \to \infty \\
= \frac{1}{2} l_1(\theta).
\]

\[(F5)\]
\[
\int \hat{s}_i \theta(y) \cdot \hat{s}_i \theta(y)^T dy \\
= \int \left[ \frac{1}{2} s_i \theta(y) \cdot l^{(i)'}(\theta; y) \right] \left[ \frac{1}{2} s_i \theta(y) \cdot l^{(i)'}(\theta; y) \right]^T dy \quad \text{by (F1)} \\
= \frac{1}{4} \mathcal{E} \left[ l^{(i)'}(\theta; Y_{ij}) \cdot l^{(i)'}(\theta; Y_{ij})^T \right] \\
= \frac{1}{4} \mathcal{E} \left[ l^{(ij)'}(\theta) \cdot l^{(ij)'}(\theta)^T \right] \\
= \frac{1}{4} l_1^{(i)}(\theta) \quad \text{by Equation (5.4)}.
\]

\[(F6)\]
\[
\int \hat{s}_i \theta(y) \cdot s_i \theta(y) dy \\
= \int \left[ \frac{1}{4} l^{(i)'}(\theta; y) \cdot l^{(i)'}(\theta; y)^T + \frac{1}{2} l^{(i)''}(\theta; y) \right] f_i \theta(y) dy \quad \text{by (F2)} \\
= \frac{1}{4} \mathcal{E} \left[ l^{(i)'}(\theta; Y_{ij}) \cdot l^{(i)'}(\theta; Y_{ij})^T \right] - \frac{1}{2} \mathcal{E} \left[ -l^{(i)''}(\theta; Y_{ij}) \right] \\
= \frac{1}{4} \mathcal{E} \left[ l^{(ij)'}(\theta) \cdot l^{(ij)'}(\theta)^T \right] - \frac{1}{2} \mathcal{E} \left[ l^{(ij)''}(\theta) \right] \\
= \frac{1}{4} l_1^{(i)}(\theta) - \frac{1}{2} l_1^{(i)}(\theta) \quad \text{by Equation (5.4)} \\
= -\frac{1}{4} l_1^{(i)}(\theta)
\]

\[(F7)\] \(T(g) = \theta_0\) is a zero of \(\hat{H}(\theta; g)\). That is, \(\hat{H}(T(g); g) = \hat{H}(\theta_0; g) = 0\). (Note that, \(\hat{H}(\theta_0; g) = 0\) does not imply \(\hat{H}(\theta_0; g_n) = 0\).)

\[(F8)\] \(T(\hat{g}_n)\) is a zero of \(\hat{H}(\theta; \hat{g}_n)\). That is, \(\hat{H}(T(\hat{g}_n); \hat{g}_n) = 0\). (Note that, \(\hat{H}(T(\hat{g}_n); \hat{g}_n) = 0\) does not imply \(\hat{H}(T(\hat{g}_n); \hat{g}_n) = 0\).)
To show the asymptotic normality of the GMHD estimator (Theorem 5.2.7), the following two lemmas are needed.

**Lemma 5.2.5** Let $f_{i,\theta}$ and $g_i$ for all $i = 1, 2, \ldots, m$ be supported on $\{0, 1, 2, \ldots\}$ and $\hat{g}_n$, be the empirical density estimator of $g_i$ defined by Equation (2.2). Suppose that Assumption (A4) Equation (5.20) holds at $\theta_0$ and that $\hat{H}(\theta_0; g) = 0$. If Assumption (A8) holds, then

$$\frac{1}{n} \hat{H}(\theta_0; \hat{g}_n) = \sum_{i=1}^{m} \left( \frac{n_i}{n} \right) \left[ \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{s}_i, \theta_0 (Y_{ij}) \cdot g_i^{-1/2} (Y_{ij}) \right] + o_p(n^{-1/2}).$$

Three tools below are useful in establishing Lemma 5.2.5.

**Tools:**

1. $g_i(y) (1 - g_i(y))$ for all $i = 1, 2, \ldots, m$.

**Proof:**

(T1) $E \left[ \left( \hat{g}_n^{1/2} (y) - g_i^{1/2} (y) \right)^2 \right] \leq \sqrt{g_i(y)(1 - g_i(y))} / n_i$ for all $i = 1, 2, \ldots, m$.

Proof:

$$E \left[ \left( \hat{g}_n^{1/2} (y) - g_i^{1/2} (y) \right)^2 \right] \leq E \left[ \left( \hat{g}_n^{1/2} (y) - g_i^{1/2} (y) \right) \left( \hat{g}_n^{1/2} (y) + g_i^{1/2} (y) \right) \right]$$

$$= E \left[ \left( \hat{g}_n (y) - g_i (y) \right) \right]$$

$$\leq E \left[ \left( \hat{g}_n (y) - g_i (y) \right)^2 \right]^{1/2} \text{ by Liapounov's inequality}$$

$$= \left[ g_i(y)(1 - g_i(y)) \right]^{1/2} / n_i$$

because $\hat{g}_n (y)$ is a binomial proportion random variable.

(T2) $n_i^{1/4} \left( \hat{g}_n^{1/2} (y) - g_i^{1/2} (y) \right) \overset{p}{\to} 0$ for all $i = 1, 2, \ldots, m$.

Proof: First, recognize that

$$\sqrt{n_i} (\hat{g}_n (y) - g_i (y)) \overset{D}{\to} N \left( 0, g_i(y)(1 - g_i(y)) \right).$$

Let $\psi(x) = x^{1/2}$. Then, by Lehmann’s (1983) Theorem 5.1.5, Equation (5.22) implies

$$\sqrt{n_i} (\psi(\hat{g}_n (y)) - \psi(g_i (y))) \overset{D}{\to} N \left( 0, \left[ \psi'(g_i (y)) \right]^2 g_i(y)(1 - g_i(y)) \right).$$
This is,
\[(5.23) \quad \sqrt{n_i} \left( g_i^{1/2}(y) - g_i^{1/2}(y) \right) \xrightarrow{p} N \left( 0, \frac{(1 - g_i(y))}{4} \right). \]

The result follows by Slutsky's theorem (Lehmann 1983, Theorem 5.1.4) when multiplying the left hand side of (5.23) by \( n^{-1/4} \) because \( n^{-1/4} \xrightarrow{p} 0. \)

(T3) \( E \left[ \sqrt{n_i} \left( g_i^{1/2}(y) - g_i^{1/2}(y) \right)^2 \right] \xrightarrow{} 0 \) as \( n_i \to \infty \) for all \( i = 1, 2, \ldots, m. \)

Proof: Let \( W_n_i(y) = n_i^{1/4} \left( g_i^{1/2}(y) - g_i^{1/2}(y) \right) \). Then, by Tool (T2) \( W_n_i(y) \xrightarrow{p} 0 \), which implies \( W_n_i^2(y) = n_i \left( g_i^{1/2}(y) - g_i^{1/2}(y) \right)^2 \xrightarrow{p} 0 \) by Billingsley (1986), page 344, Corollary 2 of Theorem 25.7. Now, if it can be shown that \( W_n_i^2(y) \) is uniformly integrable, then by Billingsley (1986), page 348, Theorem 25.12, \( W_n_i^2(y) \xrightarrow{p} 0 \) implies \( E \left[ W_n_i^2(y) \right] \xrightarrow{} 0 \), i.e., \( \lim_{n_i \to \infty} E \left[ W_n_i^2(y) \right] = E \left[ \lim_{n_i \to \infty} W_n_i^2(y) \right] = 0. \) To establish the uniform integrability of \( W_n_i^2(y) \), Billingsley's (1986) Theorem 25.12 indicates that it suffices to prove \( \sup_{n_i} E \left[ \left| W_n_i^2(y) \right|^{1+\varepsilon} \right] < \infty \) for some positive \( \varepsilon \). The proof is now given.

\[
E \left[ \left| W_n_i^2(y) \right|^{1+\varepsilon} \right] = E \left[ n_i^{1/2} \left( g_i^{1/2}(y) - g_i^{1/2}(y) \right)^2 \right]^{1+\varepsilon} \\
= n_i^{1+\varepsilon/2} E \left[ \left( g_i^{1/2}(y) - g_i^{1/2}(y) \right)^2 \right]^{1+\varepsilon} \\
\leq n_i^{1+\varepsilon/2} E \left[ \left( g_i^{1/2}(y) - g_i^{1/2}(y) \right) \left( g_i^{1/2}(y) + g_i^{1/2}(y) \right) \right]^{1+\varepsilon} \\
= n_i^{1+\varepsilon/2} E \left[ \left( g_i^{1/2}(y) - g_i^{1/2}(y) \right) \right]^{1+\varepsilon} \\
\leq n_i^{1+\varepsilon/2} E \left[ \left( g_i^{1/2}(y) - g_i^{1/2}(y) \right)^2 \right]^{1+\varepsilon} \\
= n_i^{1+\varepsilon/2} E \left[ \left( g_i^{1/2}(y) - g_i^{1/2}(y) \right)^2 \right]^{1+\varepsilon} \\
= n_i^{1+\varepsilon/2} E \left[ \left( g_i(y)(1 - g_i(y)) \right) \right]^{1+\varepsilon} \\
= \left[ g_i(y)(1 - g_i(y)) \right]^{1+\varepsilon/2} < \infty.
\]
Proof of Lemma 5.2.5: First note that

\[
\frac{1}{n_i} \sum_{j=1}^{n_i} \hat{s}_{i, \theta_0} (Y_{ij}) \cdot g_i^{-1/2}(Y_{ij}) = \sum_{y=0}^{\infty} \hat{s}_{i, \theta_0}(y) \cdot g_i^{-1/2}(y) \cdot \hat{g}_n(y)
\]

because \( \hat{g}_n(y) \) ignores the terms that are empty cells. Now, consider the remainder term

\[
R_n = - \frac{1}{n} \hat{H}(\theta_0; \hat{g}_n) - \sum_{i=1}^{m} \left( \frac{n_i}{n} \right) \left[ \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{s}_{i, \theta_0} (Y_{ij}) \cdot g_i^{-1/2}(Y_{ij}) \right]
\]

\[
= - \sum_{i=1}^{m} \left( \frac{n_i}{n} \right) \sum_{y=0}^{\infty} \hat{s}_{i, \theta_0}(y) \left\{ g_i^{-1/2}(y) \cdot \hat{g}_n(y) - 2 \hat{g}_{n_i}^{-1/2}(y) \right\}
\]

\[
= - \sum_{i=1}^{m} \left( \frac{n_i}{n} \right) \sum_{y=0}^{\infty} \hat{s}_{i, \theta_0}(y) \cdot g_i^{-1/2}(y) \left\{ \hat{g}_n(y) - 2 g_i^{1/2}(y) \cdot \hat{g}_{n_i}^{-1/2}(y) \right\}
\]

\[
= - \sum_{i=1}^{m} \left( \frac{n_i}{n} \right) \sum_{y=0}^{\infty} \hat{s}_{i, \theta_0}(y) \cdot g_i^{-1/2}(y) \left( \hat{g}_{n_i}^{-1/2}(y) - g_i^{1/2}(y) \right)^2,
\]

where the 4th equation and the 5th equation are equal because

\[
\sum_{i=1}^{m} \left( \frac{n_i}{n} \right) \left[ -2 \int \hat{s}_{i, \theta_0}(y) \cdot g_i^{1/2}(y) dy \right] = \frac{1}{n} \hat{H}(\theta_0; g) = 0
\]

by assumption. The result follows by showing that \( \sqrt{n} R_n \xrightarrow{p} 0 \) as \( n \to \infty \).

Let \( R_{n,k} \) be the \( k \)th component of \( R_n \). Then, by Markov's inequality,

\[
P \left( \left| \sqrt{n} R_{n,k} \right| > \varepsilon \right) \leq \frac{E \left[ \left| \sqrt{n} R_{n,k} \right| \right]}{\varepsilon},
\]

so it suffices to show that \( E \left[ \left| \sqrt{n} R_{n,k} \right| \right] \to 0 \), as \( n \to \infty \). Now,

\[
E \left[ \left| \sqrt{n} R_{n,k} \right| \right] = E \left[ \sum_{i=1}^{m} \sqrt{\frac{n_i}{n}} \sum_{y=0}^{\infty} \hat{s}_{i, \theta_0}^{(k)}(y) \cdot g_i^{-1/2}(y) \left( \sqrt{n_i} \left( \hat{g}_{n_i}^{1/2}(y) - g_i^{1/2}(y) \right)^2 \right) \right]
\]

\[
\leq E \left[ \sum_{i=1}^{m} \sqrt{\frac{n_i}{n}} \sum_{y=0}^{\infty} \left| \hat{s}_{i, \theta_0}^{(k)}(y) \right| \cdot g_i^{-1/2}(y) \left( \sqrt{n_i} \left( \hat{g}_{n_i}^{1/2}(y) - g_i^{1/2}(y) \right)^2 \right) \right]
\]

\[
= \sum_{i=1}^{m} \sqrt{\frac{n_i}{n}} \sum_{y=0}^{\infty} \left| \hat{s}_{i, \theta_0}^{(k)}(y) \right| \cdot g_i^{-1/2}(y) \sqrt{n_i} \left( \hat{g}_{n_i}^{1/2}(y) - g_i^{1/2}(y) \right)^2.
\]

(5.24)
Since

\begin{align*}
(5.25) \quad \left| \hat{s}_{i, \theta_0}(y) \right| g_i^{-1/2}(y) E \left[ \sqrt{n_i} \left( \hat{g}_{n_i}^{1/2}(y) - g_i^{1/2}(y) \right) \right] \\
= \left| \hat{s}_{i, \theta_0}(y) \right| g_i^{-1/2}(y) \sqrt{n_i} \left( 1 - g_i(y) \right) / n_i \text{ by Tool (T1)} \\
= \left| \hat{s}_{i, \theta_0}(y) \right| g_i^{-1/2}(y) (1 - g_i(y))^{1/2} \\
\leq \left| \hat{s}_{i, \theta_0}(y) \right|,
\end{align*}

\( \hat{s}_{i, \theta_0} \) is integrable by Assumption (A8), and (5.25) goes to 0 as \( n_i \to \infty \) (provided that \( E \left[ \sqrt{n_i} \left( \hat{g}_{n_i}^{1/2}(y) - g_i^{1/2}(y) \right) \right] \to 0 \) as \( n_i \to \infty \) by Tool (T3)), therefore, according to the Lebesgue dominated convergence theorem,

\[
\sum_{y=0}^{\infty} \left| \hat{s}_{i, \theta_0}(y) \right| g_i^{-1/2}(y) E \left[ \sqrt{n_i} \left( \hat{g}_{n_i}^{1/2}(y) - g_i^{1/2}(y) \right) \right] \to 0
\]

as \( n_i \to \infty \) (all at the same rate), which implies (5.24) tends to 0 as \( n \to \infty \), and hence \( E \left[ \sqrt{n} R_{n,k} \right] \to 0 \) as \( n \to \infty \).

\[ \Box \]

**Lemma 5.2.6** Suppose that Assumption (A4) holds for all \( \theta \in \text{int} (\Theta) \) and suppose that Assumptions (A5)-(A9) are satisfied. Then, for every sequence of vectors of densities \( \{ \hat{g}_{n_i} \}_{i=1}^{\infty} = \{ (\hat{g}_{n_{i,v}}, \hat{g}_{n_{i,v}}, \ldots, \hat{g}_{n_{i,m}}) \}_{i=1}^{\infty} \) satisfying \( \left\| \hat{g}_{n_i}^{1/2} - g_i^{1/2} \right\|_2 \to 0 \) for all \( i = 1, 2, \ldots, m \),

\[
T(\hat{g}_n) - \theta_0 = \left[ \sum_{i=1}^{m} \lambda_i \hat{H}(\theta_0; g_i) \right]^{-1} \left[ -\frac{1}{n} \hat{H}(\theta_0; \hat{g}_n) \right] + o_p(n^{-1/2}).
\]

To establish Lemma 5.2.6, the following two tools are necessary.

**Tools:**

(T4) Under the assumptions of Lemma 5.2.6,

\[
\sqrt{n} \left[ -\frac{1}{n} \hat{H}(\theta_0; \hat{g}_n) \right] \xrightarrow{D} N_p(0, V_{\theta_0}),
\]

where

\[
V_{\theta_0} = \frac{1}{4} I_1(\theta_0) - \frac{1}{4} \sum_{i=1}^{m} \lambda_i \left[ \hat{H}(\theta_0; g_i) \right] \left[ \hat{H}(\theta_0; g_i) \right]^T.
\]
for any \( g \in \Psi \) and
\[
V_{\theta_0} = \frac{1}{4} I_1(\theta_0)
\]
if \( g = f_{\theta_0} \in \mathcal{F} \).

Proof: According to Lemma 5.2.5,
\[
(5.26) \quad -\frac{1}{n} \dot{\mathcal{H}}(\theta_0; \hat{g}_n) = \sum_{i=1}^{m} \left( \frac{n_i}{n} \right) \left[ \frac{1}{n_i} \sum_{j=1}^{n_i} \dot{s}_{i,\theta_0}(Y_{ij}) \cdot g_i^{-1/2}(Y_{ij}) \right] + o_p(n^{-1/2}).
\]
Let \( v_{i,\theta_0} = \mathcal{E}_{g_i} \left[ \dot{s}_{i,\theta_0}(Y_{ij}) \cdot g_i^{-1/2}(Y_{ij}) \right] = \sum_{y=0}^\infty \dot{s}_{i,\theta_0}(y) \cdot g_i^{1/2}(y) = -\frac{1}{2} \dot{\mathcal{H}}(\theta_0; g_i) \). Then, Equation (5.26) can be written to
\[
-\frac{1}{n} \dot{\mathcal{H}}(\theta_0; \hat{g}_n) = \sum_{i=1}^{m} \left( \frac{n_i}{n} \right) \left[ \frac{1}{n_i} \sum_{j=1}^{n_i} \dot{s}_{i,\theta_0}(Y_{ij}) \cdot g_i^{-1/2}(Y_{ij}) - v_{i,\theta_0} \right] + o_p(n^{-1/2})
\]
because \( \sum_{i=1}^{m} \left( \frac{n_i}{n} \right) v_{i,\theta_0} = \sum_{i=1}^{m} \left( \frac{n_i}{n} \right) \left[ -\frac{1}{2} \dot{\mathcal{H}}(\theta_0; g_i) \right] = -\frac{1}{2} \dot{\mathcal{H}}(\theta_0; g) = 0 \) by assumption.

Now, by multivariate central limit theorem,
\[
\sqrt{n_i} \left[ \frac{1}{n_i} \sum_{j=1}^{n_i} \dot{s}_{i,\theta_0}(Y_{ij}) \cdot g_i^{-1/2}(Y_{ij}) - v_{i,\theta_0} \right] \xrightarrow{D} N_p \left( 0, \text{COV}_{g_i} \left[ \dot{s}_{i,\theta_0}(Y_{ij}) \cdot g_i^{-1/2}(Y_{ij}) \right] \right),
\]
where
\[
\text{COV}_{g_i} \left[ \dot{s}_{i,\theta_0}(Y_{ij}) \cdot g_i^{-1/2}(Y_{ij}) \right] = \mathcal{E}_{g_i} \left[ \dot{s}_{i,\theta_0}(Y_{ij}) \cdot \dot{s}_{i,\theta_0}^T(Y_{ij}) \cdot g_i^{-1}(Y_{ij}) \right] - v_{i,\theta_0} \cdot v_{i,\theta_0}^T
= \sum_{y=0}^\infty \dot{s}_{i,\theta_0}(y) \cdot \dot{s}_{i,\theta_0}^T(y) - v_{i,\theta_0} \cdot v_{i,\theta_0}^T
= \frac{1}{4} I^{(i)}_1(\theta_0) - v_{i,\theta_0} \cdot v_{i,\theta_0}^T \quad \text{by (F5)}
= \frac{1}{4} I^{(i)}_1(\theta_0) - \frac{1}{4} \left[ \dot{\mathcal{H}}(\theta_0; g_i) \right] \left[ \dot{\mathcal{H}}(\theta_0; g_i) \right]^T.
\]
Therefore,
\[
\sqrt{n} \left[ -\frac{1}{n} \dot{\mathcal{H}}(\theta_0; \hat{g}_n) \right] = \sum_{i=1}^{m} \sqrt{\frac{n_i}{n}} \left[ \sqrt{n_i} \left( \frac{1}{n_i} \sum_{j=1}^{n_i} \dot{s}_{i,\theta_0}(Y_{ij}) \cdot g_i^{-1/2}(Y_{ij}) - v_{i,\theta_0} \right) \right] + o_p(1)
\]
by Slutsky's theorem.

If \( g = f_{\theta_0} \in \mathcal{F} \), \( \hat{H}(\theta_0; g_i) = 0 \) for all \( i = 1, 2, \ldots, m \) by (F3), so

\[
\sqrt{n} \left[ -\frac{1}{n} \hat{H}(\theta_0; \hat{g}_n) \right] \xrightarrow{D} N_p \left( 0, \frac{1}{4} I_1(\theta_0) \right).
\]

\( \Box \)

(T5) Under the assumptions of Lemma 5.2.6,

\[
(5.27) \quad \frac{1}{n} \check{H}(\theta_0; \hat{g}_n) \xrightarrow{p} \sum_{i=1}^{m} \lambda_i \check{H}(\theta_0; g_i) \quad \text{(componentwise),}
\]

for any \( g \in \Psi \) and

\[
(5.28) \quad \frac{1}{n} \check{H}(\theta_0; \hat{g}_n) \xrightarrow{p} \frac{1}{2} I_1(\theta_0) \quad \text{(componentwise),}
\]

if \( g = f_{\theta_0} \in \mathcal{F} \).

Proof: Because \( \frac{1}{n} \check{H}(\theta_0; g) = \sum_{i=1}^{m} \frac{n_i}{n} \check{H}(\theta_0; g_i) \xrightarrow{p} \sum_{i=1}^{m} \lambda_i \check{H}(\theta_0; g_i) \), it suffices to show that for every fixed \( k \) and \( \ell \), the components of matrices \( \check{H}(\theta_0; g_n) \) and \( \check{H}(\theta_0; g) \) satisfy

\[
P_{\theta_0} \left( \left| \frac{1}{n} \check{H}_{k\ell}(\theta_0; \hat{g}_n) - \frac{1}{n} \check{H}_{k\ell}(\theta_0; g) \right| < \varepsilon \right) \rightarrow 1, \quad \text{as } n \to \infty.
\]

Now,

\[
\frac{1}{n} \check{H}_{k\ell}(\theta_0; \hat{g}_n) - \frac{1}{n} \check{H}_{k\ell}(\theta_0; g)
\]

\[
= \left| \sum_{i=1}^{m} \left\{ -2 \int s^{(k\ell)}_{i, \theta_0}(y) \left[ \hat{g}_{n_i}^{1/2}(y) - g_i^{1/2}(y) \right] dy \right\} \right|
\]

\[
\leq 2 \sum_{i=1}^{m} \frac{n_i}{n} \left| \int s^{(k\ell)}_{i, \theta_0}(y) \left[ \hat{g}_{n_i}^{1/2}(y) - g_i^{1/2}(y) \right] dy \right|
\]

\[
\leq 2 \sum_{i=1}^{m} \frac{n_i}{n} \left\| s^{(k\ell)}_{i, \theta_0}(y) \right\|_2 \left\| \hat{g}_{n_i}^{1/2}(y) - g_i^{1/2}(y) \right\|_2
\]

\[
xrightarrow{p} 2 \sum_{i=1}^{m} \lambda_i \left\| s^{(k\ell)}_{i, \theta_0}(y) \right\|_2 \cdot 0 = 0.
\]
Here, \( \tilde{s}_{i,\theta_0}^{(k\ell)}(y) \in L^2 \) for all \( k, \ell = 1, 2, \ldots, p \) by assumption, so \( \left\| \tilde{s}_{i,\theta_0}^{(k\ell)}(y) \right\|_2 < \infty \) for all \( k, \ell = 1, 2, \ldots, p \).

If \( g = f_{\theta_0} = \mathcal{F} \),

\[
\sum_{i=1}^{m} \lambda_i \tilde{H}(\theta_0; g_i) = \sum_{i=1}^{m} \lambda_i \left[ \frac{1}{2} l_1^{(i)}(\theta_0) \right] = \frac{1}{2} l_1(\theta_0)
\]

by (F4), which gives the result (5.28).

**Proof of Lemma 5.2.6:** Recall from (F8) that \( T(\hat{g}_n) \) is a zero of \( \tilde{H}(\theta; \hat{g}_n) \). Then,

\[
0 = \frac{1}{n} \tilde{H}(T(\hat{g}_n); \hat{g}_n) = \sum_{i=1}^{n} \frac{n_i}{n} \left[ -2 \int \hat{s}_{i,T(\hat{g}_n)}(y) \cdot \hat{g}_{n_i}^{1/2}(y) dy \right]
\]

\[
= \sum_{i=1}^{n} \frac{n_i}{n} \left\{ -2 \int [\hat{s}_{i,\theta_0}(y) + \hat{s}_{i,\theta_0}(y)(T(\hat{g}_n) - \theta_0) + V_n(y)(T(\hat{g}_n) - \theta_0)] \hat{g}_{n_i}^{1/2}(y) dy \right\}
\]

\[
= \sum_{i=1}^{n} \frac{n_i}{n} \left\{ -2 \int \hat{s}_{i,\theta_0}(y) \cdot \hat{g}_{n_i}^{1/2}(y) dy 
+ \left[ -2 \int \hat{s}_{i,\theta_0}(y) \cdot \hat{g}_{n_i}^{1/2}(y) dy - 2 \int V_n(y) \cdot \hat{g}_{n_i}^{1/2}(y) dy \right] (T(\hat{g}_n) - \theta_0) \right\}
\]

\[
= \frac{1}{n} \tilde{H}(\theta_0; \hat{g}_n) + \left\{ \frac{1}{n} \tilde{H}(\theta_0; \hat{g}_n) - \sum_{i=1}^{n} \frac{n_i}{n} \left[ 2 \int V_n(y) \cdot \hat{g}_{n_i}^{1/2}(y) dy \right] \right\} (T(\hat{g}_n) - \theta_0),
\]

where every component of \( V_n \) satisfying \( \left\| V_n^{(k\ell)} \right\|_2 \to 0 \) as \( n \to \infty \) because \( T(\hat{g}_n) \to \theta_0 \). Let

\[ C_n = \sum_{i=1}^{n} \frac{n_i}{n} \left[ -2 \int V_n(y) \cdot \hat{g}_{n_i}^{1/2}(y) dy \right] \]

and rearrange the above expression, it yields

\[
(T(\hat{g}_n) - \theta_0) = \left[ \frac{1}{n} \tilde{H}(\theta_0; \hat{g}_n) + C_n \right]^{-1} \left[ -\frac{1}{n} \tilde{H}(\theta_0; \hat{g}_n) \right]
\]

\[
= \left\{ \sum_{i=1}^{m} \lambda_i \tilde{H}(\theta_0; g_i) \right\}^{-1} \left[ -\frac{1}{n} \tilde{H}(\theta_0; \hat{g}_n) \right]
+ D_n \left[ -\frac{1}{n} \tilde{H}(\theta_0; \hat{g}_n) \right],
\]

where

\[
D_n = \left[ \frac{1}{n} \tilde{H}(\theta_0; \hat{g}_n) + C_n \right]^{-1} - \left\{ \sum_{i=1}^{m} \lambda_i \tilde{H}(\theta_0; g_i) \right\}^{-1}.
\]

It remains to show that \( D_n \left[ -\frac{1}{n} \tilde{H}(\theta_0; \hat{g}_n) \right] = o_p(n^{-1/2}) \) and that is equivalent to prove that

\[
\sqrt{n} D_n \left[ -\frac{1}{n} \tilde{H}(\theta_0; \hat{g}_n) \right] \xrightarrow{p} 0
\]
as \( n \to \infty \). First note that \( C_n \) satisfies \( \|C_n\|_2 \to 0 \) as \( n \to \infty \) because \( V_n \) satisfies \( \|V_n\|_2 \to 0 \) and \( \frac{\hat{r}_n}{\lambda_1} \to 1 \) as \( n \to \infty \). Then, use the result (5.27) from Tool (T5) and apply Billingsley (1986), page 344, Corollary 2 of Theorem 25.7 to imply that

\[
\left[ \frac{1}{n} \bar{H}(\theta_0; \hat{g}_n) \right]^{-1} \xrightarrow{p} \left[ \sum_{i=1}^m \lambda_i \bar{H}(\theta_0; g_i) \right]^{-1}.
\]

The result, \( D_n \) satisfies \( \|D_n\|_2 \to 0 \) as \( n \to \infty \), then follows. Finally, the proof is completed according Tool (T4) and the Slutsky's theorem.

**Remark:** An immediate result following Lemma 5.2.6 is the asymptotic equivalence of the GMHD estimator and the GML estimator under the assumed model, which is now described. First note that, by Lemma 5.2.5,

\[
\sqrt{n} \left[ -\frac{1}{n} \bar{H}(\theta_0; \hat{g}_n) \right] = \sqrt{n} \left\{ \sum_{i=1}^m \left( \frac{n_i}{n} \right) \left[ \frac{1}{n_i} \sum_{j=1}^{n_i} \tilde{s}_i,\theta_0(Y_{ij}) \cdot g_i^{-1/2}(Y_{ij}) \right] + o_p(n^{-1/2}) \right\}
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^m \sum_{j=1}^{n_i} \tilde{s}_i,\theta_0(Y_{ij}) \cdot g_i^{-1/2}(Y_{ij}) + o_p(1)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^m \sum_{j=1}^{n_i} \left[ \frac{1}{2} s_i,\theta_0(Y_{ij}) \cdot l^{(ij)\top}(\theta_0) \right] \cdot g_i^{-1/2}(Y_{ij}) + o_p(1) \quad \text{by (F1)}
\]

which is equal to \( \frac{1}{2\sqrt{n}} \sum_{i=1}^m \sum_{j=1}^{n_i} l^{(ij)\top}(\theta_0) + o_p(1) = \frac{1}{2\sqrt{n}} l'(\theta_0) + o_p(1) \) if \( g = f_{\theta_0} \in \mathcal{F} \).

Furthermore, by result (5.29), \( \sum_{i=1}^m \lambda_i \bar{H}(\theta_0; g_i) = \frac{1}{2} I_1(\theta_0) \) if \( g = f_{\theta_0} \in \mathcal{F} \). Therefore, following Lemma 5.2.6, if \( g = f_{\theta_0} \in \mathcal{F} \),

\[
\sqrt{n} (T(\hat{g}_n) - \theta_0) = \left[ \sum_{i=1}^m \lambda_i \bar{H}(\theta_0; g_i) \right]^{-1} \left\{ \sqrt{n} \left[ -\frac{1}{n} \bar{H}(\theta_0; \hat{g}_n) \right] \right\} + o_p(1)
\]

\[
= \left[ \frac{1}{2} I_1(\theta_0) \right]^{-1} \left[ \frac{1}{2\sqrt{n}} l'(\theta_0) + o_p(1) \right] + o_p(1)
\]

\[
= [I_1(\theta_0)]^{-1} \left( \frac{1}{\sqrt{n}} l'(\theta_0) \right) + [I_1(\theta_0)]^{-1} o_p(1) + o_p(1)
\]

\[
= [I_1(\theta_0)]^{-1} \left( \frac{1}{\sqrt{n}} l'(\theta_0) \right) + o_p(1)
\]
This is saying that \( \sqrt{n}(T(\hat{\theta}_n) - \theta_0) \) has the same limit distribution as that of 
\[ [I_1(\theta_0)]^{-1} \left( \frac{1}{\sqrt{n}} l'(\theta_0) \right) \] if the assumed model is true. Refer to Section 5.1.2, this implies that the GMHD estimator is asymptotically equivalent to the GML estimator when the assumed model is true.

**Theorem 5.2.7** Let \( f_{i, \theta} \) and \( g_i \) for all \( i = 1, 2, \ldots m \) be supported on the nonnegative integers. Suppose Assumptions (A4)-(A9) hold. Then, if \( T(\hat{\theta}_n) \) is a consistent estimator of \( \theta_0 \),

\[
\sqrt{n}(T(\hat{\theta}_n) - \theta_0) \xrightarrow{D} N_p(0, \Sigma_{\theta_0}),
\]

where

\[
\Sigma_{\theta_0} = \left[ \sum_{i=1}^{m} \lambda_i \mathcal{H}(\theta_0; g_i) \right]^{-1} \left[ \frac{1}{4} I_1(\theta_0) - \frac{1}{4} \sum_{i=1}^{m} \lambda_i \left[ \mathcal{H}(\theta_0; g_i) \right] \left[ \mathcal{H}(\theta_0; g_i) \right]^T \right] \left[ \sum_{i=1}^{m} \lambda_i \mathcal{H}(\theta_0; g_i) \right]^{-1}
\]

for any \( g \in \Psi \) and

\[
\Sigma_{\theta_0} = [I_1(\theta_0)]^{-1}
\]

if \( g = f_{\theta_0} \in \mathcal{F} \).

**Proof:** According to Lemma 5.2.6,

\[
\sqrt{n}(T(\hat{\theta}_n) - \theta_0) = \left[ \sum_{i=1}^{m} \lambda_i \mathcal{H}(\theta_0; g_i) \right]^{-1} \left\{ \sqrt{n} \left[ -\frac{1}{n} \mathcal{H}(\theta_0; \hat{\theta}_n) \right] \right\} + o_p(1).
\]

Here, by Tool (T4),

\[
\sqrt{n} \left[ -\frac{1}{n} \mathcal{H}(\theta_0; \hat{\theta}_n) \right] \xrightarrow{D} N_p \left( 0, \frac{1}{4} I_1(\theta_0) - \frac{1}{4} \sum_{i=1}^{m} \lambda_i \left[ \mathcal{H}(\theta_0; g_i) \right] \left[ \mathcal{H}(\theta_0; g_i) \right]^T \right)
\]

for any \( g \in \Psi \) and

\[
\sqrt{n} \left[ -\frac{1}{n} \mathcal{H}(\theta_0; \hat{\theta}_n) \right] \xrightarrow{D} N_p \left( 0, \frac{1}{4} I_1(\theta_0) \right)
\]

if \( g = f_{\theta_0} \in \mathcal{F} \).

Also, \( \sum_{i=1}^{m} \lambda_i \mathcal{H}(\theta_0; g_i) \), which is the limit of \( \frac{1}{n} \mathcal{H}(\theta_0; \hat{\theta}_n) \), does not depend on the sample sizes. Therefore, as \( n \to \infty \), \( \left[ \sum_{i=1}^{m} \lambda_i \mathcal{H}(\theta_0; g_i) \right]^{-1} \) does not change for any \( g \in \Psi \) and according to result (5.29), it in particular equals \( \left[ \frac{1}{2} I_1(\theta_0) \right]^{-1} \) if \( g = f_{\theta_0} \in \mathcal{F} \).
Then, the result of the theorem is immediate by (multivariate) Slutsky’s theorem.

**Remark:** Beran (1977) establishes the asymptotic normality of the MHD estimator for continuous data under some very strong assumptions. The conditions that Beran imposes are not only on the model, on the kernel density estimator, but also on the underlying distribution of the data. For example, assumption (iv) of Beran’s Theorem 4 (1977) requires that the density of the data distribution, \( g \), to be positive on some specified compact support. On the other hand, the asymptotic normality of the MHD estimator for count data is shown by Simpson (1987) under much weaker assumptions. The conditions imposed by Simpson’s Theorem 2 (1987) are only on the model, but no longer on the data distribution. Theorem 5.2.7 inherits this characteristic. The most stringent condition, \( s_{\theta,\theta} \in L^1 \) (Assumption (A8)), is satisfied by Poisson and log-series models.

To end this section, the asymptotic properties of the GML estimator and the GMHD estimator in GLM setting are considered. Recall that GML estimation and GMHD estimation apply directly to GLMs since each vector of explanatory variables \( x_i \) determines a subpopulation through a link function \( \ell \) such that \( \mu_i = \ell^{-1}(\eta_i) = \ell^{-1}(x_i'\beta) \) and \( \beta \) are common parameters of all subpopulations. Therefore, it is expected that the results in Sections 5.1 and 5.2 are applicable for GLMs, where \( \theta = \beta \), as long as the conditions of the theorems are satisfied. The following are some conditions that require special attention when applying the results in Section 5.1 and 5.2 to GLMs.

(i) To apply the results of Theorem 5.2.3, one must assure that the parameter space \( B \) of \( \beta \) can be embedded to a compact set because for most GLMs, link functions are chosen in a way that \( \beta_j \)'s are allowed to take on any values on the real line. Consider that \( \mathcal{F}_i = \{ f_{\mu_i(\beta)}(y) : \beta \in B \} \) is the model family of distributions for the \( i^{th} \) subpopulation, where, for most cases, the parametric space \( B = \prod_{k=1}^{p} (-\infty, \infty) \). Then, the embedding can be done using logistic transformation. Let each \( \beta_k' = \frac{\beta_k}{1+e^{-\beta_k}} \) for \( k = 1, 2, \ldots, p \). Then, each model family can be represented as \( \mathcal{F}_i = \{ f_{\mu_i(\beta')}(y) : \beta' \in B' \} \), where \( B' = \prod_{k=1}^{p} (0, 1) \). As \( \beta \to 0 \) or \( \beta \to 1 \), \( \frac{1}{n} \mathcal{H}(\beta';g) = \sum_{i=1}^{m} \frac{n_i}{n} \| f_{\mu_i(\beta')} - g_i \|_2^2 = 2 \).
Therefore, \( \frac{1}{n}H(\beta';g) \) can be extended to a function \( \frac{1}{n}H(\cdot;g) \) on \( B = \Pi_{k=1}^p [0,1] \), which is compact, and the extended function achieves a minimum in \( B \). In fact, the minimum must occur in \( \text{int}(B') \) since \( 0 \leq \frac{1}{n}H(\beta;g) \leq 2 \) for every \( \beta' \in B' \) and \( \frac{1}{n}H(\beta;g) \equiv 2 \) is impossible. Consequently, the conclusions of Theorem 5.2.3 remain valid for the GLMs.

(ii) In application of the asymptotic normality results of the GML estimator (Theorem 5.1.2) and of the GMHD estimator (Theorem 5.2.7), an extra factor that needs to be considered is the nature of the explanatory variables, whose values in practice may be chosen by design or may arise at random. One needs to ensure that the asymptotic variances \( [I_1(\beta_0)]^{-1} \) of the GML estimator and

\[
\left[ \sum_{i=1}^m \lambda_i \tilde{H}(\beta_0;g_i) \right]^{-1} \left[ \frac{1}{4} I_1(\beta_0) - \frac{1}{4} \sum_{i=1}^m \lambda_i \left[ \tilde{H}(\beta_0;g_i) \left[ \tilde{H}(\beta_0;g_i) \right]^2 \right] \right]^{-1} \sum_{i=1}^m \lambda_i \tilde{H}(\beta_0;g_i)
\]

of the GMHD estimator remain finite and positive definite under the consideration of the explanatory variables.

Now, consider a GLM whose family of distributions is in the canonical form (Equation (1.13)). Given that the canonical link is used, the log-likelihood function for the \( i^{th} \) subpopulation is

\[
l(y_i; \eta_i, \varphi) = \left\{ \left[ y_i - b(\eta_i) \right] / a_i(\varphi) + c(y; \varphi) \right\}.
\]

Then, it is easy to check that

\[
I_1(\beta) = \sum_{i=1}^m \lambda_i \cdot \frac{b''(\eta_i)}{a_i(\varphi)} \cdot x_i x_i^\top = X^\top V X,
\]

where \( V = \text{diag} \left( \lambda_1 \frac{b''(\eta_1)}{a_1(\varphi)}, \lambda_2 \frac{b''(\eta_2)}{a_2(\varphi)}, \ldots, \lambda_m \frac{b''(\eta_m)}{a_m(\varphi)} \right) \), which is positive definite. Since \( m \) is fixed and \( \frac{\Delta}{n} \to \lambda_i \) by definition and \( X \), for the interest of this dissertation, is chosen by design, hence, \( I_1(\beta) \) is finite and positive definite. That is, the asymptotic normality of the GML estimator remains valid.
For GMHD estimation,
\[
\sum_{i=1}^{m} \lambda_i \hat{H}(\beta; g_i) = \sum_{i=1}^{m} \lambda_i \left\{ \int \left[-\frac{1}{2} \left( y - b'(\eta_i) \right)^2 + \frac{b''(\eta_i)}{a_i(\varphi)} \right] s_{\mu_i(\beta)}(y) \cdot g_i^{1/2}(y) \, dy \right\} x_i x_i^T,
\]
which, under the assumed model, equals
\[
\frac{1}{2} \sum_{i=1}^{m} \lambda_i \cdot \frac{b''(\eta_i)}{a_i(\varphi)} \cdot x_i x_i^T = \frac{1}{2} I_1(\beta),
\]
and
\[
\frac{1}{4} I_1(\beta_0) - \frac{1}{4} \sum_{i=1}^{m} \lambda_i \left[ \hat{H}(\beta_0; g_i) \right] \left[ \hat{H}(\beta_0; g_i) \right]^T = \frac{1}{4} \sum_{i=1}^{m} \lambda_i \left\{ \frac{b''(\eta_i)}{a_i(\varphi)} - \left[ \int \frac{y - b'(\eta_i)}{a_i(\varphi)} s_{\mu_i(\beta)}(y) \cdot g_i^{1/2}(y) \, dy \right]^2 \right\} x_i x_i^T,
\]
which, under the model, is equal to
\[
\frac{1}{4} \sum_{i=1}^{m} \lambda_i \cdot \frac{b''(\eta_i)}{a_i(\varphi)} \cdot x_i x_i^T = \frac{1}{4} I_1(\beta).
\]

With fixed \( m \) and the conditions that \( \frac{m}{n} \to \lambda_i \) and \( X \) is a design matrix, both \( \sum_{i=1}^{m} \lambda_i \hat{H}(\beta; g_i) \) and \( \frac{1}{4} I_1(\beta_0) - \frac{1}{4} \sum_{i=1}^{m} \lambda_i \left[ \hat{H}(\beta_0; g_i) \right] \left[ \hat{H}(\beta_0; g_i) \right]^T \) exist. The latter is positive definite because it is the asymptotic variance of \( \sqrt{n} \left[-\frac{1}{n} \hat{H}(\theta_0; \hat{g}_n)\right] \) according to Tool (T4). Hence, the asymptotic variance-covariance matrix of the GMHD estimator of \( \beta \) is finite and positive definite, which implies the validity of Theorem 5.2.7.

5.3 Robustness Properties of GMMPHD Estimators in GLMs

Hampel’s (1974) influence function is often used to analyze the local robustness of an estimator. It measures how much an estimator is influenced under small contamination at any point. A robust estimator is usually required to have a bounded influence function.

For iid models, suppose that \( g \) is the true density of the data, \( F = \{ f_\theta : \theta \in \Theta \} \) is the assumed model family for \( g \), and \( K \) is the support of the model distribution. If \( T \) is the functional of an estimator on the space of densities \( G \), then Hampel’s influence function is defined to be
\[
IF(z; T; g) = \left. \frac{\partial T(g_{\alpha, z})}{\partial \alpha} \right|_{\alpha = 0} = \left. \frac{\partial T((1 - \alpha)g + \alpha \Delta_z)}{\partial \alpha} \right|_{\alpha = 0},
\]
where $\Delta_z$ is the indicator function for $z$ that puts probability one on the point mass $z \in K$. Here, $IF(z; T; g)$ may be a $p \times 1$ vector if the parameter space $\Theta$ is of dimensional $p$ and its $j^{th}$ component represents the influence of $z$ on the $j^{th}$ parameter component $\theta_j$.

Beran (1977) and Lindsay (1994) demonstrate that if the assumed model is true, the MHD estimator and the ML estimator have the same influence function:

$$IF(z; T; f_\theta) = T(f_\theta, z) - \theta = T((1 - \alpha)f_\theta + \alpha \Delta_z) - \theta = \left[ I_1(\theta) \right]^{-1} l'(\theta; z),$$

where $I_1(\theta)$ is the Fisher information matrix of a single observation and $l'(\theta; y) = \frac{\partial}{\partial \theta} \log(f_\theta(y))$ is the score function. This influence function is usually not bounded because the score function usually is not.

A discussion provided by Lindsay (1994) gives a detailed analysis of why the traditional influence function is a very misleading robust measure for MHD type of density-based estimators. Beran (1977) suggests the use of the $\alpha$-influence function as an alternative measure of robustness for MHD estimator. Cutler and Cordero-Braña (1996) use a slightly modified $\alpha$-influence function of Beran's for the application in finite mixture models and illustrate the $\alpha$-influence function for a specific example (a mixture model of two univariate normal distributions) by evaluating the $\alpha$-influence function numerically. For the models with count data, the $\alpha$-influence curve for an iid model may be defined by

$$\alpha - IC(z) = \frac{T(f_\theta, z) - \theta}{\alpha} = \frac{T((1 - \alpha)f_\theta + \alpha \Delta_z) - \theta}{\alpha},$$

where $\alpha \in [0, 1]$, $\Delta$ is the indicator function (rather than the uniform distribution function used by Beran), and $z$ is a positive integer in $K$. Note that $IF(z; T; f_\theta) = \lim_{\alpha \to 0} \alpha - IC(z)$. Beran (1977) points out that the $\alpha$-influence function is bounded for any $\alpha$ and that $\lim_{z \to \infty} \alpha - IC(z) = 0$.

To explore the robustness of GML and GMHD estimators, we extend the idea of influence function and $\alpha$-influence function for the iid case to the more general case where data come from two or more subpopulations. Suppose that $g_1, g_1, \ldots, g_m$ are true densities of the $m$ subpopulations from which data come and that each $\mathcal{F}_i = \{ f_{i, \theta} : \theta \in \Theta \}$ is the assumed
model family for $g_i$. Assume that $\alpha$ is the same for all subpopulations and let each

$$g_{\alpha,z_i} = (1 - \alpha) g_i + \alpha \Delta z_i.$$ 

Then, if $T$ is the functional of an estimator on the densities $g_1, g_2, \ldots, g_m$, the generalized influence function may be defined by

$$IF(z; T; g) := IF(z_1, z_2, \ldots, z_m; T; g_1, g_2, \ldots, g_m) = \left. \frac{\partial T(g_{\alpha,z_1}, g_{\alpha,z_2}, \ldots, g_{\alpha,z_m})}{\partial \alpha} \right|_{\alpha=0} := \left. \frac{\partial T(g_{\alpha,z})}{\partial \alpha} \right|_{\alpha=0},$$

which is equal to

$$IF(z; T; f_{\theta}) := IF(z_1, z_2, \ldots, z_m; T; f_{1,\theta}, f_{2,\theta}, \ldots, f_{m,\theta}) = \left. \frac{\partial T(f_{\theta,\alpha,z_1}, f_{\theta,\alpha,z_2}, \ldots, f_{\theta,\alpha,z_m})}{\partial \alpha} \right|_{\alpha=0} := \left. \frac{\partial T(f_{\theta,\alpha,z})}{\partial \alpha} \right|_{\alpha=0}$$

if the assumed models are true, where

$$f_{\theta,\alpha,z_i} = (1 - \alpha) f_{i,\theta} + \alpha \Delta z_i.$$ 

Section 5.3.1 establishes that at the true model the GMHD estimator has the same unbounded influence function as that of the GML estimator.

The generalized $\alpha$-influence function, correspondently, may be defined by

$$\alpha - IC(z) = \frac{T(f_{\theta,\alpha,z}) - \theta}{\alpha}$$

so that the condition $IF(z; T; f_{\theta}) = \lim_{\alpha \to 0} \alpha - IC(z)$ is satisfied.

For models with data coming from more than one population, the $\alpha$-influence function defined above is difficult to illustrate because it depends on a set of values of $z_i$'s. Therefore, in studying the robustness of a certain estimator, it would be convenient to consider the special cases where all $z_i$'s depend on a common factor $d$ that controls the degree of the overall contamination. For example, consider the case where

$$z_i = \mu_i + d \sigma_i \text{ for each } i$$
and illustrate the influence of the contamination through

$$\alpha - IC(d) = \frac{T(f_{\theta, \alpha, d}) - \theta}{\alpha}$$

instead. For any \( \alpha \), \( \alpha - IC(d) \) is bounded because \( \lim_{d \to \infty} \alpha - IC(d) = 0 \). Note that each \( z_i \) needs to be rounded to the nearest integer in the support of the model distribution if models are discrete.

The rest of this section is organized as follows. Section 5.3.1 explains how the influence functions of GML and GMHD estimators are derived. Section 5.3.2 illustrates the \( \alpha \)-influence functions of GML and GMHD estimators for Poisson and binomial GLMs studied in this dissertation followed by the corresponding empirical \( \alpha \)-influence results of GML and GMPHD estimators (using optimal \( h \), \( h=0.5 \), and \( h=1.0 \)) in Section 5.3.3.

5.3.1 Influence Function of the GMPHD Estimator

As with the iid models, it is expected that the GML and GMHD estimators have the same influence function if the assumed model is true. The derivations of the influence functions for both estimators are straightforward but tedious. Consider the definition of the GML estimator. The GML functional \( T_{\text{GML}} \) can be defined by

$$T_{\text{GML}}(g) = T_{\text{GML}}(g_1, g_2, \ldots, g_m) = \arg \max_{\theta \in \Theta} \sum_{i=1}^{m} \lambda_i E_{g_i} \left[ \log f_i, \theta (Y) \right],$$

or, equivalently, by the estimating equation

$$0 = \sum_{i=1}^{m} \lambda_i E_{g_i} \left[ t^{(i)}'(T_{\text{GML}}(g); Y) \right],$$

where \( t^{(i)}(\theta; y) \) (Section 5.2, Notation 32) is the score function of the \( i^{th} \) subpopulation.

The GMHD functional is already given in Section 5.2 by Equation (5.16). Recall that the GMHD functional \( T_{\text{GMHD}} \) is defined by

$$T_{\text{GMHD}}(g) := T_{\text{GMHD}}(g_1, g_2, \ldots, g_m) = \arg \min_{\theta \in \Theta} \sum_{i=1}^{m} \lambda_i \left\| f_{i, \theta}^{1/2} - g_i^{1/2} \right\|_2^2$$

$$= \arg \max_{\theta \in \Theta} \sum_{i=1}^{m} \lambda_i \int f_{i, \theta}^{1/2}(y) \cdot g_i^{1/2}(y) \, dy, \,$$
or, equivalently, by the $T(g)$ satisfying the estimating equation

$$0 = \sum_{i=1}^{m} \lambda_i \int \hat{s}_{i,T(g)}(y) \cdot g_{i}^{1/2}(y) \, dy. \tag{5.31}$$

Here, $\hat{s}_{i,\theta}$ is given by Notation 17 in Section 5.2. Then, the influence functions of GML and GMHD estimators can be derived after some tedious differentiation of the estimating equations (5.30) and (5.31), individually. It is easy to verify that under the true model

$$IF(z; T_{GML}; f_\theta) = IF(z; T_{GMHD}; f_\theta) = \left[ \sum_{i=1}^{m} \lambda_i I_i(x)^{(i)}(\theta) \right]^{-1} \left[ \sum_{i=1}^{m} \lambda_i I_i'(x; z_i) \right]$$

which is again unbounded for most cases.

5.3.2 $\alpha$-Influence Curve of the GMPHD Estimator in GLMs

Analytical results of $\alpha$-influence functions are difficult to obtain for most estimators, but numerical results can easily be computed for some models. In this section, $\alpha$-influence functions of QL (i.e., GML) and GMHD estimators are calculated numerically and illustrated for all the Poisson and binomial GLMs used in this dissertation against Type I contamination.

Consider the GLMs with count data. The QL and GMHD functionals have the following forms.

$$T_{\text{QL}}(g) := T_{\text{QL}}(g_1, g_2, \ldots, g_m) = \arg \max_{b \in B} \sum_{i=1}^{m} \lambda_i E_{g_i} \left[ \log f_{\mu_i}(b)(Y_i) \right]$$

and

$$T_{\text{GMHD}}(g) := T_{\text{GMHD}}(g_1, g_2, \ldots, g_m) = \arg \max_{b \in B} \sum_{i=1}^{m} \lambda_i \sum_{t} \left[ \log f_{\mu_i}(b)(t) \right] g_i(t)$$

First, construct the $\alpha$-influence curves for Type I contamination models defined by Equation (4.2). Recall that in this dissertation it is assumed that $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0.1$ and
\( d_1 = d_2 = \ldots = d_m = 5 \) for all Type I contamination models. Therefore, the \( \alpha \)-influence curves for Type I contamination models can be constructed by letting each

\[
    f_{\mu_i(\beta), \alpha, d}(t) = (1 - \alpha) f_{\mu_i(\beta)}(t) + \alpha \Delta z_i(t),
\]

where \( z_i = \mu_i(\beta) + d \sigma_i(\beta) \) rounded to the nearest integer in the support of the model distribution and \( \sigma_i \) is the standard deviation of the \( i^{th} \) subpopulation. For example, \( \sigma_i = \sqrt{\mu_i} \) for Poisson GLMs and \( \sigma_i = \sqrt{N_i \pi_i(1 - \pi_i)} = \sqrt{N_i \left( \frac{k_i}{N} \right) \left( 1 - \frac{k_i}{N} \right)} \) for binomial GLMs.

Then,

\[
    T_{QL}(f_{\mu(\beta), \alpha, d}) := T_{QL}(f_{\mu(\beta), \alpha, d}, f_{\mu_2(\beta), \alpha, d}, \ldots, f_{\mu_m(\beta), \alpha, d})
\]

\[
    = \arg \max_{b \in B} \sum_{i=1}^{m} \lambda_i \sum_{t} \left[ \log f_{\mu_i(b)}(t) \right] f_{\mu_i(\beta), \alpha, d}(t)
\]

\[
    = \arg \max_{b \in B} \sum_{i=1}^{m} \lambda_i \sum_{t} \left[ \log f_{\mu_i(b)}(t) \right] \left[ (1 - \alpha) f_{\mu_i(\beta)}(t) + \alpha \Delta z_i(t) \right]
\]

\[
    = \arg \max_{b \in B} \sum_{i=1}^{m} \lambda_i \left\{ (1 - \alpha) \sum_{t} \left[ \log f_{\mu_i(b)}(t) \right] f_{\mu_i(\beta)}(t) + \alpha \left[ \log f_{\mu_i(b)}(z_i) \right] \right\}
\]

and

\[
    T_{GMHD}(f_{\mu(\beta), \alpha, d}) := T_{GMHD}(f_{\mu(\beta), \alpha, d}, f_{\mu_2(\beta), \alpha, d}, \ldots, f_{\mu_m(\beta), \alpha, d})
\]

\[
    = \arg \max_{b \in B} \sum_{i=1}^{m} \lambda_i \sum_{t} f_{\mu_i(b)}^{1/2}(t) \cdot f_{\mu_i(\beta), \alpha, d}^{1/2}(t)
\]

\[
    = \arg \max_{b \in B} \sum_{i=1}^{m} \lambda_i \sum_{t} f_{\mu_i(b)}^{1/2}(t) \left[ (1 - \alpha) f_{\mu_i(\beta)}(t) + \alpha \Delta z_i(t) \right]^{1/2}
\]

\[
    = \arg \max_{b \in B} \sum_{i=1}^{m} \lambda_i \left\{ \sum_{t \neq z_i} f_{\mu_i(b)}^{1/2}(t) \left[ (1 - \alpha) f_{\mu_i(\beta)}(t) \right]^{1/2} + f_{\mu_i(b)}^{1/2}(z_i) \left[ (1 - \alpha) f_{\mu_i(\beta)}(z_i) + \alpha \right]^{1/2} \right\}
\]

\[
    = \arg \max_{b \in B} \sum_{i=1}^{m} \lambda_i \left\{ (1 - \alpha)^{1/2} \sum_{t \neq z_i} f_{\mu_i(b)}^{1/2}(t) \cdot f_{\mu_i(\beta)}^{1/2}(t) + f_{\mu_i(b)}^{1/2}(z_i) \left[ (1 - \alpha) f_{\mu_i(\beta)}(z_i) + \alpha \right]^{1/2} \right\}
\]

For any fixed values of \( \beta \) and \( \alpha \), \( T_{QL}(f_{\mu(\beta), \alpha, d}) \) and \( T_{GMHD}(f_{\mu(\beta), \alpha, d}) \) can be computed numerically over a fine grid of values of \( d \) and their \( \alpha \)-influence curves \( \alpha - IC(d) = \frac{T(f_{\mu(\beta), \alpha, d} - \beta)}{\alpha} \) can be plotted against \( d \).

Figure 5.1 illustrates the \( \alpha - IC(d) \) of the QL estimator (solid line) and of the GMHD estimator (long dashed line) against Type I contamination for three Poisson GLMs, where
Figure 5.1. α-IC of QL and GMHD estimators for Poisson GLMs under Type I contamination.
each $\alpha-IC(d)$ is calculated over $d = 0.1, 0.2, \ldots, 9.9, 10.0$. In these models every $\alpha-IC(d)$ is a $4 \times 1$ vector so each element of $\alpha-IC(d)$ is plotted versus $d$ for each one of the parameter components, $\beta_0, \beta_1, \beta_2,$ and $\beta_3$. It shows that for every parameter component, the influence of Type I contamination on GMHD estimator reaches its maximum at values of $d$ between 2 and 3 and then decreases to 0 as $d \to \infty$. This nice robustness property of the GMHD estimator does not hold for the QL estimator whose $\alpha-IC(d)$ appears to grow without bound. Note that the lack of smoothness of the $\alpha-IC(d)$'s is because the $\alpha-IC(d)$ is not a continuous function of $d$ in the sense that the $z_i$'s are rounded integers.

The $\alpha-IC$ results for five Type I contaminated binomial GLMs are given in Figures 5.2 and 5.3. P and N models have similar results to those of Poisson models, but the $\alpha-IC$'s of the three B models do not decrease to 0 as $d \to \infty$. The latter is because the $z_i$'s, which are restricted to be less than or equal to the $N_i$'s for the binomial, do not really tend to 0 as $d$ tends to $\infty$. For B models where normally the $N_i$'s are small and the $\pi_i$'s are moderate, the values of $z_i$ will be replaced by $N_i$ after a certain value of $d$. For example, in the B1 model the $z_i$'s are equal to the common $N_i (= 5)$ as $d$ exceeds about 2.5. In the B2 model the $z_i$'s are equal to the common $N_i (= 15)$ as $d$ exceeds about 3.8. In the B3 model every $z_i$ is equal to the $N_i$ (unequal $N_i$'s) as $d$ exceeds about 3.9. This is also the reason why the $\alpha-IC$ of the QL estimator is bounded for B models because the $z_i$'s have no chance to get any worse after a certain value of $d$.

The Type II contamination model define by Equation (4.3) is formulated in joint probability distribution form, which makes both $T_{QL}(f_{\mu(\beta),k,d})$ and $T_{GMHD}(f_{\mu(\beta),k,d})$ very difficult to be computed even numerically. Therefore, the robustness of QL and GMHD estimators against Type II contamination is analyzed in Section 5.3.3 by means of the empirical $\alpha$-influence curves.

Note that the $\alpha-IC$ is an asymptotic version of the empirical $\alpha-IC$. The latter is considered as an estimate of the former. Unlike the empirical $\alpha-IC$, the $\alpha-IC$ is computed without data so empty cells are meaningless for the $\alpha-IC$. Hence, asymptotically, the
Figure 5.2. α-IC of QL and GMHD estimators for binomial GLMs under Type I contamination (B1, P, and N models)
Alpha-IC for Type I Contamination (Alpha = 0.1)

Binomial means approximately equal to 10
B2 Model (Ni's are all Equal)

Binomial means approximately equal to 10
B3 Model (Ni's are Different)

Figure 5.3. α-IC of QL and GMHD estimators for binomial GLMs under Type I contamination (B2 and B3 models)
GMPHD estimator with any value of $h$ should have the the same $\alpha$-influence curve as that of the GMHD estimator. This is the reason why only GMHD estimators are considered in this subsection. The empirical $\alpha$-influence curve on the other hand is computed based on data, so it can be constructed for any member in the class of GMPHD estimators.

5.3.3 Empirical $\alpha$-Influence Curve of the GMPHD Estimator in GLMs

In this subsection, empirical $\alpha$-influence curves of QL estimators, of GMPHD estimators (using optimal $h$ and $h = 0.5$), and of GMHD estimators ($h = 1.0$) are computed by simulation for all Poisson and binomial GLMs in this dissertation against both Type I and Type II contaminations.

To construct the empirical $\alpha - IC(d)$ for any model, data are generated repeatedly from the model under contamination and under no contamination 1000 times, individually. Assume that $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0.1$ for all Type I contaminated models and $k = 2$ for all Type II contaminated models yielding the overall contamination factor $\alpha$ equal to 0.1 for Type I contamination and equal to $\frac{2}{21}$ for Type II contamination. For each sample and for each $d = 0.0, 0.5, \ldots, 9.5, 10.0$, compute both the estimator under contamination ($\alpha = 0.1$ or $k = 2$) and that under no contamination ($\alpha = 0$ or $k = 0$). Then, the empirical $\alpha - IC(d)$ is defined by

\[ \text{Empirical } \alpha - IC(d) = \frac{E[\hat{\beta}_{\alpha,d}] - E[\hat{\beta}]}{\alpha}, \]

where $\hat{\beta}_{\alpha,d}$ and $\hat{\beta}$ denote the estimate derived under contamination and that under no contamination, respectively. Here, $E[\hat{\beta}_{\alpha,d}]$ and $E[\hat{\beta}]$ are the sample means of 1000 values of $\hat{\beta}_{\alpha,d}$ and of $\hat{\beta}$ from replications. Note that for GMPHD estimation, $E[\hat{\beta}_{\alpha,d}]$ and $E[\hat{\beta}]$ must be computed using the same $h$, so for the GMPHD estimator using optimal $h$, the optimal $h$ derived according to the contaminated data is used.

The empirical $\alpha - IC$ defined by Equation (5.32) can be decomposed as the following.

\[ \frac{E[\hat{\beta}_{\alpha,d}] - E[\hat{\beta}]}{\alpha} = \frac{E[\hat{\beta}_{\alpha,d}] - \beta}{\alpha} - \frac{E[\hat{\beta}] - \beta}{\alpha}. \]
in which the first term on the right-hand side can be seen as a measure containing both
the influence of the contamination (the term on the left-hand side) and the bias caused by
estimation (the second term on the right-hand side). Since empirical $\alpha - IC$ is a measure
of robustness, it should only contain the influence of the contamination. The definition
of empirical $\alpha - IC$ given by Equation (5.32) eliminates the estimation bias and hence a
correct estimate of the theoretical $\alpha - IC$.

Figures 5.4 and 5.5 compare the measures of $\frac{\varepsilon[\hat{\beta}_{\alpha,d}]-\beta}{\alpha}$ and $\frac{\varepsilon[\hat{\beta}_{\alpha,d}]-\varepsilon[\hat{\beta}]}{\alpha}$ for the Poisson
GLM with $\mu_i$'s approximately equal to 2 under Type I contamination for common
sample sizes of 1, 20, and 100. Each figure has results for the QL estimator (solid line),
the GMPHD estimator using optimal $h$ (dotted line), the GMPHD estimator using $h = 0.5$
(short dashed line), and the GMHD estimator (long dashed line). The biasedness of the
GMPHD estimators for small samples is shown in Figure 5.4. One can see that the GMHD
estimators are very biased at small samples and that their biases can be reduced by down-
weighting the empty cells. When the common sample size is large ($n_c = 100$), the estimation
bias vanishes for all GMPHD estimators in the class. Comparing Figure 5.5 to the first row
of the Figure 5.1, it suggests that the empirical $\alpha - IC$ defined in Equation (5.32) is an
unbiased estimate of the theoretical $\alpha - IC$ and that it is a very accurate estimate of the
theoretical $\alpha - IC$ when $n_c$ is as large as 100.

The figures in the rest of this subsection contain the results of empirical $\alpha - IC$ for all
models considered in this dissertation. They are Figures 5.6 and 5.7 for Poisson GLMs and
Figures 5.8, 5.9, 5.10, and 5.11 for binomial GLMs. A common sample size of 100 is used
in all cases. One can see that the results of Type I contamination in Figure 5.6, 5.8, and
5.9 are almost identical to the results of $\alpha - IC$ in Figures 5.1, 5.2, and 5.3 except that the
$\alpha - IC$ is evaluated over a finer grid of values of $d$.

Results for Type II contamination are given in Figures 5.7, 5.10, and Figures 5.11. The
behavior of the empirical $\alpha - IC$ for Type II contamination is very similar to that for Type
I contamination except that the magnitudes of the influence are slightly different.
Figure 5.4. Empirical $\alpha$-IC before eliminating the bias caused by estimation of QL estimator, the optimal GMPHD estimator, the GMPHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLM with means approximately equal to 2 under Type I contamination for $n_c$ of sizes 1, 20, and 100.
Empirical Alpha-IC for Type I Contamination (Alpha = 0.1)

Poisson means approximately equal to 2

Figure 5.5. Empirical $\alpha$-IC of QL estimator, the optimal GMPHD estimator, the GMPHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLM with means approximately equal to 2 under Type I contamination for $n_c$ of sizes 1, 20, and 100
Empirical Alpha-IC for Type I Contamination ($\alpha = 0.1$)
$n_1=n_2=\ldots=n_m=100$

Poisson means approximately equal to 2

Poisson means approximately equal to 10

Poisson means approximately equal to 50

Figure 5.6. Empirical $\alpha$-IC of QL estimator, the optimal GMPHD estimator, the GMPHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLMs under Type I contamination for $n_c$ of size 100
Empirical Alpha-IC for Type II Contamination (Alpha = 2/21)
n1=n2=...=nm=100

Poisson means approximately equal to 2

Poisson means approximately equal to 10

Poisson means approximately equal to 50

Figure 5.7. Empirical $\alpha$-IC of QL estimator, the optimal GMPHD estimator, the GMPHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLMs under Type II contamination for $n_c$ of size 100
Empirical Alpha-IC for Type I Contamination (Alpha = 0.1)
n1=n2=...=nm=100

Binomial means approximately equal to 2
B1 Model

Binomial means approximately equal to 10
P Model

Binomial means approximately equal to 50
N Model

Figure 5.8. Empirical $\alpha$-IC of QL estimator, the optimal GMPHD estimator, the GMPHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLMs under Type I contamination for $n_c$ of size 100 (B1, P, and N models)
Empirical Alpha-IC for Type I Contamination (Alpha = 0.1)
n1=n2=...=nm=100

Binomial means approximately equal to 10
B2 Model (Ni's are all Equal)

Binomial means approximately equal to 10
B3 Model (N_i's are Different)

Figure 5.9. Empirical $\alpha$-IC of QL estimator, the optimal GMPHD estimator, the GMPHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLMs under Type I contamination for $n_c$ of size 100 (B2 and B3 models)
Empirical Alpha-IC for Type II Contamination (Alpha = 2/21)
n1=n2=...=nm=100

Binomial means approximately equal to 2
B1 Model

Binomial means approximately equal to 10
P Model

Binomial means approximately equal to 50
N Model

Figure 5.10. Empirical $\alpha$-IC of QL estimator, the optimal GMPHD estimator, the GMPHD estimator using $h = 0.5$, and the ordinary GMHD estimator for the Poisson GLMs under Type II contamination for $n_c$ of size 100 (B1, P, and N models)
Empirical Alpha-IC for Type II Contamination (Alpha = 2/21)

\( n_1 = n_2 = \ldots = n_m = 100 \)

Binomial means approximately equal to 10

B2 Model (Ni's are all Equal)

Binomial means approximately equal to 10

B3 Model (Ni's are Different)

Figure 5.11. Empirical \( \alpha \)-IC of QL estimator, the optimal GMPHD estimator, the GMPHD estimator using \( h = 0.5 \), and the ordinary GMHD estimator for the Poisson GLMs under Type II contamination for \( n_c \) of size 100 (B2 and B3 models)
For the three binomial B GLMs, as with the theoretical $\alpha - IC$, the behavior of their empirical $\alpha - IC$s (in Figures 5.8, 5.9, 5.10, and 5.11) explains the phenomenon where GMPHD estimators at $d = d_1 = d_2 = \ldots = d_m = 5$ are not very robust against both Type I and Type II contamination. (See Figures 4.31, 4.32, 4.33, and 4.34 for binomial GMPHD estimation results under contamination.)
CHAPTER 6

CONCLUSIONS, PROBLEMS AND FUTURE RESEARCH

6.1 Conclusions

In this dissertation, GMPHD estimation and GPHDV tests are developed as robust and efficient alternatives to the traditional QL estimation and GLR tests (deviance tests) for discrete GLMs. The idea is to generalize the application of MHD estimation and HDV tests, which are robust alternatives to the ML estimation and LR tests, for a single random sample to data coming from two or more populations (including GLMs). Because cell samples are usually small in most applications of GLMs, the penalized procedures used in MPHHD estimation (Harris and Basu, 1994) and in PHDV tests (Basu, Harris, and Basu, 1995) are also considered in the generalization to improve the small sample performance of the GMPHD estimation and GPHDV tests. An extensive simulation study for Poisson and for binomial GLMs is carried out to evaluate these new procedures.

The implementation of GMPHD estimation involves the choice of penalty weights. Simulation results indicate that the suggested penalty weight 0.5 for MPHHD estimation is not optimal (in terms of mean squared error) for small sample estimation. A parametric bootstrap (PBoot) procedure is proposed to "estimate" the optimal penalty weight for the GMPHD estimation. Results suggest that the GMPHD estimates obtained using PBoot-estimated penalty weights are very competitive to the QL estimates when data are not contaminated and are much more robust than the QL estimates under contamination.

Based on simulation results, even with the penalized procedure, the test statistics of GPHDV tests do not converge fast enough (as subsample sizes increase) to their limiting $\chi^2$ distributions to make $\chi^2$ critical values usable. A nonparametric bootstrap (NPBoot) algorithm is proposed to compute critical values for the GPHDV tests for goodness-of-fit and for nested models. Results suggest that this NPBoot algorithm generates very accurate
critical values for GPHDV tests even when data are contaminated.

As expected, the GMPHD estimator and GPHDV tests inherit the nice asymptotic and robustness properties of the MHD estimator and HDV tests. The asymptotic results of the GMPHD estimator, such as consistency and asymptotic efficiency, are established analytically in this dissertation. The GMPHD estimates are shown to be asymptotically as efficient as QL estimates if the assumed model is true. The local robustness of the GMPHD estimation procedure is analyzed by the $\alpha$-influence function, which is computed numerically for a specific model, and by the empirical $\alpha$-influence curve, which is obtained by simulation. The superior robustness of GMPHD estimation compared to QL estimation is shown clearly for infinite and finite samples. The asymptotic and robustness properties of the GPHDV tests are demonstrated by simulation. The GPHDV tests are shown to be asymptotically equivalent to GLR tests when the null hypothesis is true and much more robust than GLR tests when using the NPBoot critical values.

6.2 Current Concerns

There are still some concerns of GMPHD estimation and GPHDV tests that need to be improved in the future. They are described as follows.

(1) The parametric bootstrap algorithm for choosing optimal penalty weight is not robust because, under contamination, the parametric bootstrap procedure is not able to generate bootstrap samples that contain the contamination information of the original sample. In another words, the PBoot procedure would choose a penalty weight as if the data were not contaminated. This sometimes causes the GMPHD estimator to slightly lose efficiency when data are contaminated.

(2) For GLMs with one observation per subpopulation, the levels of NPB-GPHDV tests are often very inaccurate, even when data are not contaminated. This situation did not happen for PB-GPHDV tests, so it is the NPBoot algorithm that fails to obtain accurate critical values for this particular case.
(3) NPB-GPHDV tests are not very robust against Type II contamination for all sizes of samples.

(4) Both GMPHD estimation and NPB-GPHDV test procedures are not robust for binomial B models under Type I and Type II contamination.

6.3 Future Research

The followings are some of the things that we plan to do in the future.

(1) Find analytical methods for identifying the optimal penalty weight for GMPHD estimation.

(2) Improve the IRWLS algorithm for computing GMPHD estimators so it may be used as an alternative to the generic Powell's method.

(3) Derive the sampling distribution for the GPHDV test statistics if possible or at least improve the nonparametric bootstrap method of computing critical values for the one-observation-per-cell case and for all cases under Type II contamination.

(4) Extend the application of GMHD estimation and the GHDV tests to continuous GLMs, such as beta GLMs.
REFERENCES


APPENDICES
APPENDIX A

SIMULATION RESULTS FOR HYPOTHESIS TESTING OF 5% AND 1%

NOMINAL LEVELS OF SECTION 4.2.3.1
Tests for Goodness-of-fit (Nominal Level: 0.05)
No Contamination
Poisson Means Approximately Equal to 2

Figure A.1. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)

No Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure A.2. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.05)
Type I Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure A.3. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type I contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)
Type I Contamination
Poisson Means Approximately Equal to 2

![Graph showing observed levels of GLR tests, GPHDVT tests, ordinary GHDV tests, PB-GPHDVT tests, and NPB-GPHDVT tests vs. sample sizes for comparing two nested Poisson GLMs.]

Figure A.4. Plots of observed levels of the GLR tests, the GPHDVT tests using optimal $h$, the GPHDVT tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDVT tests using optimal $h$, and the NPB-GPHDVT tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type I contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.05)
Type II Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure A.5. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type II contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)

Type II Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure A.6. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type II contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.01)

No Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure A.7. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.01)
Figure A.8. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.01)
Tests for Goodness-of-fit (Nominal Level: 0.01)

Type I Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Sample Sizes (n1=n2=...=nm)

Figure A.9. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type I contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)
Type I Contamination
Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure A.10. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type I contamination; Nominal level 0.01)
Tests for Goodness-of-fit (Nominal Level: 0.01)
Type II Contamination
Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure A.11. Plots of observed levels of the GLR tests, the GPHDV tests using optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, the PB-GPHDV tests using optimal \( h \), and the NPB-GPHDV tests using optimal \( h \) vs. common sample sizes \( n_c \) for testing goodness-of-fit of Poisson GLMs. (Type II contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)

Type II Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure A.12. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, the PB-GPHDV tests using optimal $h$, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type II contamination; Nominal level 0.01)
APPENDIX B

SIMULATION RESULTS FOR HYPOTHESIS TESTING OF 5% AND 1%

NOMINAL LEVELS OF SECTION 4.2.3.2
Figure B.1. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)
No Contamination
Poisson Means Approximately Equal to 2

Figure B.2. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.05)

Type I Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure B.3. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (Type I contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)
Type I Contamination
Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure B.4. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type I contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)

Type II Contamination
Poisson Means Approximately Equal to 2

Figure B.5. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type II contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.01)
No Contamination
Poisson Means Approximately Equal to 2

Figure B.6. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of Poisson GLMs. (No contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)
No Contamination

Poisson Means Approximately Equal to 2

Poisson Means Approximately Equal to 10

Poisson Means Approximately Equal to 50

Figure B.7. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for comparing two nested Poisson GLMs. (No contamination; Nominal level 0.01)
Figure B.8. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for testing goodness-of-fit of Poisson GLMs. (Type I contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)
Type I Contamination
Poisson Means Approximately Equal to 2

Figure B.9. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type I contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)

Type II Contamination

Poisson Means Approximately Equal to 2

Sample Sizes ($n_1 = n_2 = \ldots = n_m$)

Poisson Means Approximately Equal to 10

Sample Sizes ($n_1 = n_2 = \ldots = n_m$)

Poisson Means Approximately Equal to 50

Sample Sizes ($n_1 = n_2 = \ldots = n_m$)

Figure B.10. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested Poisson GLMs. (Type II contamination; Nominal level 0.01)
APPENDIX C

SIMULATION RESULTS FOR HYPOTHESIS TESTING OF 5% AND 1%

NOMINAL LEVELS OF SECTION 4.3.3.1
Figure C.1. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.05)
No Contamination
Binomial Means Approximately Equal to 10
B2 Model (N's are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (N's are Different)

Figure C.2. Plots of observed levels of the GLR tests, the GPHDV tests using optimal h, the GPHDV tests using h = 0.5, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal h vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.05)
Figure C.3. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)
No Contamination

Binomial Means Approximately Equal to 10
B2 Model (Ni's are all Equal)

Sample Sizes (n1=n2= ... = nm)

GPHDV tests using optimal h, the GPHDV tests using h = 0.5, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal h vs. common sample sizes n_c for comparing two nested binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.05)
Type I Contamination

Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure C.5. Plots of observed levels of the GLR tests, the GPHDV tests using optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using optimal \( h \) vs. common sample sizes \( n_c \) for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.05)
Type I Contamination

Binomial Means Approximately Equal to 10
B2 Model (Ni's are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (Ni's are Different)

Figure C.6. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.05)
Figure C.7. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)

Type I Contamination

Binomial Means Approximately Equal to 10

B2 Model (Ni's are all Equal)

Binomial Means Approximately Equal to 10

B3 Model (Ni's are Different)

Figure C.8. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.05)
Type II Contamination
Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure C.9. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.05)
Figure C.10. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)  
Type II Contamination

Binomial Means Approximately Equal to 2  
B1 Model

Binomial Means Approximately Equal to 10  
P Model

Binomial Means Approximately Equal to 50  
N Model

Figure C.11. Plots of observed levels of the GLR tests, the GPHDV tests using optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using optimal \( h \) vs. common sample sizes \( n_c \) for comparing two nested binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)
Type II Contamination
Binomial Means Approximately Equal to 10
B2 Model (N's are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (N's are Different)

Figure C.12. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.05)
Figure C.13. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_e$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.01)
Figure C.14. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)
No Contamination
Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure C.15. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)
No Contamination
Binomial Means Approximately Equal to 10
B2 Model (Ni's are all Equal)

Tests for Nested Models (Nominal Level: 0.01)
No Contamination
Binomial Means Approximately Equal to 10
B3 Model (Ni's are Different)

Figure C.16. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.01)
Tests for Goodness-of-fit (Nominal Level: 0.01)
Type I Contamination

Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure C.17. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.01)
Tests for Goodness-of-fit (Nominal Level: 0.01)
Type I Contamination
Binomial Means Approximately Equal to 10
B2 Model (N's are all Equal)

Figure C.18. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)
Type I Contamination

Binomial Means Approximately Equal to 2
B1 Model

Figure C.19. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)
Type I Contamination
Binomial Means Approximately Equal to 10
B2 Model (N1's are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (N1's are Different)

Figure C.20. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.01)
Tests for Goodness-of-fit (Nominal Level: 0.01)
Type II Contamination
Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure C.21. Plots of observed levels of the GLR tests, the GPHDV tests using optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using optimal \( h \) vs. common sample sizes \( n_c \) for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.01)
Figure C.22. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)
Type II Contamination

Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure C.23. Plots of observed levels of the GLR tests, the GPHDV tests using optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)
Type II Contamination
Binomial Means Approximately Equal to 10
B2 Model (N's are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (N's are Different)

Figure C.24. Plots of observed levels of the GLR tests, the GPHDV tests using optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using optimal \( h \) vs. common sample sizes \( n_c \) for comparing two nested binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.01)
APPENDIX D

SIMULATION RESULTS FOR HYPOTHESIS TESTING OF 5% AND 1%

NOMINAL LEVELS OF SECTION 4.3.3.2
Tests for Goodness-of-fit (Nominal Level: 0.05)
No Contamination

Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure D.1. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.05)
No Contamination
Binomial Means Approximately Equal to 10
B2 Model (N_i's are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (N_i's are Different)

Figure D.2. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)

No Contamination
Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure D.3. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.05)
Figure D.4. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.05)
Type I Contamination

Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure D.5. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.05)
Type I Contamination
Binomial Means Approximately Equal to 10
B2 Model (N's are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (N's are Different)

Figure D.6. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)
Type I Contamination
Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure D.7. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)
Type I Contamination

Binomial Means Approximately Equal to 10
B2 Model (Nis are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (Nis are Different)

Figure D.8. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.05)
Figure D.9. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.05)
Tests for Nested Models (Nominal Level: 0.05)
Type II Contamination
Binomial Means Approximately Equal to 10
B2 Model (N's are all Equal)

Figure D.10. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for comparing two nested binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.05)
Tests for Goodness-of-fit (Nominal Level: 0.01)
No Contamination

Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure D.11. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.01)
Tests for Goodness-of-fit (Nominal Level: 0.01)
No Contamination
Binomial Means Approximately Equal to 10
B2 Model (Ni’s are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (Ni’s are Different)

Figure D.12. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal h, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)
No Contamination

Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure D.13. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; No contamination; Nominal level 0.01)
Figure D.14. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; No contamination; Nominal level 0.01)
Tests for Goodness-of-fit (Nominal Level: 0.01)
Type I Contamination

Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure D.15. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for testing goodness-of-fit of binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.01)
Tests for Goodness-of-fit (Nominal Level: 0.01)
Type I Contamination
Binomial Means Approximately Equal to 10
B2 Model (N1's are all Equal)

Binomial Means Approximately Equal to 10
B3 Model (N1's are Different)

Figure D.16. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for testing goodness-of-fit of binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)
Type I Contamination

Binomial Means Approximately Equal to 2
B1 Model

Sample Sizes (n1=n2=...=nm)

Binomial Means Approximately Equal to 10
P Model

Sample Sizes (n1=n2=...=nm)

Binomial Means Approximately Equal to 50
N Model

Sample Sizes (n1=n2=...=nm)

Figure D.17. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal \( h \), the GPHDV tests using \( h = 0.5 \), the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal \( h \) vs. common sample sizes \( n_c \) for comparing two nested binomial GLMs. (B1, P, and N models; Type I contamination; Nominal level 0.01)
Figure D.18. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type I contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)
Type II Contamination

Binomial Means Approximately Equal to 2
B1 Model

Binomial Means Approximately Equal to 10
P Model

Binomial Means Approximately Equal to 50
N Model

Figure D.19. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B1, P, and N models; Type II contamination; Nominal level 0.01)
Tests for Nested Models (Nominal Level: 0.01)

Type II Contamination

Binomial Means Approximately Equal to 10

B2 Model (N’s are all Equal)

Figure D.20. Plots of observed levels of the GLR tests, the GPHDV tests using PBoot optimal $h$, the GPHDV tests using $h = 0.5$, the ordinary GHDV tests, and the NPB-GPHDV tests using PBoot optimal $h$ vs. common sample sizes $n_c$ for comparing two nested binomial GLMs. (B2 and B3 models; Type II contamination; Nominal level 0.01)
CURRICULUM VITAE

Hu ey Yan

(August 2000)

Education:

BS in Statistics, Fu-Jen Catholic University,
Taipei, Taiwan, Republic of China. (6/86). GPA: 3.30

MS in Statistics, Utah State University,
Logan, Utah, USA. (8/91). GPA: 3.87

Ph.D. in Statistics, Utah State University,
Logan, Utah, USA. (expected 5/01). GPA: 3.92 (current)

Experience:

Graduate Grading Assistant, (Grader and Tutor), 9/89-5/90
Department of Mathematics and Statistics, Utah State University.

Graduate Grading Assistant, (Recitation Leader), 9/90-5/92
Department of Mathematics and Statistics, Utah State University.

Graduate Teaching Assistant, 9/92-5/96
Department of Mathematics and Statistics, Utah State University.
Statistical Consultant, 8/00-present
Department of Mathematics and Statistics, Utah State University

Honors and Awards:
Ph.D. Graduate Student Researcher of the Year Award,
Department of Mathematics and Statistics, Utah State University, 2001
International Teaching Assistant of the Year Award,
Utah State University, 1996
Robins Award for Teaching Assistant of the Year,
Utah State University, 1995
Teaching Assistant of the Year Award,
College of Science, Utah State University, 1995
Outstanding Teaching Award,
Department of Mathematics and Statistics, Utah State University, 1993
Elected as a member of the Honor Society of Phi Kappa Phi, 1993
Outstanding Student Award, Academics,
Department of Mathematics and Statistics, Utah State University, 1990

Professional Memberships:
American Mathematical Society (AMS)
American Statistical Association (ASA)
Institute of Mathematical Statistics (IMS)