ABSTRACT

Title of dissertation: CONTRIBUTIONS TO BAYESIAN STATISTICAL MODELING IN PUBLIC POLICY RESEARCH

Kevin D. Dayaratna, Doctor of Philosophy, 2014

Dissertation directed by: Professor Benjamin Kedem
Department of Mathematics

This dissertation improves existing Bayesian statistical methodologies and applies these improvements to a variety of important public policy questions. The manuscript is divided into six chapters. The first chapter provides an overview of the various chapters of the dissertation. The second chapter improves existing Bayesian binary logistic regression methodologies using polynomial expansions as an alternative to existing Markov Chain Monte Carlo (MCMC) methods. Our improvements make the estimation technique quite useful for a variety of applications. We also demonstrate the methodology to be considerably faster than existing MCMC methods. These computational gains are quite useful for models analyzing large data sets involving high-dimensional parameter spaces. We apply this methodology to a child poverty data set to analyze the potential causes of child poverty. The next chapter improves upon a well-known technique in semiparametric modeling known as density ratio estimation. This methodology is useful in principle; however, it suffers from one primary limitation - The technique has thus far been incapable

of modeling individual-level heterogeneity. Modeling heterogeneity is important as there is often no a priori reason to believe that different individuals (or observations) in a data set will behave in an identical manner. We ameliorate this limitation in the third chapter of this dissertation by adapting density ratio estimation methods to accommodate individual-level heterogeneity. We apply this new methodology to an analysis of the efficacy of medical malpractice reform across the country. In the fourth chapter of this dissertation, we shift our focus toward improving Bayesian credible interval estimation via semiparametric density ratio estimation. We do so by applying an innovative adaptation of the methodology, known as out of sample fusion, to posterior samples from a hierarchical Bayesian linear model looking at the efficacy of the welfare reform of the 1990s. In the fifth chapter, we extend the application of this methodology to credible interval estimation of a hierarchical generalized linear model used for analyzing terrorism data in a number of major conflicts across the globe. We use our results to offer some prescriptive policy suggestions regarding counterterrorism policy. The final chapter concludes the dissertation and offers a number of suggestions for further research. We emphasize that the modeling contributions presented in this dissertation are useful in myriads of other applied problems beyond just the public policy applications presented here.

CONTRIBUTIONS TO BAYESIAN STATISTICAL MODELING IN PUBLIC POLICY RESEARCH

by

Kevin D. Dayaratna

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Benjamin Kedem, Chair/Advisor
Professor Tobias Von Petersdorff
Professor Yuan Liao
Professor David Hamilton
Professor P.K. Kannan (Dean's Representative)

Dedication

To Daddy

Acknowledgments

My experience in Graduate School at the University of Maryland is one that will remain with me for the rest of my life. There are so many people I'd like to acknowledge that these first few pages cannot accommodate them all. Firstly, I would like to thank both of my parents, Dr. Lama Dayaratna and Mrs. Esme Dayaratna, as well as my two siblings, Arnal and Sandra, for their endless support and encouragement over the course of my entire life. My accomplishments would almost surely not have been possible without them. I'd also like to of course thank my wife Sashi for her tremendous support as well. Others I'd also like to thank include Ron Matteotti and my in-laws Mr. Gamini Abeysekara and Mrs. Kalyani Abeysekara. I'd also like to thank many of my other relatives including Asantha Kempitya, Obashini Kempitiya, Dharshika Kempitiya Sooriyabandara, Gihan Sooriyabandara, Hemali de Silva, Harsha Rajapaksha, Udeshi Hargett, Nirmala Abeysinghe, and Channa Abeysinghe to name a few.

There are also of course plenty of others outside my family whom I would like to thank. Firstly, from the bottom of my heart, I would like to express my utmost gratitude to my adviser Professor Benjamin Kedem. Professor Kedem has not only been the best adviser I could have ever asked for but has also been a great friend and colleague. Much of this dissertation builds upon many of his excellent ideas, and he has provided tremendous input and guidance to improve upon previous drafts of this dissertation.

I would also like to thank Professor Paul Smith who admitted me to the

program and served as my professor for some of my coursework. In addition, I would like to thank a number of other professors in the department including Professors Eric Slud, Abram Kagan, Leonid Koralov, Sandra Cerrai, Tobias von Petersdorff, Yuan Liao, and David Hamilton to name a few. Outside Maryland's Mathematics department, I would like to thank my colleague Professor P.K. Kannan of the Robert H. Smith School of Business, whom I had the privilege of publishing other research with outside this dissertation [22]. I would also like to thank Professor Jordan Boyd-Graber of the iSchool for introducing me to the concept of non-parametric Bayesian modeling used in Chapter 5 of this dissertation. Another faculty member I would like to thank is Professor Cindy Clement of the Economics department for giving me the opportunity to serve as a Lecturer in her department for each of the last three years.

Additionally, there are numerous classmates at Maryland that I would also like to thank including Eddie Kim, Xuan Yao, Xuan Liu, Joyce Hsiao, Kwame Okrah, Hisham Talukder, Arseniy Zakharov, Joseph Paulson, David Shaw, Daniel Han, Jong Jun Lee, Clare Wickman, Jim Greene, William Waldron, Catherine Ochalek, Jeff Tan, Wenqing Hu, Wen Zhou, Lemeng Pan, Yue Tian, Zi Ding, Sean Burke, Joel Witten, Peggy Tseng, Gauri Kulkarni, and Anshuman Sinha among others. I would also like to especially thank Paul Tschirhart of the University's Writing Fellows Program for his advice about LaTeX coding and help formatting this dissertation. I would also like to thank the Mathematics department's great staff including Celeste Regalado, Linette Berry, Haydee Hidalgo, and Alverda McCoy.

Outside the University of Maryland, there are also many people I would like to

thank. Firstly, I am indebted to Professor Steven J. Miller of Williams College. Steve has essentially served as my "mathematics mentor" since I was an undergraduate at U.C. Berkeley. Over the years, we have become friends and colleagues, and I've had the privilege of co-authoring a number of publications with him [23, 24, 108]. He has also provided very insightful comments to this dissertation. I would also like to thank Professor Eric Bradlow of the University of Pennsylvania for helpful input on the Chapter 2 of this dissertation.

There are many people from my undergraduate days at the University of California, Berkeley I would also like to thank. Special thanks go to my good friend Gagan Tara Nanda whom I studied mathematics with throughout the course of my studies at Berkeley. I'd like to thank my undergraduate adviser, Professor L. Craig Evans, who was instrumental in getting me interested in continuing to study higher-level mathematics. I'd also like to of course thank a number of professors that I had the privilege of learning from at Berkeley including Professors Richard Borcherds, Maciej Zworski, Michael Hutchings, and John Neu. Studying mathematics from these and other professors interested me greatly in pursuing graduate-level studies. I would also like to thank one of my political science professors, Professor Darren Zook, for getting me interested in terrorism analysis and ethnic conflict, which form a component of this dissertation.

There are plenty of friends I'd also like to thank. I'd like to thank Alex Levy who has been my good friend for nearly twenty years and the best man at my wedding. I'd also like to thank Hovannes Abramyan, Mike Taylor, and Richard Nazareth who not only served as groomsmen at my wedding but have also been great friends since my time at U.C. Berkeley. I would especially like to thank Hovannes and his colleague Mark Crain for providing the data that they used in their work with the Pacific Research Institute for Chapter 3 of this dissertation. Other U.C. Berkeley friends I would like to thank include Carl Densing, Elaine Gin, Yuriy Pasko, Mike McFarlane, Dennis Bedford, Cameron Huey, Laura Russell, Nat Lipanovich, Erica Batres, Erik Johannessen, Steve Berkovich, Michael Klein, Sam Lee, Eddie Kim, Taiwan Udtamadilok, Rosanna Leroe-Munoz, David Timmons, and Rebekah Hsieh among others. I'd also like to thank A.J. Hergenroder, Annie Moskovian, Jocelyn Hsiao, Ben Ouyang, Laurie Kaufman, Rebekka Levy, Chris Luise, and Pamitha Weeresinghe. Special thanks also go to the Rhodes, Pillai, and Pathirana families for their friendship and encouragement over the years.

Finally, I would like to say thank you to many of my colleagues at The Heritage Foundation. In particular, I'd like to thank my colleagues in the Center for Data Analysis - namely, Rea Hederman, Salim Furth, Drew Gonshorowski, Patrick Tyrrell, James Sherk, David Kreutzer, David Muhlhausen, John Ligon, Rachel Greszler, and Filip Jolevski. I'd also like to thank Derrick Morgan, Bill Beach, Robert Rector, Leslie Merkle, Nina Owcharenko, Alyene Senger, Jennifer Marshall, Rachel Sheffield, Bob Moffit, David Azerrad, Katie Tubb, Elinor Renner, and Pamela Ouzts. I would especially like to thank Rea Hederman and Robert Rector for providing the data used in Chapters 2 and 4 of this dissertation as well as for insightful discussions. I would also like to think Jim Carafano for helpful comments on Chapter 5 of this dissertation on statistical inferences for counterterrorism policy.

Contents

List of Figures

Chapter 1: Introduction

1.1 Introduction

In public policy research, questions abound about the direction of our country. In recent years, rigorous statistical analysis has been a useful tool for answering many of these questions [42]. This dissertation looks at a number of such policy questions improving existing statistical methodologies in the process.

For example, in our next chapter, we look at the determinants of child poverty and improve existing Bayesian estimation techniques to do so. Understanding the root causes of poverty is a fundamental question that has consistently plagued policy researchers for decades. With nearly $16 trillion spent on federal welfare programs since President Lyndon Johnson began his War on Poverty and with the Obama Administration expected to spend over another $10 trillion over the course of the next decade, millions of American children continue to live in poverty [131]. Any meaningful policy proposals should be based on addressing the causes of poverty and not just the apparent symptoms.

Although a number of studies have attempted to identify these causes, no studies, to our knowledge, have done so including individual-level heterogeneity in the associated statistical models. Incorporating heterogeneity is very important as different families will almost surely respond differently to potential causes. Unfortunately, however, for the binary logistic regression models typically called upon to help answer this question, positing a distribution for heterogeneity does not really allow for closed-form inferences. As a result, researchers often have to resort to numerical techniques such as Markov Chain Monte Carlo (MCMC) methods to estimate the associated models. These methods can be difficult and time consuming to obtain convergence, particularly for large data sets.

We address this issue in Chapter 2 by presenting an alternative Bayesian estimation technique for binary logistic regression using polynomial expansions that allows for closed form inferences, enabling researchers to make direct inferences about the population. Miller, Bradlow, and Dayaratna (2006) also presented a similar method; however, their result was quite limited as it was restricted to using a single-sided prior distribution [108]. We assuage this limitation by allowing the researcher to draw from one of the most flexible and commonly used prior distributions - the normal distribution. After deriving our polynomial expansions, we present a number of numerical simulations to illustrate the usefulness and advantages of our approach. We also estimate our model on a large poverty data set from the Current Population Survey to study the determinants of child poverty. In particular, we find that marital status of parents, parental age, parental education, whether the parents are working full time, and the number of children living in a household all significantly influence whether a child grows up in poverty. We discuss the resulting policy implications and conclude with a discussion of potential avenues for future research.

In Chapter 3, we examine the efficacy of medical malpractice reforms instituted throughout the country over the course of the last decade. In the process, we improve on existing semiparametric density ratio estimation methodologies. Density ratio estimation (DRE) is a well-known semiparametric modeling technique that has been around for decades. Although the DRE method has proven to be very useful in statistical modeling, it suffers from one primary limitation. In particular, the method has thus far been incapable of modeling individual-level heterogeneity.

We ameliorate this limitation in Chapter 3 to enable DRE methods to model individual-level heterogeneity. We perform a series of numerical simulations, along with goodness of fit computations, to illustrate the efficacy of our approach. We show that this new approach outperforms existing semiparametric density ratio estimation methods. After our numerical simulations, we apply our approach to medical malpractice loss data from the previous decade to quantify the probability of extreme losses. Our results indicate the success of some recently instituted medical malpractice reforms.

Subsequently, in our fourth chapter, we shift our focus toward welfare policy, conducting a rigorous Bayesian analysis of the welfare reform of the 1990s, providing some improvements to Bayesian credible interval estimation techniques in the process. First popularized by Ronald Reagan in his classic 1964 "A Time for Choosing" speech, the concept of welfare reform has been a hot topic of public policy research for decades [127]. The Personal Responsibility and Work Authorization Act of 1996 was one of the nation's most comprehensive efforts at welfare reform [118]. The law's primary aim was to transform one of America's major welfare programs away from a system fostering dependency and into a program providing temporary assistance to enable people to become contributing members of society.

In Chapter 4, we utilize hierarchical Bayesian linear modeling to rigorously quantify the impact of this law. In the process, we improve upon existing Bayesian interval estimation methods by calling upon semiparametric DRE method thus far used only for frequentist statistical modeling. We find that the welfare reform of the 1990s was quite successful in getting people back to work and can be improved upon even further. We conclude by discussing the resulting policy implications.

In Chapter 5, we shift our focus to foreign policy, particularly toward combating terrorism. Terrorism has been around for generations and understanding how to fight terrorists has been a question vexing policymakers throughout the world. In Chapter 5, we rigorously analyze terrorist attack data from four major conflicts across the globe - from Afghanistan, Iraq, Sri Lanka, and Northern Ireland - to help policymakers answer this very question. We utilize both parametric and non-parametric (hence generalized) Bayesian logistic regression techniques to do so, and again improve upon Bayesian credible interval estimation in the process by using semiparametric DRE methods. We thus extend the application of the semiparametric DRE method used in Chapter 4 for linear models to this generalized linear model. Our study helps shed light on the factors that influence the success of terrorist attacks in each conflict, providing policymakers with advice about how to more effectively combat this very dangerous enemy.

Finally, in Chapter 6, we discuss some conclusions as well as potential avenues for future research. We emphasize that the modeling contributions presented here are applicable to myriads of other fields beyond just the public policy applications looked at in this dissertation.

Chapter 2: A Rigorous Examination of the Determinants of Child Poverty in America via Closed-Form Bayesian Inferences with Implications for Welfare Reform

2.1 *Introduction*

2.1.1 *What are the Causes of Child Poverty?*

Ever since President Lyndon Johnson began his War on Poverty in 1964, the U.S. Federal Government has spent an exorbitant amount (estimated to be as much as \$15.9 trillion) on welfare programs and on other forms of cash assistance to the poor. Such programs include means-tested welfare programs such as Temporary Assistance for Needy Families (TANF), the Earned Income Tax Credit (EITC), Supplemental Security Income (SSI), food stamps, public housing, and Medicaid among others [131]. As millions of Americans continue to live in poverty today, some researchers have argued that many of these programs have been largely ineffective and have the potential to trap people in poverty. In his 1984 seminal work *Losing Ground*, for example, Charles Murray of the American Enterprise Institute suggested that many of these welfare programs were actually encouraging dependence and ironically crippling the people they were intended to help [111]. Many such researchers have consequently argued that meaningful public policy proposals on these issues should be grounded on an understanding of the root causes of poverty and not just the mere symptoms [130, 136].

Although some studies have attempted to identify these causes, no such research, to our knowledge, have incorporated individual-level heterogeneity in its statistical modeling [130, 145]. Incorporating heterogeneity is of utmost importance in these models as there is no a priori reason to believe that all families will respond to potential causes in the same manner. From a statistical perspective, assuming homogeneous response coefficients in a model can lead to the researcher ignoring potential variability and can consequently result in misleading statistical inferences.

We fill this gap in the extant literature by examining the potential causes of child poverty via Bayesian logistic regression. The contribution of our study is two fold. From a policy perspective, we look at the determinants of child poverty, while incorporating individual-level heterogeneity into our model. Our incorporation of individual-level heterogeneity allows us to make more informative statistical inferences compared to methods that simply just assume homogeneity. Methodologically, our approach makes Bayesian estimation involving large data sets, such as the child poverty data set examined here, much more feasible.

2.1.2 Improvements to Bayesian Computation

From fields ranging from economics to public policy to medicine to professional sports, logistic regression has become one of the most widely used tools in applied statistical research. For example, logistic regression models have been used to help understand the determinants of depression and suicide, examine consumer choice in marketing research, model medical outcomes pertaining to various illnesses, and shed light on baseball player hitting performance among others [1, 7, 11, 58–60, 133, 164, 167].

With improvements in statistical computing power over the course of the past two decades, incorporating heterogeneity in these and other models has become increasingly common in the applied statistical literature. Researchers have incorporated heterogeneity by choosing amongst a variety of methodologies, including a parametric Bayesian approach, a non-parametric Bayesian approach, or even a frequentist finite mixture approach [49, 72, 74, 100]. Parametric Bayesian models are often used to incorporate individual-level heterogeneity. Generally, with limited data per individual in a data set, assuming a different parametrization for each individual renders a model statistically unidentifiable, making estimation virtually impossible. Researchers will typically assume that these individual response coefficients are all drawn from their own lower-dimensional probability distribution. They can then estimate these models from an empirical Bayesian perspective or from a fully Bayesian perspective by imposing priors on the parameters of the heterogeneity distributions themselves [49, 110].

Although "nice" in principle, incorporating individual-level heterogeneity is often concomitant with the drawbacks of the computational complexity associated with numerical computation. For example, numerical methods such as quadrature, simulated maximum likelihood, and Markov Chain Monte Carlo (MCMC) methods can be difficult and time consuming to estimate, especially for large data sets involving high-dimensional parameter spaces [50, 138]. Additionally, commonly-used MCMC methods suffer from the drawback of sensitivity to starting values and can consequently result in a significant amount of simulation error.

As a result, a number of researchers have approached alternative techniques to make Bayesian computation more feasible. For example, Everson and Bradlow (2002) used polynomial expansions to approximate the posterior distributions of the beta-binomial random variables using a class of prior distributions previously considered non-conjugate [39]. Similarly, Bradlow et al (2002) used polynomial expansions to improve on researchers' ability to make posterior inferences about the negative binomial distribution [9]. In subsequent research, McShane et al (2008) used similar techniques to improve on Weibull count model estimation [105].

Miller et al (2006) used polynomial expansions to solve the problem looked at in this research, namely for binary logistic regression [108]. Their approach, however, suffered from a serious limitation by requiring that the prior distribution be single-sided. Consequently, their result, although nice in principle, is very limited in scope as it is often difficult to know a priori the signs of the coefficients beforehand. We assuage this limitation by looking at the same problem but instead allowing the researcher to draw from the one of the richest and most commonly used two-sided prior distributions - The normal distribution.

In particular, we derive a marginalized likelihood for the binary logit model using polynomial expansions. We begin by reviewing the Miller et al (2006) approach and its subsequent limitations. We then proceed by ameliorating these limitations, assuming that the response coefficients are drawn from independent normal distributions. We subsequently generalize this result by allowing correlations amongst the coefficients. Afterwards, we allow for dependence of the prior distributions on various covariates and then subsequently allow for the choice of non-normal prior distributions.

Our model can be estimated via the method of maximum marginal likelihood (MML) from which empirical Bayesian inferences can be made, allowing us to make direct inferences about the population [110, 147]. We present a number of numerical simulations to illustrate the usefulness of our approach as well as its advantages over existing MCMC methods. We also utilize our approach to answer an important question in public policy research; particularly, identifying the causes of child poverty.

2.2 Problem Formulation

Consider a data set obtained from $i \in \{1,\ldots,I\}$ individuals (units) having $j \in \{1,\ldots,J\}$ categories (e.g. illness, product brand, etc) measured on $t \in \{1,\ldots,N_i\}$ occasions (repeated measures). As is standard, we define

$$y_{ijt} = \begin{cases} 1 & \text{if outcome occurs for individual } i \text{ pertaining to category } j \text{ at time } t \\ 0 & \text{otherwise,} \end{cases} \tag{2.1}$$

where $p_{ijt} = \text{Prob}(y_{ijt} = 1)$ is the probability of a particular outcome occurring (e.g. living in poverty, purchase of a product, mortality of a patient) for the i^{th} individual pertaining to the j^{th} category on the t^{th} occasion. Additionally, let $p = 1,\ldots,P$ represent a set of attributes pertaining to the covariates, with corresponding values $x_{ijt,p} \geq 0$ such that $X_{ijt}^T = (x_{ijt,1},\ldots,x_{ijt,P})$. To account for residual effects not manifested in the coefficient estimates for the explanatory variables, we can allow $x_{ijt,1} = 1$ defining category-level intercepts.

Multiplying over all individuals, categories, and occasions, we obtain the standard logit likelihood of the data, $Y = (y_{ijt})$:

$$P(Y|\beta) = \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i} \frac{e^{y_{ijt}X_{ijt}^T\beta_i}}{1+e^{X_{ijt}^T\beta_i}}, \tag{2.2}$$

where $\beta_i = (\beta_{i,1},\ldots,\beta_{i,P})$ is the coefficient vector for the i^{th} individual with p^{th} variable-specific coefficient, $\beta_{i,p}$ and $\beta = (\beta_1,\ldots,\beta_I)$.

Allowing our model to accommodate heterogeneity is quite important in modeling real world phenomena, as there is no reason to believe that all individuals will behave in an identical manner. For example, in modeling consumer purchase behavior, customers purchasing a product will almost surely differ in how they respond to prices or promotions. In modeling baseball hitting performance, different sluggers will respond differently to different pitching styles. Additionally, in public policy research, looked at in this study, different families will respond differently to various factors that may or may not contribute to them living in poverty.

We can model heterogeneity across individuals by allowing each $\beta_{i,p}$ to be drawn from probability distributions. To start out, we discuss the Miller et al (2006) approach and the limitations of their result. Subsequently, we ameliorate these limitations and make the mode much for useful for applying to real-world problems.

2.2.1 Polynomial Expansions of the Binary Logit Model

We are interested in the following marginalized likelihood which we intend to maximize over our parameter space:

$$P(Y|\Omega) = \int_{\beta} P(Y|\beta) N(\beta|\Omega) \mathrm{d}\beta. \tag{2.3}$$

In the above equation, $P(Y|\beta)$ is our standard logit likelihood with a prior distribution $N(\beta|\Omega)$ and Ω represents the parameters of our prior distribution. As mentioned above, we have non-negative explanatory variables $X_{ijt}^T = (x_{ijt,1}, \ldots, x_{ijt,P})$ and binary dependent variables $Y = (y_{ijt})$. We intend to maximize the above marginalized likelihood over our prior distribution's parameter space. Specifically, our logit likelihood is:

$$P(Y|\beta) \; = \; \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{t=1}^{N_i} \frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}}, \tag{2.4}$$

It is the heterogeneity across $i = 1, ..., I$ individuals in their $\beta_{i,p}$ coefficients that we are modeling by allowing these parameters to follow $N(\beta|\Omega)$. Due to the fact that the β_i appears in both the numerator and denominator of (2.4), performing the integration in (2.3) analytically for most choices of heterogeneity distributions without any numerical approximations is essentially impossible.

We can take a series expansion approach to this problem and rewrite $P(Y|\beta)$ as follows:

$$\begin{aligned}
P(Y|\beta) &= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i} \frac{e^{y_{ijt}X_{ijt}^T\beta_i}}{1+e^{X_{ijt}^T\beta_i}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i} e^{y_{ijt}X_{ijt}^T\beta_i} \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i} \frac{1}{1+e^{X_{ijt}^T\beta_i}} \\
P(Y|\beta) &= P_1(Y|\beta)P_2(Y|\beta). \qquad (2.5)
\end{aligned}$$

We refer to the second factor above as $P_2(Y|\beta)$ although it does not depend on Y. If we assume $X_{ijt}^T\beta_i < 0$, we can expand $P_2(Y|\beta)$ via a geometric series expansion as follows [108]:

$$\begin{aligned}
P_2(Y|\beta) &= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i} \frac{1}{1+e^{X_{ijt}^T\beta_i}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{k_{ijt}X_{ijt}^T\beta_i}. \qquad (2.6)
\end{aligned}$$

Putting together the pieces, we therefore have when $X_{ijt}^T\beta_i < 0$:

$$
\begin{aligned}
P(Y|\beta) &= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\frac{e^{y_{ijt}X_{ijt}^T\beta_i}}{1+e^{X_{ijt}^T\beta_i}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i} e^{y_{ijt}X_{ijt}^T\beta_i}\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\frac{1}{1+e^{X_{ijt}^T\beta_i}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i} e^{y_{ijt}X_{ijt}^T\beta_i}\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{k_{ijt}X_{ijt}^T\beta_i} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{y_{ijt}X_{ijt}^T\beta_i+k_{ijt}X_{ijt}^T\beta_i} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{(y_{ijt}+k_{ijt})X_{ijt}^T\beta_i} \\
P(Y|\beta) &= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{(y_{ijt}+k_{ijt})\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}}. \qquad (2.7)
\end{aligned}
$$

If, on the other hand, we assume $X_{ijt}^T\beta_i > 0$, we can also use a geometric series expansion:

$$
\begin{aligned}
P_2(Y|\beta) &= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\frac{1}{1+e^{X_{ijt}^T\beta_i}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\frac{e^{-X_{ijt}^T\beta_i}}{1+e^{-X_{ijt}^T\beta_i}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i} e^{-X_{ijt}^T\beta_i}\frac{1}{1+e^{-X_{ijt}^T\beta_i}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i} e^{-X_{ijt}^T\beta_i}\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{-k_{ijt}X_{ijt}^T\beta_i} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{-X_{ijt}^T\beta_i-k_{ijt}X_{ijt}^T\beta_i} \\
P_2(Y|\beta) &= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{-(1+k_{ijt})\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}}. \qquad (2.8)
\end{aligned}
$$

And again putting together the pieces:

$$\begin{aligned}
P(Y|\beta) &= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\frac{e^{y_{ijt}X_{ijt}^T\beta_i}}{1+e^{X_{ijt}^T\beta_i}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}e^{y_{ijt}X_{ijt}^T\beta_i}\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\frac{1}{1+e^{X_{ijt}^T\beta_i}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}e^{y_{ijt}X_{ijt}^T\beta_i}\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{-(1+k_{ijt})X_{ijt}^T\beta_i} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{y_{ijt}X_{ijt}^T\beta_i-(1+k_{ijt})X_{ijt}^T\beta_i} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{(y_{ijt}-1-k_{ijt})X_{ijt}^T\beta_i} \\
P(Y|\beta) &= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{(y_{ijt}-1-k_{ijt})\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}}. \qquad (2.9)
\end{aligned}$$

In the next section we utilize these series expansions to derive closed-form expressions from which we can make Bayesian inferences.

2.2.2 *Closed-Form Bayesian Inference via Polynomial Expansions as Described In Miller et al (2006)*

Ideally, one would like to allow each $\beta_{i,p}$ to follow two sided prior distribution, under such circumstances we would have a combination of both of the above situations, as well as when $X_{ijt}^T\beta_i = 0$:

$$\begin{aligned}
P(Y|\beta) &= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\frac{e^{y_{ijt}X_{ijt}^T\beta_i}}{1+e^{X_{ijt}^T\beta_i}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\left[\frac{e^{y_{ijt}X_{ijt}^T\beta_i}}{1+e^{X_{ijt}^T\beta_i}}(1)\right] \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\left[\frac{e^{y_{ijt}X_{ijt}^T\beta_i}}{1+e^{X_{ijt}^T\beta_i}}\left[I(X_{ijt}^T\beta_i>0)+I(X_{ijt}^T\beta_i<0)+I(X_{ijt}^T\beta_i=0)\right]\right] \\
P(Y|\beta) &= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\left[\frac{e^{y_{ijt}X_{ijt}^T\beta_i}}{1+e^{X_{ijt}^T\beta_i}}I(X_{ijt}^T\beta_i>0)+\frac{e^{y_{ijt}X_{ijt}^T\beta_i}}{1+e^{X_{ijt}^T\beta_i}}I(X_{ijt}^T\beta_i<0)\right. \\
&\quad\left.+\ \frac{e^{y_{ijt}X_{ijt}^T\beta_i}}{1+e^{X_{ijt}^T\beta_i}}I(X_{ijt}^T\beta_i=0)\right].
\end{aligned}$$

This can be rewritten as:

$$\begin{aligned}
P(Y|\beta) &= \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\left[\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{(y_{ijt}-1-k_{ijt})\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}}I(\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}>0)\right. \\
&\quad+\ \sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{(y_{ijt}+k_{ijt})\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}}I(\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}<0) \\
&\quad\left.+\ \frac{e^{y_{ijt}\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}}}{1+e^{\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}}}I(\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}=0)\right].
\end{aligned} \tag{2.10}$$

As a result, the marginalized likelihood is:

$$\begin{aligned}
P(Y|\Omega) &= \int_{\beta}P(Y|\beta)N(\beta|\Omega)\mathrm{d}\beta \\
&= \int_{\beta}\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\frac{e^{y_{ijt}X_{ijt}^T\beta_i}}{1+e^{X_{ijt}^T\beta_i}}N(\beta|\Omega)\mathrm{d}\beta \\
&= \int_{\beta}\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\frac{e^{y_{ijt}X_{ijt}^T\beta_i}}{1+e^{X_{ijt}^T\beta_i}}\mathrm{d}P_{\beta_{i,p}} \\
P(Y|\Omega) &= \int_{\beta}P(Y|\beta)\mathrm{d}P_{\beta_{i,p}}
\end{aligned}$$

where $P_{\beta_{i,p}}$ is the measure induced by $\beta_{i,p}$ on measurable space $(S_{i,p}, F_{i,p})$.

When Miller et al (2006) looked at this problem, the authors attempted to integrate each $\beta_{i,p}$ individually for every potential value of i and p [108]. For a two-sided heterogeneity distribution, such as a normal heterogeneity distribution, this marginalization would involve:

$$\begin{aligned} P(Y|\Omega) &= \int_\beta P(Y|\beta)N(\beta|\Omega)\mathrm{d}\beta \\ &= \int_\beta \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\left[\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{(y_{ijt}-1-k_{ijt})\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}}I(\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}>0)\right. \\ &\quad + \sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{(y_{ijt}+k_{ijt})\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}}I(\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}<0) \\ &\quad \left.+ \frac{e^{y_{ijt}\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}}}{1+e^{\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}}}I(\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}=0)\right]\cdot\prod_{p=1}^{P}\frac{1}{\sqrt{2\pi}\sigma_p}\cdot e^{\frac{-(\beta_{i,p}-\mu_p)^2}{2\sigma_p^2}}\mathrm{d}\beta_{\mathrm{i,p}} \end{aligned} \quad (2.11)$$

As the range of $\beta_{i,p}$ is the entire real line, the limits of the integration space differ for the first and second integrals depending on whether $\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}<0$ or $\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}>0$. Miller et al (2006) noted that integrating over both spaces would result in "numerous, complicated subdivisions of the integration space." These subdivisions, they argued, rendered the integration "untenable" and precluded the derivation of "tractable closed-form expansions." As a result, the authors restricted their model to adhere to only one of the above cases, particularly $X_{ijt}^T\beta_i<0$[1] and required that the density $N(\beta|\Omega)$ being integrated over be a one-sided probability distribution. Making the assumption that $N(\beta|\Omega)$ was composed of independent gamma distributions $g(\beta_{i,p}|b_p,n_p)$ with parameters b_p and n_p (i.e. $N(\beta|\Omega)=\prod_{p=1}^{P}g(\beta_{i,p}|b_p,n_p)$), they derived the marginalized likelihood as follows:

[1] Miller et al (2006) actually parametrized their logit likelihood in a slightly different functional form as $P(Y|\beta)=\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\frac{e^{-y_{ijt}X_{ijt}^T\beta_i}}{1+e^{-X_{ijt}^T\beta_i}}$ and hence equivalently assumed that $X_{ijt}^T\beta_i>0$ and $\beta_{i,p}\geq 0\;\forall\; i,p$.

$$
\begin{aligned}
P(Y|\Omega) &= \int_\beta P(Y|\beta)N(\beta|\Omega)\mathrm{d}\beta \\
&= \int_\beta \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{N_i}\left[\sum_{k_{ijt}=0}^{\infty}(-1)^{k_{ijt}}e^{(y_{ijt}+k_{ijt})\sum_{p=1}^{P}x_{ijt,p}\beta_{i,p}}\prod_{p=1}^{P}g(\beta_{i,p}|b_p,n_p)\mathrm{d}\beta_{\mathrm{i,p}}\right] \\
&= \prod_{i=1}^{I}\sum_{k_{i11}=0}^{\infty}\cdots\sum_{k_{iJN_i}=0}^{\infty}(-1)^{\sum_{j=1}^{J}\sum_{t=1}^{N_i}k_{ijt}}\prod_{p=1}^{P}\int_{\beta_{i,p}=0}^{\infty}e^{-\sum_{j=1}^{J}\sum_{t=1}^{N_i}(y_{ijt}+k_{ijt})x_{ijt,p}\beta_{i,p}} \\
&\quad \cdot\frac{1}{b_p\Gamma(n_p)}\left(\frac{\beta_{i,p}}{b_p}\right)^{n_p-1}e^{-\beta_{i,p}/b_p}\mathrm{d}\beta_{\mathrm{i,p}} \\
&= \prod_{i=1}^{I}\sum_{k_{i11}=0}^{\infty}\cdots\sum_{k_{iJN_i}=0}^{\infty}(-1)^{\sum_{j=1}^{J}\sum_{t=1}^{N_i}k_{ijt}}\prod_{p=1}^{P}\left(\frac{1}{1+b_p\sum_{j=1}^{J}\sum_{t=1}^{N_i}(y_{ijt}+k_{ijt})x_{ijt,p}}\right)^{n_p}
\end{aligned}
$$

Having made assumptions that the explanatory variables $X_{ijt,p}$ were restricted to the set of non-negative integers, Miller et al (2006) borrowed some tools from analytic number theory to rewrite the above equation in terms of solutions to a system of Diophantine equations, which made estimating the model significantly more feasible from a computational perspective [109]. The interested reader is referred to Miller et al (2006) for a complete discussion of this methodology [108].

2.2.3 Bayesian Inference via Series Expansions Using a Two-Sided Heterogeneity Distribution

Although the Miller et al (2006) result is elegant mathematically, it is not particularly useful to implement in practice as in most applications it is generally unrealistic to a priori assume that the regression coefficients all have the same sign. However, for the case when $J = 1$ and $N_i = 1$, a simple transformation of variables leads to very clean and tractable integration, allowing us to integrate within distinct regions along the real line. Restricting J and N_i in this manner is quite reasonable for many applied statistical problems including cross-sectional data analysis with a single category (such as the child poverty application looked at later in this study), longitudinal analysis of a single individual (such as the baseball player hitting streak analysis conducted in Albrght (1993)), or analysis where the heterogeneity can be assumed across all observations of the data set (such as the terrorist attack data analysis conducted in Kyung et al (2012) or the data set used in the analysis of medical outcomes in Wisner (1990)) [1, 90, 108, 164].

Specifically, if we make the assumption that $p_i = \text{Prob}(y_i = 1)$ is the probability of a particular outcome occurring (e.g. living in poverty, patient mortality, purchase incidence, etc) for the i^{th} individual and again let $p = 1, \ldots, P$ represent a set of attributes describing the covariates, with corresponding values $x_{i,p}$ such that $X_i^T = (x_{i,1}, \ldots, x_{i,P})$ and take the product across all individuals i, the likelihood function is:

$$P(Y|\beta) = \prod_{i=1}^{I} \frac{e^{y_i X_i^T \beta_i}}{1 + e^{X_i^T \beta_i}}, \tag{2.12}$$

where $\beta_i = (\beta_{i,1}, \ldots, \beta_{i,P})$ and β are defined as before. Upon making these assumptions, we can recall the marginalization presented in (2.11):

$$\begin{aligned} P(Y|\Omega) &= \int_\beta P(Y|\beta) N(\beta|\Omega) \mathrm{d}\beta \\ &= \int_\beta \prod_{i=1}^{I} \Bigg[\sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i - 1 - k_i) \sum_{p=1}^{P} x_{i,p}\beta_{i,p}} I(\sum_{p=1}^{P} x_{i,p}\beta_{i,p} > 0) \\ &\quad + \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i + k_i) \sum_{p=1}^{P} x_{i,p}\beta_{i,p}} I(\sum_{p=1}^{P} x_{i,p}\beta_{i,p} < 0) + \frac{e^{y_i \sum_{p=1}^{P} x_{i,p}\beta_{i,p}}}{1 + e^{\sum_{p=1}^{P} x_{i,p}\beta_{i,p}}} I(\sum_{p=1}^{P} x_{i,p}\beta_{i,p} = 0) \Bigg] \\ &\quad \cdot \prod_{p=1}^{P} \frac{1}{\sqrt{2\pi}\sigma_p} e^{\frac{-(\beta_{i,p} - \mu_p)^2}{2\sigma_p^2}} \mathrm{d}\beta_{\mathrm{i,p}}. \end{aligned} \tag{2.13}$$

In particular, since we are assuming that the $\beta_{i,p}$ follow independent normal distributions for $p = 1, ..., P$ (i.e. $\beta_{i,p} \sim N(\mu_p, \sigma_p^2)$) it follows that $z_i = \sum_{p=1}^{P} x_{i,p}\beta_{i,p} \sim N(\sum_{p=1}^{P} x_{i,p}\mu_p, \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2)$. Therefore, if we define P_{z_i} as the measure induced by z_i on measurable space (T_i, G_i) having density with respect to Lebesgue measure:

$$f(z_i) = \frac{1}{\sqrt{2\pi \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}} e^{\frac{-(z_i - \sum_{p=1}^{P} x_{i,p}\beta_{i,p})^2}{2\sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}},$$

then:

$$\begin{aligned}
P(Y|\Omega) &= \int_\beta P(Y|\beta)N(\beta|\Omega)\mathrm{d}\beta \\
&= \int_\beta \prod_{i=1}^{I} \frac{e^{y_i X_i^T \beta_i}}{1+e^{X_i^T \beta_i}} N(\beta|\Omega)\mathrm{d}\beta \\
&= \int_\beta \prod_{i=1}^{I} \frac{e^{y_i X_i^T \beta_i}}{1+e^{X_i^T \beta_i}} \mathrm{d}P_{\beta_{i,p}} \\
P(Y|\Omega) &= \int_\beta P(Y|\beta)\mathrm{d}P_{\beta_{i,p}}.
\end{aligned}$$

Applying our transformation we can see that [149]:

$$\begin{aligned}
P(Y|\Omega) &= \int_\beta P(Y|\beta)\mathrm{d}P_{\beta_{i,p}} \\
&= \int_{T_I} \dots \int_{T_1} \prod_{i=1}^{I} P(y_i|z_i)\mathrm{d}P_{z_i} \\
&= \prod_{i=1}^{I} \int_{T_i} P(y_i|z_i)\mathrm{d}P_{z_i} \\
&= \prod_{i=1}^{I} \int_{z_i} P(y_i|z_i) \frac{1}{\sqrt{2\pi \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}} e^{\frac{-(z_i - \sum_{p=1}^{P} x_{i,p}\mu_p)^2}{2\sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}} \mathrm{d}z_i \\
P(Y|\Omega) &= \prod_{i=1}^{I} H_i, \qquad (2.14)
\end{aligned}$$

where:

$$\begin{aligned} H_i &= \int_{-\infty}^{\infty} \left[\sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i - 1 - k_i) z_i} I(z_i > 0) \right. \\ &\quad + \left. \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i + k_i) z_i} I(z_i < 0) + \frac{e^{y_i z_i}}{1 + e^{z_i}} I(z_i = 0) \right] \\ &\quad \cdot \frac{1}{\sqrt{2\pi \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}} e^{\frac{-(z_i - \sum_{p=1}^{P} x_{i,p} \mu_p)^2}{2 \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}} \mathrm{d}z_i. \end{aligned}$$

We can decompose H_i into a sum of three integrals, $H_{i,1}$, $H_{i,2}$, and $H_{i,3}$ where $H_i = H_{i,1} + H_{i,2} + H_{i,3}$.

Before we proceed, we present a simple integration lemma for integrating an exponential against a normal distribution with mean μ and variance σ^2.

Lemma 2.2.1 (Integrating an Exponential Against a Normal Distribution)**.**

$$\int_{c}^{\infty} e^{kx} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = e^{k\mu + \frac{k^2\sigma^2}{2}} \Phi\left(\frac{k\sigma^2 - c + \mu}{\sigma} \right) \tag{2.15}$$

$$\int_{-\infty}^{c} e^{kx} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = e^{k\mu + \frac{k^2\sigma^2}{2}} \Phi\left(-\frac{k\sigma^2 - c + \mu}{\sigma} \right) \tag{2.16}$$

where $\Phi(x)$ is the normal cumulative distribution function.

Proof:

$$\begin{aligned}
\int_c^\infty e^{kx}\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-\mu)^2}{2\sigma^2}}\mathrm{d}x &= \int_c^\infty \frac{1}{\sqrt{2\pi}\sigma}e^{kx-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x \\
&= \int_c^\infty \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-1}{2\sigma^2}(x-\mu)^2+kx}\mathrm{d}x \\
&= \int_c^\infty \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-1}{2\sigma^2}[[x-(\sigma^2k+\mu)]^2-(\sigma^2k+\mu)^2+\mu^2]}\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-1}{2\sigma^2}[-(\sigma^2k+\mu)^2+\mu^2]}\int_c^\infty e^{\frac{-1}{2\sigma^2}[x-(\sigma^2k+\mu)]^2}\mathrm{d}x \\
&= e^{\frac{\sigma^2k^2}{2}+k\mu}\left[1-\Phi\left(\frac{c-(\sigma^2k+\mu)}{\sigma}\right)\right] \\
\int_c^\infty e^{kx}\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-\mu)^2}{2\sigma^2}}\mathrm{d}x &= e^{k\mu+\frac{k^2\sigma^2}{2}}\Phi\left(\frac{k\sigma^2-c+\mu}{\sigma}\right). \qquad (2.17)
\end{aligned}$$

The computation of the second integral is quite similar. As a result of (Lemma 2.2.1),

$$\begin{aligned}
H_{i,1} &= \int_0^\infty \sum_{k_i=0}^\infty (-1)^{k_i} e^{(y_i-1-k_i)z_i}\frac{1}{\sqrt{2\pi\sum_{p=1}^P x_{i,p}^2\sigma_p^2}}e^{\frac{-(z_i-\sum_{p=1}^P x_{i,p}\mu_p)^2}{2\sum_{p=1}^P x_{i,p}^2\sigma_p^2}}\mathrm{d}z_i \\
&= \frac{1}{\sqrt{2\pi\sum_{p=1}^P x_{i,p}^2\sigma_p^2}}\sum_{k_i=0}^\infty (-1)^{k_i}\int_0^\infty e^{(y_i-1-k_i)z_i}e^{\frac{-(z_i-\sum_{p=1}^P x_{i,p}\mu_p)^2}{2\sum_{p=1}^P x_{i,p}^2\sigma_p^2}}\mathrm{d}z_i \\
H_{i,1} &= \sum_{k_i=0}^\infty (-1)^{k_i} e^{\frac{(y_i-1-k_i)^2\sum_{p=1}^P x_{i,p}^2\sigma_p^2}{2}+(y_i-1-k_i)\sum_{p=1}^P x_{i,p}\mu_p} \\
&\quad \cdot\Phi\left(\frac{(y_i-1-k_i)\sum_{p=1}^P x_{i,p}^2\sigma_p^2+\sum_{p=1}^P x_{i,p}\mu_p}{\sqrt{\sum_{p=1}^P x_{i,p}^2\sigma_p^2}}\right). \qquad (2.18)
\end{aligned}$$

$$
\begin{aligned}
H_{i,2} &= \int_{-\infty}^{0} \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i+k_i)z_i} \frac{1}{\sqrt{2\pi \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}} e^{\frac{-(z_i - \sum_{p=1}^{P} x_{i,p}\mu_p)^2}{2\sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}} \mathrm{d}z_i \\
&= \frac{1}{\sqrt{2\pi \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}} \sum_{k_i=0}^{\infty} (-1)^{k_i} \int_{-\infty}^{0} e^{(y_i+k_i)z_i} e^{\frac{-(z_i - \sum_{p=1}^{P} x_{i,p}\mu_p)^2}{2\sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}} \mathrm{d}z_i \\
H_{i,2} &= \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{\frac{((y_i+k_i))^2 \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}{2} + (y_i+k_i)\sum_{p=1}^{P} x_{i,p}\mu_p} \\
&\quad \cdot \Phi\left(-\frac{(y_i+k_i)\sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2 + \sum_{p=1}^{P} x_{i,p}\mu_p}{\sqrt{\sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}} \right).
\end{aligned}
\tag{2.19}
$$

$H_{i,3} = 0$ as it is an integral against a density on a set of Lebesgue measure zero. As a result, as $P(Y|\Omega) = \prod_{i=1}^{I} H_i$, we estimate our model via maximum likelihood estimation by maximizing $\log P(Y|\Omega)$ which is equivalent to:

$$
\begin{aligned}
\log P(Y|\Omega) &= \log \prod_{i=1}^{I} H_i \\
&= \sum_{i=1}^{I} \log(H_i) \\
&= \sum_{i=1}^{I} \log(H_{i,1} + H_{i,2}),
\end{aligned}
\tag{2.20}
$$

where $H_{i,1}$ and $H_{i,2}$ are defined as above. We state this result as a theorem as it is the main result of this chapter.

Theorem 2.2.2 (Marginalized Logit Likelihood Assuming Independent Normal Prior Distributions)**.** *The log marginalized likelihood of (2.12) assuming independent normal heterogeneity distributions, based on a convergent series approximation to (2.3), is provided by:*

$$\log P(Y|\Omega) = \sum_{i=1}^{I} \log(H_{i,1} + H_{i,2}), \tag{2.21}$$

where:

$$H_{i,1} = \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{\frac{(y_i-1-k_i)^2 \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}{2} + (y_i-1-k_i) \sum_{p=1}^{P} x_{i,p}\mu_p} \cdot\Phi\left(\frac{(y_i-1-k_i)\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2 + \sum_{p=1}^{P} x_{i,p}\mu_p}{\sqrt{\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2}}\right) \tag{2.22}$$

$$H_{i,2} = \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{\frac{((y_i+k_i))^2 \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2}{2} + (y_i+k_i) \sum_{p=1}^{P} x_{i,p}\mu_p} \cdot\Phi\left(-\frac{(y_i+k_i)\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2 + \sum_{p=1}^{P} x_{i,p}\mu_p}{\sqrt{\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2}}\right) \tag{2.23}$$

Theorem 2.2.2 provides the marginalized likelihood, and we can estimate our parameters μ_p and σ_p^2 for $p = 1, ..., P$ by maximizing the above equation and compute associated p-values to determine the statistical significance of the resulting estimates. This marginalization reduces the parameter space from one of IP dimensions to $2P$ dimensions, making model estimation on large data sets considerably easier.

2.3 Simulations

2.3.1 Numerical Simulations

We illustrate the efficacy of our method by performing a series of numerical simulations. In particular, we conducted a series of simulations for $p = 2, 3$, and 4 attributes, allowing $I = 1000$, $JN_i = 1$, and 300 terms in the series expansion (i.e. 1000 households, 1 category, and 1 occasion). These simulations are a subset of a larger number of simulations conducted that is available upon request. For each vector of parameters $(\mu_1,...,\mu_P,...,\sigma_1^2,...,\sigma_P^2)$, we performed 25 simulations. For each such simulate, we simulated $I = 1000$ values of $(x_{i,1},...,x_{i,P})$, rescaled them by dividing by a constant to allow for enough 0/1 variation of y_i, numerically approximated $P(Y|\Omega)$ via (2.20), and maximized the resulting marginal likelihood as a function of Ω.

These simulations were run using MATLAB on an AMD 2.2 GHz Triple Core Processor with 8 GB of RAM. Our results are summarized in Tables 2.1-2.9, consisting of the true values of $(\mu_1,...,\mu_P,...,\sigma_1^2,...,\sigma_P^2)$, the mean and standard deviation of each of these values, and the corresponding t-statistics. The simulations corresponding to $p = 4$ attributes is split up into several tables (Tables 2.5-2.9) due to space constraints. In order to determine whether the values were in numerical correspondence with the true values, we conducted t-tests for each of the parameters. This resulted in comparing the computed t-statistics to a t-distribution with 24 degrees of freedom, as a result of performing 25 simulations for each set of parameter estimates. After instituting conservative but commonly accepted Bonferroni corrections, we used critical values of 3.791 for $p = 2$ attributes, 4.051 for $p = 3$ attributes, and 4.244 for $p = 4$ attributes to compare the calculated t-statistics against. All of our values are well below these critical values. These significance tests therefore do not suggest any meaningful disparity between the estimated parameter values and the true values. As these runs were based on simulated data from a plethora of normal distributions, these simulations demonstrate the strong accuracy of our polynomial approximations as well as the efficacy of our new technique. Additionally, the final few terms of the tails of the truncated series approximations were essentially zero (according to MATLAB output) for a variety of choices of truncation levels. This fact suggested that, at least for these simulations, the truncated series expansions based on 300 terms were reasonable approximations to the marginalized likelihood.

Tab. 2.1: Numerical Simulations for $P = 2$ attributes

(μ_1, μ_2)	(σ_1^2, σ_2^2)	$(\overline{\mu_1}, \overline{\mu_2})$	$(\overline{\sigma_1^2}, \overline{\sigma_2^2})$	$(\sigma_{\mu_1}, \sigma_{\mu_2})$	$(\sigma_{\sigma_1^2}, \sigma_{\sigma_2^2})$
(-8, -3)	(3, 8)	(-7.970, -2.972)	(2.556, 7.658)	(1.581, 1.417)	(1.195, 1.387)
(-4, -14)	(5, 6)	(-3.728, -14.175)	(4.772, 6.511)	(1.188, 1.236)	(1.273, 1.037)
(-7, -5)	(5, 3)	(-7.300, -5.078)	(4.691, 2.710)	(1.434, 1.426)	(1.582, 1.366)
(5, -4)	(4, 5)	(4.865, -3.842)	(4.265, 5.239)	(1.001, 1.210)	(1.432, 1.150)
(-5, 4)	(4, 5)	(-4.801, 4.123)	(4.385, 5.072)	(1.408, 1.318)	(1.252, 1.252)
(5, 4)	(4, 5)	(5.062, 3.893)	(4.429, 5.312)	(1.377, 1.279)	(1.441, 1.086)
(8, -3)	(3, 8)	(8.069, -3.232)	(2.771, 8.290)	(1.742, 1.816)	(0.969, 0.636)
(-8, 3)	(3, 8)	(-7.601, 2.904)	(2.942, 8.389)	(1.587, 1.414)	(0.745, 0.885)
(8, 3)	(3, 8)	(8.182, 2.970)	(2.173, 8.484)	(1.329, 1.362)	(1.387, 1.873)
(4, -14)	(5, 6)	(3.779, -14.461)	(5.259, 5.584)	(1.077, 1.353)	(1.242, 1.303)
(-4, 14)	(5, 6)	(-3.945, 13.877)	(5.167, 5.557)	(0.945, 1.135)	(1.088, 1.012)
(7, -5)	(5, 3)	(7.061, -4.359)	(5.005, 3.416)	(1.253, 1.111)	(1.686, 1.374)
(-7, 5)	(5, 3)	(-7.666, 5.115)	(4.806, 2.929)	(1.036, 1.319)	(1.344, 1.542)
(7, 5)	(5, 3)	(7.327, 4.565)	(4.786, 2.678)	(1.410, 1.240)	(0.858, 1.323)

2.3.2 Comparisons to MCMC Methods

We also assessed the efficacy of our series expansion approach by comparing the speed of the approach to that of existing MCMC methods [50]. For our MCMC baseline comparison, we used the *bayesm* package in R (Rossi and McCullouch 2006) and to ensure a strict "apples-to-apples" comparison of computing times, we also translated our series expansion approach from MATLAB to R. We ran the

Tab. 2.2: t-statistics for $P = 2$ attributes

(μ_1, μ_2)	(σ_1^2, σ_2^2)	t-stat (μ_1)	t-stat (μ_2)	t-stat (σ_1^2)	t-stat (σ_2^2)
(-8, -3)	(3, 8)	0.095	0.098	1.859	1.232
(-4, -14)	(5, 6)	1.147	-0.708	0.897	-2.461
(-7, -5)	(5, 3)	-1.047	-0.272	0.976	1.063
(5, -4)	(4, 5)	-0.672	0.652	-0.924	-1.040
(-5, 4)	(4, 5)	0.706	0.466	-1.538	-0.289
(5, 4)	(4, 5)	0.224	-0.418	-1.490	-1.437
(8, -3)	(3, 8)	0.197	-0.640	1.180	-2.275
(-8, 3)	(3, 8)	1.257	-0.339	0.392	-2.201
(8, 3)	(3, 8)	-0.683	0.111	2.981	-1.291
(4, -14)	(5, 6)	-1.025	-1.704	-1.042	1.597
(-4, 14)	(5, 6)	0.292	-0.542	-0.765	2.187
(7, -5)	(5, 3)	0.242	2.887	-0.013	-1.514
(-7, 5)	(5, 3)	-3.212	0.435	0.723	0.230
(7, 5)	(5, 3)	1.159	-1.756	1.246	1.216

Tab. 2.3: Numerical Simulations for $P = 3$ attributes

(μ_1, μ_2, μ_3)	(σ_1^2, σ_2^2, σ_3^2)	($\overline{\mu_1}$, $\overline{\mu_2}$, $\overline{\mu_3}$)	($\overline{\sigma_1^2}$, $\overline{\sigma_2^2}$, $\overline{\sigma_3^2}$)	(σ_{μ_1}, σ_{μ_2}, σ_{μ_3})	(σ_{σ_1}, σ_{σ_2}, σ_{σ_3})
(-5, -6, -7)	(3, 4, 3)	(-5.010, -6.298, -6.883)	(3.268, 3.973, 3.232)	(1.591, 1.453, 1.704)	(1.037, 1.091, 0.941)
(5, -6, -7)	(3, 4, 3)	(5.030, -5.690, -7.658)	(3.028, 4.098, 2.679)	(1.788, 1.897, 1.382)	(1.397, 1.128, 1.357)
(-5, 6, -7)	(3, 4, 3)	(-5.283, 6.043, -6.691)	(2.685, 3.947, 3.189)	(1.159, 1.369, 1.617)	(1.288, 0.748, 0.923)
(-5, -6, 7)	(3, 4, 3)	(-4.803, -5.962, 7.247)	(2.930, 4.096, 2.885)	(1.242, 1.541, 1.396)	(1.250, 0.882, 1.139)
(5, -6, 7)	(3, 4, 3)	(4.745, -5.889, 7.093)	(3.130, 3.925, 3.263)	(1.295, 1.600, 1.501)	(1.054, 1.038, 0.766)
(5, 6, 7)	(3, 4, 4)	(4.964, 6.290, 6.987)	(2.972,4.339, 4.123)	(1.542, 1.423, 1.593)	(0.970, 1.617, 1.397)
(-15, -4, -6)	(2, 7, 4)	(-15.347, -3.751, -6.367)	(2.116, 7.619, 3.472)	(2.045, 1.894, 2.162)	(1.009, 2.977, 2.961)
(15, -4, -6)	(2, 7, 4)	(14.352, -3.913, -5.930)	(2.005, 7.058, 4.202)	(1.708, 1.568, 1.579)	(0.825, 2.255, 1.830)
(-9, -8, 4)	(3, 3, 5)	(-9.203, -7.712, 3.560)	(3.024, 3.637, 4.689)	(1.567, 1.737, 1.919)	(1.011, 1.888, 1.167)
(-15, -4, 6)	(2, 7, 4)	(-14.975, -3.882, 5.610)	(1.888, 6.995, 4.364)	(1.890, 1.528, 1.790)	(0.981, 2.478, 2.290)
(15, -4, 6)	(2, 7, 4)	(14.601, -3.780, 6.152)	(1.761, 6.503, 4.535)	(2.174, 1.919, 1.853)	(0.940, 2.714, 2.688)
(15, 4, 6)	(2, 7, 4)	(14.818, 4.131, 5.782)	(2.169, 6.344, 4.371)	(1.891, 2.127, 1.729)	(0.986, 2.706, 2.787)
(-9, -8, -4)	(3, 3, 5)	(-9.039, -8.245, -3.967)	(2.721, 2.975, 5.219)	(1.641, 1.504, 1.533)	(1.989, 2.321, 2.394)
(-9, 8, -4)	(3, 3, 5)	(-9.112, 8.124, -4.107)	(3.191, 3.364, 4.949)	(1.205, 1.686, 1.434)	(1.022, 0.762, 0.598)
(-9, -8, 4)	(3, 3, 5)	(-9.203, -7.712, 3.560)	(3.024, 3.637, 4.689)	(1.567, 1.737, 1.919)	(1.011, 1.888, 1.167)
(9, -8, 4)	(3, 3, 5)	(9.699, -7.759, 3.298)	(3.104, 2.810, 4.954)	(1.591, 1.361, 1.421)	(0.557, 0.782, 0.557)
(-9, 8, 4)	(3, 3, 5)	(-9.176, 7.820, 4.318)	(2.949, 3.260, 5.068)	(1.440, 1.403, 1.431)	(1.157, 1.054, 0.993)
(9, 8, 4)	(3, 3, 5)	(9.394, 8.074, 3.907)	(2.709, 2.310, 4.487)	(1.693, 1.531, 1.629)	(2.403, 2.344, 2.304)

Tab. 2.4: t-statistics for $P = 3$ attributes

(μ_1, μ_2, μ_3)	(σ_1^2, σ_2^2, σ_3^2)	t-stat (μ_1)	t-stat (μ_2)	t-stat (μ_3)	t-stat (σ_1^2)	t-stat (σ_2^2)	t-stat (σ_3^2)
(-5, -6, -7)	(3, 4, 3)	-0.030	-1.024	0.342	-1.293	0.125	-1.231
(5, -6, -7)	(3, 4, 3)	0.083	0.818	-2.382	-0.101	-0.433	1.185
(-5, 6, -7)	(3, 4, 3)	-1.221	0.156	0.955	1.222	0.352	-1.023
(-5, -6, 7)	(3, 4, 3)	0.791	0.123	0.883	0.279	-0.544	0.507
(5, -6, 7)	(3, 4, 3)	-0.984	0.347	0.310	-0.614	0.364	-1.717
(5, 6, 7)	(3, 4, 4)	0.118	-1.016	0.041	0.147	-1.048	-0.442
(-15, -4, -6)	(2, 7, 4)	-0.848	0.657	-0.849	-0.572	-1.040	0.892
(15, -4, -6)	(2, 7, 4)	-1.897	0.279	0.223	-0.031	-0.129	-0.552
(-9, -8, 4)	(3, 3, 5)	0.648	-0.829	1.146	0.121	1.686	-1.333
(-15, -4, 6)	(2, 7, 4)	0.066	0.385	-1.089	0.570	0.010	-0.794
(15, -4, 6)	(2, 7, 4)	-0.917	0.573	0.409	1.273	0.916	-0.996
(15, 4, 6)	(2, 7, 4)	-0.481	0.307	-0.630	-0.856	1.213	-0.666
(-9, -8, -4)	(3, 3, 5)	0.118	0.814	-0.107	-0.702	-0.054	0.456
(-9, 8, -4)	(3, 3, 5)	0.465	-0.369	0.372	0.932	2.392	-0.423
(-9, -8, 4)	(3, 3, 5)	0.648	-0.829	1.146	0.121	1.686	-1.333
(9, -8, 4)	(3, 3, 5)	-2.197	-0.887	2.472	0.936	-1.215	-0.411
(-9, 8, 4)	(3, 3, 5)	0.611	0.640	-1.111	-0.222	1.235	0.342
(9, 8, 4)	(3, 3, 5)	-1.163	-0.242	0.285	-0.605	-1.472	-1.113

Tab. 2.5: Numerical Simulations for $P = 4$ attributes

$(\mu_1,\ \mu_2,\ \mu_3,\ \mu_4)$	$(\sigma_1^2,\ \sigma_2^2,\ \sigma_3^2,\ \sigma_4^2)$
(-5, -14, -3, -2)	(3,6,4,6)
(-10,-12,-5,-5)	(3,3,3,4)
(-7,-5,-5,-3)	(4,6,5,4)
(5,-14,-3,-5)	(3,6,4,6)
(-5,14,-3,-5)	(3,6,4,6)
(-5,-14,-3,5)	(3,6,4,6)
(5,14,-3,5)	(3,6,4,6)
(5,-14,3,-5)	(3,6,4,6)
(5,14,3,5)	(3,6,4,6)
(10,-12,-5,-5)	(3,4,3,4)
(-10,12,-5,-5)	(3,4,3,4)
(-10,-12,5,-5)	(3,4,3,4)
(-10,-12,-5,5)	(3,4,3,4)
(10,-12,5,-5)	(3,4,3,4)
(-10,12,-5,5)	(3,4,3,4)
(10,12,5,5)	(3,4,3,4)
(7,-5,-5,-3)	(4,6,5,4)
(-7,5,-5,-3)	(4,6,5,4)
(-7,-5,5,-3)	(4,6,5,4)
(7,-5,5,-3)	(4,6,5,4)
(-7,5,-5,3)	(4,6,5,4)
(7,5,5,3)	(4,6,5,4)

Tab. 2.6: Numerical Simulations for $P = 4$ attributes (continued)

$(\overline{\mu_1}, \overline{\mu_2}, \overline{\mu_3}, \overline{\mu_4})$	$(\overline{\sigma_1^2}, \overline{\sigma_2^2}, \overline{\sigma_3^2}, \overline{\sigma_4^2})$
(-4.957, -13.727, -3.254, -2.323)	(2.368, 6.014, 3.736, 5.074)
(-9.651, -11.951, -5.249, -4.794)	(2.753, 3.035, 2.806, 3.984)
(-7.269, -4.924, -5.028, -3.163)	(3.523, 5.720, 4.367, 3.839)
(5.088, -14.096, -3.556, -4.840)	(2.437, 5.938, 4.067, 6.112)
(-5.618, 14.514, -2.732, -4.960)	(2.832, 6.048, 3.760, 5.845)
(-4.707, -14.606, -2.770, 5.084)	(2.694, 5.589, 4.116, 6.014)
(-5.120, 14.078, -3.168, 5.304)	(3.410, 5.987, 4.348, 5.793)
(4.987, -14.345, 3.140, -4.963)	(2.412, 5.604, 3.876, 5.818)
(5.062, 14.542, 2.360, 5.343)	(2.597, 5.537, 3.947, 5.476)
(9.624, -11.746, -5.127, -5.109)	(2.659, 4.061, 3.088, 3.973)
(-10.161, 12.526, -4.934, -5.120)	(3.124, 4.026, 3.069, 4.073)
(-10.167, -12.214, 4.796, -4.899)	(2.685, 3.774, 3.066, 3.636)
(-9.792, -11.717, -5.133, 4.834)	(3.243, 4.206, 3.034, 3.243)
(10.076, -11.867, 5.134, -4.972)	(2.969, 4.024, 2.934, 4.088)
(-9.833, 12.462, -5.085, 4.123)	(2.964, 3.948, 3.008, 3.934)
(10.432, 11.902, 5.244, 4.756)	(3.054, 4.280, 2.398, 3.640)
(7.045, -5.497, -4.710, -3.204)	(4.347, 6.055, 4.975, 3.883)
(-7.097, 5.174, -5.122, -2.976)	(3.875, 5.943, 4.816, 3.799)
(-6.715, -5.053, 5.164, -3.247)	(4.037, 6.064, 4.898, 4.066)
(7.361, -4.920, 4.834, -2.846)	(3.907, 5.966, 5.073, 3.997)
(-7.234, 4.965, -5.234, 3.456)	(4.026, 5.923, 5.091, 4.172)
(6.766, 5.419, 5.249, 3.109)	(3.757, 5.210, 4.864, 4.174)

Tab. 2.7: Numerical Simulations for $P = 4$ attributes (continued)

$(\sigma_{\mu_1}, \sigma_{\mu_2}, \sigma_{\mu_3}, \sigma_{\mu_4})$	$(\sigma_{\sigma_1^2}, \sigma_{\sigma_2^2}, \sigma_{\sigma_3^2}, \sigma_{\sigma_4^2})$
(1.956, 1.762, 1.672, 1.712)	(1.231, 1.878, 1.525, 1.599)
(1.938, 1.898, 1.712, 1.827)	(2.301, 2.506, 2.298, 2.501)
(1.652, 1.479, 1.743, 1.574)	(1.845, 2.094, 2.099, 2.007)
(1.595, 1.482, 1.471, 1.601)	(1.593, 0.967, 0.737, 1.153)
(1.277, 1.576, 1.337, 1.517)	(1.365, 0.934, 0.749, 0.565)
(1.382, 1.612, 1.570, 1.670)	(0.914, 1.146, 0.939, 0.492)
(1.707, 1.542, 1.451, 1.287)	(1.502, 0.906, 1.061, 0.875)
(1.440, 1.811, 1.365, 1.349)	(1.132, 0.789, 0.786, 0.724)
(1.380, 1.726, 1.839, 1.850)	(1.822, 2.091, 2.155, 2.075)
(1.372, 1.680, 1.645, 1.379)	(1.313, 0.848, 1.171, 0.761)
(1.588, 1.542, 1.434, 1.666)	(0.740, 0.549, 0.887, 0.660)
(1.805, 1.776, 1.723, 1.866)	(1.928, 2.108, 2.050, 2.163)
(1.798, 1.769, 1.545, 1.692)	(1.936, 1.835, 1.826, 1.678)
(1.656, 1.866, 1.496, 1.623)	(0.502, 0.479, 0.235, 0.294)
(1.564, 1.680, 1.755, 1.518)	(0.780, 0.575, 0.763, 0.775)
(1.753, 1.815, 1.588, 1.806)	(2.352, 2.435, 2.215, 2.432)
(1.138, 1.449, 1.528, 1.654)	(1.340, 0.880, 1.124, 1.040)
(1.519, 1.793, 1.945, 1.562)	(1.009, 0.880, 0.857, 0.854)
(1.681, 1.850, 1.458, 1.728)	(0.469, 0.898, 0.277, 0.689)
(1.294, 1.450, 1.624, 1.433)	(0.657, 0.267, 0.215, 0.528)
(1.376, 1.826, 1.736, 1.493)	(1.185, 0.630, 0.966, 0.527)
(1.590, 1.671, 1.557, 1.679)	(1.903, 2.122, 2.165, 2.243)

Tab. 2.8: Numerical Simulations for $P = 4$ attributes (continued)

t-stat (μ_1)	t-stat (μ_2)	t-stat (μ_3)	t-stat (μ_4)
0.111	0.774	-0.758	-0.942
0.899	0.129	-0.727	0.563
-0.813	0.257	-0.079	-0.519
0.276	-0.325	-1.891	0.500
-2.419	1.631	1.003	0.132
1.060	-1.881	0.731	0.250
-0.351	0.254	-0.580	1.180
-0.045	-0.951	0.513	0.138
0.225	1.570	-1.740	0.928
-1.370	0.756	-0.387	-0.396
-0.508	1.705	0.231	-0.360
-0.463	-0.603	-0.591	0.272
0.580	0.799	-0.431	-0.491
0.230	0.356	0.449	0.085
0.534	1.375	-0.241	-2.889
1.233	-0.269	0.768	-0.675
0.199	-1.713	0.949	-0.616
-0.320	0.486	-0.313	0.078
0.848	-0.144	0.562	-0.715
1.394	0.274	-0.512	0.538
-0.849	-0.097	-0.674	1.526
-0.737	1.255	0.801	0.324

Tab. 2.9: Numerical Simulations for $P = 4$ attributes (continued)

t-stat (σ_1^2)	t-stat (σ_2^2)	t-stat (σ_3^2)	t-stat (σ_4^2)
2.568	-0.038	0.866	2.897
0.537	-0.070	0.421	0.032
1.294	0.668	1.508	0.400
1.768	0.324	-0.452	-0.485
0.615	-0.255	1.606	1.372
1.675	1.792	-0.616	-0.147
-1.366	0.070	-1.639	1.183
2.597	2.513	0.790	1.255
1.107	1.106	0.124	1.262
1.300	-0.359	-0.376	0.175
-0.841	-0.236	-0.388	-0.555
0.817	0.537	-0.160	0.841
-0.627	-0.560	-0.094	2.255
0.306	-0.248	1.405	-1.490
0.229	0.455	-0.054	0.425
-0.115	-0.576	1.360	0.741
-1.294	-0.310	0.112	0.563
0.622	0.326	1.072	1.177
-0.392	-0.354	1.841	-0.481
0.708	0.628	-1.689	0.028
-0.108	0.608	-0.471	-1.629
0.638	1.860	0.314	-0.388

MCMC sampler for 20,000 iterations, after allowing for 5,000 burn-in iterations. In particular, we simulated data and estimated our model on this simulated data for $I = 1000$ individuals, letting $p = 2, 3$, or 4 attributes, and $JN_i = 1$. We truncated our series expansions after 300 terms, which was shown to be quite sufficient in the previous section. The computing times in minutes for each method is presented in Table 2.10.

Tab. 2.10: Comparisons of Computing Time for closed-form Series Expansion Approach versus MCMC Methods (in minutes)

	Series Expansions	MCMC Methods
$p = 2$ attributes	2.20	69.20
$p = 3$ attributes	11.29	69.60
$p = 4$ attributes	10.98	69.30

As Table 2.10 illustrates, our new technique clearly outperforms existing MCMC methods. Autocorrelations of draws from the posterior suggested that the number of iterations performed (20,000 along with 5,000 for burn-in) was necessary in order to begin having a sense of the entire posterior distribution, although running the sampler for more iterations (and hence for a longer amount of time) would be advisable in practice. Thus, it is quite clear that our non-iterative series expansion approach outperforms existing MCMC methods in computing time. These computational gains can be quite substantial for large data sets.

2.4 A Few Generalizations

2.4.1 Incorporating covariances: Positing Heterogeneity from a Multivariate Normal Distribution

Previously, we had assumed that our parameters $\beta_{i,1}, ..., \beta_{i,p}$ were drawn from independent normal distributions. It is possible, however, to generalize this assumption and allow for correlations amongst the different response coefficients. In some applications, allowing for correlations is an important assumption to make; for example, in modeling consumer choice of products, a consumer's price sensitivity may be related to sensitivities to other marketing-related covariates, such as the presence of a promotion or even the time of year.

Still retaining our initial assumption that all households are independent, we can allow $\beta_i = (\beta_{i,1}, ..., \beta_{i,p})$ to be drawn from a multivariate normal distribution with mean μ and variance-covariance matrix Σ:

$$
\begin{aligned}
P(Y|\beta) &= \prod_{i=1}^{I} \frac{e^{y_i X_i^T \beta_i}}{1 + e^{X_i^T \beta_i}}, \\
\text{where } (\beta_{i,1}, ...\beta_{i,p}) &\sim MVN(\vec{\mu}, \Sigma).
\end{aligned}
\tag{2.24}
$$

The diagonal elements of Σ represent variances $\sigma_1^2, ...\sigma_p^2$ and the off-diagonal elements $\rho_{m,n}$ for $m \neq n$ allow us to model correlations amongst the various coefficients. The result is a more general integral to evaluate. As we did earlier, we can define $z_i = \sum_{p=1}^{P} x_{i,p}\beta_{i,p}$. Now that we have covariances in our model, however, it follows that $z_i \sim N(\sum_{p=1}^{P} x_{i,p}\mu_p, \sum_{p=1}^{P} x_{i,p}^2\sigma_p^2 + \sum_{m \neq n, m,n \leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n)$.

Therefore, if we define P_{z_i} as the measure induced by z_i on measurable space (T_i, G_i) having density with respect to Lebesgue measure:

$$f(z_i) = \frac{1}{\sqrt{2\pi \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2 + \sum_{m \neq n, m,n \leq P} x_{i,m} x_{i,n} \rho_{m,n} \sigma_m \sigma_n}} e^{\frac{-(z_i - \sum_{p=1}^{P} x_{i,p}\beta_{i,p})^2}{2\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2 + \sum_{m\neq n, m,n \leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}},$$

then we can again integrate H_i as defined in (2.20):

$$\begin{aligned} H_i &= \int_{-\infty}^{\infty} \left[\sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i - 1 - k_i) z_i} I(z_i > 0) \right. \\ &\quad \left. + \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i + k_i) z_i} I(z_i < 0) + \frac{e^{y_i z_i}}{1 + e^{z_i}} I(z_i = 0) \right] f(z_i) \mathrm{d} z_i. \end{aligned}$$

We can again decompose H_i into a sum of three integrals, $H_{i,1}$, $H_{i,2}$, and $H_{i,3}$ where $H_i = H_{i,1} + H_{i,2} + H_{i,3}$. And once again, as a result of Lemma 2.2.1,

$$\begin{aligned} H_{i,1} &= \int_0^{\infty} \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i - 1 - k_i) z_i} \frac{e^{\frac{-(z_i - \sum_{p=1}^{P} x_{i,p}\beta_{i,p})^2}{2\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2 + \sum_{m\neq n, m,n \leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}}}{\sqrt{2\pi \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2 + \sum_{m \neq n, m,n \leq P} x_{i,m} x_{i,n} \rho_{m,n} \sigma_m \sigma_n}} \mathrm{d} z_i \\ &= \frac{1}{\sqrt{2\pi \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2 + \sum_{m \neq n, m,n \leq P} x_{i,m} x_{i,n} \rho_{m,n} \sigma_m \sigma_n}} \sum_{k_i=0}^{\infty} (-1)^{k_i} \int_0^{\infty} e^{(y_i - 1 - k_i) z_i} \\ &\quad \cdot e^{\frac{-(z_i - \sum_{p=1}^{P} x_{i,p}\beta_{i,p})^2}{2\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2 + \sum_{m\neq n, m,n \leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}} \mathrm{d} z_i \\ H_{i,1} &= \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{\frac{(y_i - 1 - k_i)^2 (\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2 + \sum_{m\neq n, m,n \leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n)}{2} + (y_i - 1 - k_i) \sum_{p=1}^{P} x_{i,p}\mu_p} \\ &\quad \cdot \Phi\left(\frac{(y_i - 1 - k_i)(\sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2 + \sum_{m \neq n, m,n \leq P} x_{i,m} x_{i,n} \rho_{m,n} \sigma_m \sigma_n) + \sum_{p=1}^{P} x_{i,p}\mu_p}{\sqrt{\sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2 + \sum_{m \neq n, m,n \leq P} x_{i,m} x_{i,n} \rho_{m,n} \sigma_m \sigma_n}} \right). \end{aligned} \tag{2.25}$$

$$
\begin{aligned}
H_{i,2} &= \int_{-\infty}^{0} \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i+k_i)z_i} \frac{e^{\frac{-(z_i-\sum_{p=1}^{P} x_{i,p}\beta_{i,p})^2}{2\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}}}{\sqrt{2\pi\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}} \mathrm{d}z_i \\
&= \frac{1}{\sqrt{2\pi\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}} \sum_{k_i=0}^{\infty} (-1)^{k_i} \int_{-\infty}^{0} e^{(y_i+k_i)z_i} \\
&\quad \cdot e^{\frac{-(z_i-\sum_{p=1}^{P} x_{i,p}\beta_{i,p})^2}{2\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}} \\
H_{i,2} &= \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{\frac{(y_i+k_i)^2(\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n)}{2}+(y_i+k_i)\sum_{p=1}^{P} x_{i,p}\mu_p} \\
&\quad \cdot\Phi\left(-\frac{(y_i+k_i)(\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n)+\sum_{p=1}^{P} x_{i,p}\mu_p}{\sqrt{\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}}\right).
\end{aligned} \tag{2.26}
$$

$H_{i,3} = 0$ as it is once again an integral against a density on a set of Lebesgue measure zero. As a result, as $P(Y|\Omega) = \prod_{i=1}^{I} H_i$, we can again estimate our model, now incorporating correlations amongst the coefficients, via the method of marginal maximum likelihood by maximizing $\log P(Y|\Omega)$ which is equivalent to:

$$
\begin{aligned}
\log P(Y|\Omega) &= \log \prod_{i=1}^{I} H_i \\
&= \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{N_i} \log(H_i) \\
&= \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{N_i} \log(H_{i,1} + H_{i,2}).
\end{aligned} \tag{2.27}
$$

The parameter space is now considerably larger given the need to estimate the off-diagonal coefficients $\rho_{m,n}$ but can still be optimized over using standard optimization routines.

2.4.2 *Allowing dependence on other factors*

Some researchers may want to allow the mean of β_i to depend on certain covariates $(Z_{i,1}, \ldots, Z_{i,k})$ that may represent K demographic factors for each particular individual [141]. Under such a model, we would have:

$$P(Y|\beta) = \prod_{i=1}^{I} \frac{e^{y_i X_i^T \beta_i}}{1 + e^{X_i^T \beta_i}},$$

$$\text{where } (\beta_{i,1}, ...\beta_{i,p}) \sim MVN(\vec{\Delta}, \Sigma), \tag{2.28}$$

where $\vec{\Delta} = (\sum_{k=1}^{K} Z_{i,k}\mu_{1,k}, ..., \sum_{k=1}^{K} Z_{i,k}\mu_{P,k})$ and once again, the diagonal elements of Σ represent variances $\sigma_1^2, ...\sigma_p^2$ and the off-diagonal elements $\rho_{m,n}$ for $m \neq n$ allow us to model correlations amongst the various coefficients. The density for $z_i = \sum_{p=1}^{P} x_{i,p}\beta_{i,p}$ with which we integrate our series expansions with respect to is now a bit different:

$$z_i \sim N(\sum_{p=1}^{P} x_{i,p} \sum_{k=1}^{K} Z_{i,k}\mu_{p,k}, \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2 + \sum_{m \neq n, m,n \leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n). \tag{2.29}$$

Therefore, we can now redefine P_{z_i} as the measure induced by z_i on measurable space (T_i, G_i) having density with respect to Lebesgue measure,

$$f(z_i) = \frac{1}{\sqrt{2\pi \sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2 + \sum_{m \neq n, m,n \leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}} e^{\frac{-(z_i - \sum_{p=1}^{P} x_{i,p} \sum_{k=1}^{K} Z_{i,k}\mu_{p,k})^2}{2\sum_{p=1}^{P} x_{i,p}^2 \sigma_p^2 + \sum_{m \neq n, m,n \leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}}$$

then again:

$$
\begin{aligned}
H_i &= \int_{-\infty}^{\infty}\left[\sum_{k_i=0}^{\infty}(-1)^{k_i}e^{(y_i-1-k_i)z_i}I(z_i>0)\right.\\
&\quad\left.+\sum_{k_i=0}^{\infty}(-1)^{k_i}e^{(y_i+k_i)z_i}I(z_i<0)+\frac{e^{y_iz_i}}{1+e^{z_i}}I(z_i=0)\right]f(z_i)\mathrm{d}z_i.
\end{aligned}
$$

Once again decomposing H_i into a sum of three integrals, $H_{i,1}$, $H_{i,2}$, and $H_{i,3}$ where $H_i = H_{i,1} + H_{i,2} + H_{i,3}$ and utilizing Lemma 2.2.1,

$$
\begin{aligned}
H_{i,1} &= \int_0^{\infty}\sum_{k_i=0}^{\infty}(-1)^{k_i}e^{(y_i-1-k_i)z_i}\frac{e^{\frac{-(z_i-\sum_{p=1}^P x_{i,p}\sum_{k=1}^K Z_{i,k}\mu_{p,k})^2}{2\sum_{p=1}^P x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P}x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}}}{\sqrt{2\pi\sum_{p=1}^P x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P}x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}}\mathrm{d}z_i\\
&= \frac{1}{\sqrt{2\pi\sum_{p=1}^P x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P}x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}}\sum_{k_i=0}^{\infty}(-1)^{k_i}\int_0^{\infty}e^{(y_i-1-k_i)z_i}\\
&\quad\cdot e^{\frac{-(z_i-\sum_{p=1}^P x_{i,p}\sum_{k=1}^K Z_{i,k}\mu_{p,k})^2}{2\sum_{p=1}^P x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P}x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}}\mathrm{d}z_i\\
H_{i,1} &= \sum_{k_i=0}^{\infty}(-1)^{k_i}e^{\frac{(y_i-1-k_i)^2(\sum_{p=1}^P x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P}x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n)}{2}+(y_i-1-k_i)\sum_{p=1}^P x_{i,p}\sum_{k=1}^K Z_{i,k}\mu_{p,k}}\\
&\quad\cdot\Phi\left(\frac{(y_i-1-k_i)(\sum_{p=1}^P x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P}x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n)+\sum_{p=1}^P x_{i,p}\sum_{k=1}^K Z_{i,k}\mu_{p,k}}{\sqrt{\sum_{p=1}^P x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P}x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}}\right).
\end{aligned}
\tag{2.30}
$$

$$
\begin{aligned}
H_{i,2} &= \int_{-\infty}^{0} \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i+k_i)z_i} \frac{e^{\frac{-(z_i-\sum_{p=1}^{P} x_{i,p}\sum_{k=1}^{K} Z_{i,k}\mu_{p,k})^2}{2\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}}}{\sqrt{2\pi\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}} \mathrm{d}z_i \\
&= \frac{1}{\sqrt{2\pi\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}} \sum_{k_i=0}^{\infty}(-1)^{k_i}\int_{-\infty}^{0} e^{(y_i+k_i)z_i} \\
&\quad \cdot e^{\frac{-(z_i-\sum_{p=1}^{P} x_{i,p}\sum_{k=1}^{K} Z_{i,k}\mu_{p,k})^2}{2\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}} \\
H_{i,2} &= \sum_{k_i=0}^{\infty}(-1)^{k_i} e^{\frac{(y_i+k_i)^2(\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n)}{2}+(y_i+k_i)\sum_{p=1}^{P} x_{i,p}\sum_{k=1}^{K} Z_{i,k}\mu_{p,k}} \\
&\quad \cdot\Phi\left(-\frac{(y_i+k_i)(\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n)+\sum_{p=1}^{P} x_{i,p}\sum_{k=1}^{K} Z_{i,k}\mu_{p,k}}{\sqrt{\sum_{p=1}^{P} x_{i,p}^2\sigma_p^2+\sum_{m\neq n,m,n\leq P} x_{i,m}x_{i,n}\rho_{m,n}\sigma_m\sigma_n}}\right).
\end{aligned}
\tag{2.31}
$$

As before, $H_{i,3} = 0$. We utilize these equations to maximize (2.20), which again now has a slightly larger parameter space as a result of $\beta_{i,p}$'s dependence on $Z_{i,k}$.

2.4.3 Arbitrary Priors

There is no reason to solely restrict ourselves to normal prior distributions. If we look at the pertinent integration, we can nicely generalize our result provided we have an analytic expression for the probability density function representing $z_i = \sum_{p=1}^{P} x_{i,p}\beta_{i,p}$:

$$
\begin{aligned}
H_i &= \int_{-\infty}^{\infty}\Bigg[\sum_{k_i=0}^{\infty}(-1)^{k_i} e^{(y_i-1-k_i)z_i} I(z_i>0) \\
&\quad + \sum_{k_i=0}^{\infty}(-1)^{k_i} e^{(y_i+k_i)z_i} I(z_i<0) + \frac{e^{y_i z_i}}{1+e^{z_i}} I(z_i=0)\Bigg] f(z_i)\mathrm{d}z_i.
\end{aligned}
$$

It is easy to notice that this is simply just a linear combination of incomplete moment generating functions corresponding to $f(z_i)$. Thus provided that a closed-form density for $f(z_i)$ exists:

$$H_i \quad = \quad \sum_{k_i=0}^{\infty}(-1)^{k_i} IMGF_1(y_i - 1 - k_i) + \sum_{k_i=0}^{\infty}(-1)^{k_i} IMGF_2(y_i + k_i), \tag{2.32}$$

where $IMGF_1$ and $IMGF_2$ are the incomplete moment generating functions corresponding to $f(z_i)$, integrated from 0 to ∞ and from $-\infty$ to 0 respectively. For an arbitrary prior, we can use these incomplete moment generating functions to rewrite our marginalized likelihood and maximize the function accordingly, provided that a closed-form density corresponding to $f(z_i)$ exists.

2.5 Application to Analyzing Potential Causes of Child Poverty

Understanding the potential causes of child poverty is an important issue in public policy research. With trillions of dollars being thrown at welfare programs for decades, it is paramount for policy researchers to analyze potential determinants of child poverty to evaluate the efficacy of such programs [131]. We do so in this Chapter.

2.5.1 Data

We used 2009 Current Population Survey (CPS) Data, compiled by the United States Census Bureau [156]. The CPS data is used for a variety of purposes including for providing the United States Federal Government's monthly jobs report. From the CPS Data, we extracted children under the age of 18 and information of whether they were below the poverty level (1 if they were below the poverty level and 0 otherwise), education of the head of the household (hereafter referred to as head parent), marital status of the parents, the head parent's age, whether the head parent was working full time, and how many people were living in the household. The resulting data set consisted of slightly more than 60,000 observations.

In existing policy research, these covariates have been considered significant factors in determining whether a child grows up in poverty [63, 130, 134, 145]. Although logistic regression has been used to analyze poverty data [150], no policy research, to our knowledge, has looked at this question using models that allow for individual-level heterogeneity. Allowing for heterogeneity is important as there is no a priori reason to believe that some individuals would not respond differently to these factors than others. Furthermore, ignoring heterogeneity can result in the researcher ignoring potential variability in the model and can therefore lead to misleading statistical inferences.

We therefore used the closed-form expansion derived in the previous sections to estimate a Bayesian binary logistic regression model with child poverty as a binary categorical dependent variable and all the other variables mentioned above as explanatory variables. In order to simplify our model estimation, we assumed no correlations amongst our coefficients.

Data was rescaled and coefficients were multiplied again by a constant to ensure for sufficient terms in the series expansion to allow for reasonable approximations. Due to the rapid convergence of the series expansions associated with this data set indicated by looking the final few terms of the truncated series approximations for a variety of choices of truncated terms, only 50 terms in our series expansion were necessary.

2.5.2 Estimation Results

Our estimation results are depicted in Table 2.11. The use of our closed-form polynomial expansion approach reduced the dimensionality of our problem from approximately 350,000 parameters to simply just 12 parameters. P-values were determined by performing likelihood ratio tests and utilizing Wilk's Theorem:

Tab. 2.11: Empirical Bayesian Logistic Regression Estimation Results on Child Poverty Data

Parameter	(μ_p)	p-value	(σ_p^2)	p-value
Intercept	0.955	< 0.001	0.000025	< 0.001
Parents Married	-2.006	< 0.001	2.902	< 0.001
Head Parent College Educated	-1.889	< 0.001	1.869	< 0.001
Head Parent Working Full Time	-1.711	< 0.001	2.577	< 0.001
Head Parent Age	-0.030	< 0.030	0.000025	< 0.001
Number of Children in the Household	0.345	< 0.001	0.380	< 0.001

Our results shed light on potential factors influencing child poverty. All of our coefficient estimates are highly significant. In particular, of the covariates looked at, marital status of the parents is an influential predictor of child poverty, as well as the educational level of the head parent, and whether the head parent is working full time. Parental age and the number of children living in the household are also significant factors influencing whether a child lives in poverty. Our variance estimates also indicate there is a substantial amount of heterogeneity in the population in how these factors combine to influence child poverty.

The significance of our explanatory variables coincides with common sense - Two working parents have the potential to bring in more income to a household than simply just one parent. Additionally, more educated parents have more job opportunities and therefore have greater potential to comfortably support a family than less educated parents. Older parents are typically more mature and understand the challenges associated with having children and often wait until they are ready to do so. Similarly, more responsible parents will also often wait until they are financially ready to have additional children.

2.5.3 *Policy Implications*

Our results are in line with research produced by both the Heritage Foundation and the Brookings Institution [130, 132, 145]. Both think tanks have argued that marriage is a powerful antidote to child poverty. Recent research, for example, has argued that marriage in the American society has declined in recent years while out of wedlock births have steadily increased [130]. In 1964, for example, 93% of children were born to married parents while in 2007 only 59% of children were born to married parents. On the other hand, in the mid 1960s, less than 10% of children were born out of wedlock, while in 2007, this number skyrocketed to 40.6%. As our results here indicate, children born out of wedlock are overwhelmingly far more likely to live in poverty than children born to married parents [130, 145].

These results have a number of important policy implications. Our results, along with work from both Heritage and Brookings, suggest that increasing marriage can significantly reduce child poverty [134]. In order to do so, state and local policymakers could consider establishing a campaign of public advertising and education on teenage abstinence as well as the consequences of child bearing outside of marriage. These campaigns could also work toward communicating the practical issues faced by single parents, the importance of delaying having children until one is older and more mature, as well as the importance of waiting until one finds a suitable partner before doing so [67]. These campaigns could set normative expectations for younger generations, encouraging young people to become well-educated, to delay having children until marriage, to work full time to support any children they have, and to limit their family size to what they can afford [145]. Just as policymakers have established anti-smoking, anti-drinking, and staying in school campaigns, the impact of such "pro-marriage campaigns" could be quite significant [130].

Additionally, our results also suggest that welfare programs could benefit from significant reform. Currently, many means-tested welfare programs such as food stamps, public housing, and Temporary Assistance to Needy Families (TANF) are structured in a manner that disincentivizes marriage. In particular, many of these programs have penalties for marriage because welfare benefits decline as a family's income rises. Thus, for many low-income mothers, marriage signifies a decline in governmental assistance and consequently an overall reduction in the couple's combined income. These problems with the welfare system can be ameliorated by reforming the Earned Income Tax Credit for married couples with children to counteract the anti-marriage penalties associated with welfare programs [130].

Furthermore, policymakers could also consider strict work and/or study requirements for able-bodied welfare recipients. Such "workfare" requirements not only have the potential to increase personal income but also encourage only those truly in need to apply for welfare [129]. Additionally, "workfare" helps enable welfare recipients to acquire valuable training that could be useful in finding subsequent full time work. In the 1990s, for example, AFDC programs were fundamentally restructured into TANF around these ideas, and the result was a rise in employment as well as a marked decline in child poverty. There are many other American welfare programs that could benefit from similar reforms [135]. The right reforms could transform many of these programs from the broken safety nets that they currently are into trampolines that foster growth and success.

2.6 Conclusions and Future Research

With millions of Americans remaining in poverty despite the federal government's endless spending on welfare programs, it is important that policymakers understand the fundamental causes of poverty. Our study has looked at this issue and offers a number of informative suggestions. In particular, our results suggest that policies that work toward educating young people about marriage and having a family could have a significant capacity to reduce child poverty. Additionally, work requirements for able-bodied welfare recipients could also be particularly helpful.

Methodologically, we now also have an approach for obtaining closed-form Bayesian inferences for the binary logit model utilizing polynomial expansions, allowing for the use of rich two-sided normal prior distribution. These series expansions can be made arbitrarily close depending on how many terms the researcher chooses to use in the truncated approximations. Our simulations demonstrate the efficacy of our technique as well as the fact that our method outperforms existing MCMC methods. The speed of our approach provides an attractive alternative to MCMC methods, particularly for large data sets, such as the child poverty data set used here. Our analysis of this child poverty data set suggests that marriage, parental age, parental education, and parental work status are significant factors influencing child poverty. Our findings are in line with extant policy research, and we suggest a number of policy implications and suggestions based on our results coupled with these studies.

Although we have focused our application of this model to public policy research, there is no reason that this model cannot be applied to other fields where logistic regression is used including marketing, economics, biostatistics, criminology, and professional sports among others. Methodologically, there are also many potential avenues of future research that this study should encourage. For example, we made the assumption in this study that $JN_i = 1$. Future research should look into weakening this restriction. Additionally, a potential avenue of future research is to explore other methods of polynomial expansions and compare them to the approach using geometric series expansions here. Furthermore, although we primarily concentrated on one particular model - the binary logit model, and one class of priors, the normal distribution, we hope this study spurs research on closed-form Bayesian inferences for other models as well. In particular, a nice aspect about members of the exponential family is that each member has a certain conjugate prior. It could be useful from a computational perspective to use polynomial expansions to approximate posterior distributions within this family for a choice of priors previously considered non-conjugate. Additionally, the binary logit model discussed here belongs to a larger class of generalized linear models (GLMs). A potential avenue of future research could be to utilize polynomial expansions to allow researchers to make closed-form Bayesian inferences based on other GLMs. Additionally, deriving a polynomial expansion approach for the multinomial logistic regression model, a workhorse model in applied economics research, would also be a worthy endeavor of future research [58, 99].

It is always useful to have a variety of methods to draw inferences from statistical models, especially for large data sets. We hope the polynomial expansion approach presented here adds significant value to the applied statistician's toolbox.

Chapter 3: Closed-form Bayesian Inferences for Semiparametric Density Ratio Modeling with Applications to Tort Reform

3.1 Introduction

3.1.1 Tort Reform

Medical doctors belong to one of the most highly-respected professions in America. Yet they are at risk of facing unnecessary lawsuits everyday. From surgeons to obstetricians to general practitioners, virtually all doctors fear the risk of lawsuit abuse. These risks have been considered by policymakers on both sides of the aisle to be an important component in reforming our nation's health care system [73].

One aspect of the American health care system that is worthy of attention is medical malpractice reform. The potential for fraud and abuse of the medical malpractice system unnecessarily raises health care costs by impacting physician supply and forcing doctors to engage in defensive medicine [19, 85, 151]. Medical malpractice reform falls into a larger class of reforms of the civil justice system, known as tort reform. According to *Black's Law Dictionary*, a tort is defined as "... a legal wrong committed upon the person or property independent of contract ..." [46]. In settling these disputes, compensation may be awarded to the victims.

America has always been a highly litigious country with many frivolous lawsuits and unnecessary abuses of the civil justice system. Such a climate has been shown to impose hidden costs on consumers as well as on business, including, but not limited to, the health care, automotive, agricultural, and retail sectors of the American economy [103]. Tort reform seeks to to reduce tort costs by fundamentally reforming the civil justice system to prevent abuse, thereby reducing unnecessary litigation. Recent research has illustrated that tort reform can significantly improve the business climate in America leading to more jobs, better health care, and a more prosperous economy [104].

Policymakers have sought to pursue tort reform in a number of ways. One approach has been to impose monetary caps, which limit the amount of money that a jury may award a plaintiff. These caps may apply to appeal bonds, non-economic damages, punitive damages, or monetary damage awards. Other approaches have been to direct reforms toward other aspects of the civil justice system such as class action lawsuits, imposing statutes of limitations, and requiring attorney fee limitations among others.

In this chapter, we quantify the impact of medical malpractice reforms by estimating the probabilities of extreme tort losses. Computation of these probabilities, however, is not an elementary statistical problem as different states exhibit different degrees of litigiousness. Alaska and Texas have been shown, for example, to be considerably less litigious than other states such as New York and California [104]. Additionally, different localities within states may also exhibit a certain degree of heterogeneity. For example, New York City and various towns in upstate New York may differ how litigious they are. In order to capture this type of state-level heterogeneity in our model, we improve on existing semiparametric density ratio estimation methodologies. A series of numerical simulations, along with goodness of fit computations, illustrates the efficacy of our approach.

3.1.2 Bayesian Parametric methods

With the rapid improvements in statistical computing power over the course of the last three decades, incorporating heterogeneity in parametric models has become increasingly common in the statistical literature. These days, researchers can choose from a variety of methodologies such as a parametric Bayesian approach, a non-parametric Bayesian approach, a finite mixture approach, or a combination of the finite mixture modeling coupled with Bayesian modeling [74, 90, 92, 140].

Modeling individual-level heterogeneity is often important as there is generally no a priori reason to assume that all individuals (or observations) in a data set behave in an identical manner. However, as data sets often contain limited information about each individual, it is quite difficult, if not impossible, to estimate models incorporating heterogeneity from a frequentist perspective. The Bayesian approach allows the statistician to assume individual-level parameters adhere to a lower dimensional probability distribution and perform statistical inferences based on the parameters of these lower dimensional distributions either via an empirical Bayesian approach or a fully Bayesian approach [48, 110]. By reducing the problem's dimensionality, estimation of these statistical models becomes quite feasible.

In this chapter, we adapt such a Bayesian approach to semiparametric methodology used thus far only for frequentist statistical modeling. Our approach enables these models to accommodate individual-level heterogeneity with high-dimensional parameter spaces. We discuss this semiparametric methodology in the following section.

3.1.3 Density Ratio Estimation

Density ratio estimation (DRE) methods were first suggested over thirty years ago. In their 1997 paper, following Prentice and Pyke (1979) and others, Qin and Zhang suggested instead of making the typically strict parametric assumptions about distributions of data sets, that a researcher only make assumptions about the ratios of probability densities (known as tilts) based on subsamples within the data sets. Qin and Zhang (2005) assumed exponential tilts and recommended estimation of these models via the method of empirical likelihood [116, 122, 124, 125].

Over the years, DRE has had many applications in statistical research. For example, Gilbert et al (1999) improved on DRE's methodologies and applied these improvements to understanding the efficacy of HIV vaccine trials. Several years later, Kedem et al (2008) applied DRE to time series forecasting [52, 81]. In particular, their study used DRE to develop distributional assumptions necessary to forecast mortality rates for various age groups within the United States. Kedem et al (2009) and Voulgaraki et al (2012) applied DRE to cancer research [80,160]. Both studies were able to utilize DRE to determine the significance of certain risk factors for cancer. In addition to these studies, there have been many other studies that have utilized the semiparametric benefits of the DRE approach [43, 44, 80, 119, 122].

Fokianos and Qin (2008) employed importance sampling in connection with DRE, which required the generation of artificial data [45]. Prior to that, however, all research using DRE was based on within-sample data. Specifically, data was always divided into smaller subsets (such as cohorts) and comparisons would be made between these subsets. Recent research proposed a innovative adaptation of density ratio estimation known as "out of sample fusion." Based on the idea of having one primary data set as a reference, and a secondary artificial (potentially simulated) data set, this research illustrated that more accurate inferences can be made about the primary data set by applying DRE to both samples [78, 82, 168]. Earlier in this dissertation, we illustrated that that this methodology can be particularly useful for making Bayesian inferences when applied to posterior samples.

To date, however, although a few studies have applied Bayesian methods to empirical likelihood problems, no studies have done so uniting the semiparametric

DRE method with Bayesian methods to model individual-level heterogeneity [91, 106, 146, 165]. We ameliorate this limitation in this chapter. In particular, we adapt the Bayesian approach mentioned in the previous section to the semiparametric DRE method to model individual-level heterogeneity. We apply this methodology to a data set used in a tort reform study conducted by the Pacific Research Institute to understand overall distributional properties of tort losses throughout the country [19].

3.2 Problem Formulation

Suppose that we have a data set consisting of $i = 1, ..., I$ samples and define $j = 1, ..., n_i$ observations within each i^{th} sample, such that we have the following P-dimensional vectors $\boldsymbol{x}_{i,j} = (x_{i,j,1}, x_{i,j,2}, ..., x_{i,j,P})$ with $\sum_{i=1}^{I} n_i = N$. Let $M = I + 1$ and define probability density functions g_i such that:

$$\boldsymbol{x}_{i,j} \sim g_i. \tag{3.1}$$

Additionally, we can also define $g_{I+1} \equiv g$ as our reference probability density, describing another sample of size n_{I+1}. Assume the densities g_i satisfy the following relationship regarding their ratios:

$$\frac{g_i(\boldsymbol{x_{i,j}})}{g(\boldsymbol{x_{i,j}})} \quad = \quad w(\boldsymbol{\theta_i}, \boldsymbol{x}_{i,j}), \tag{3.2}$$

where $\boldsymbol{\theta_i}$ is a vector-valued statistical parameter to be estimated. Without loss of generality, we assume exponential tilts, defining $w(\boldsymbol{\theta_i}, \boldsymbol{x}_{i,j}) \equiv e^{\alpha_i + \boldsymbol{\beta_i' h(x_{i,j})}}$, where we are currently assuming $\boldsymbol{h} : \mathbf{R}^P \to \mathbf{R}^P$. We allow $\alpha_i \sim N(\mu_\alpha, 1)$ and $\boldsymbol{\beta_i} \sim N(\boldsymbol{\mu_\beta}, \boldsymbol{\Sigma_\beta})$

where $\boldsymbol{\beta_i'} = (\beta_{i,1}, ..., \beta_{i,P})$ be our heterogeneity distributions.[1] By allowing our model's coefficients to vary for each sample, we enable our model to capture "sample-level" (or individual-level if each sample represents an individual) heterogeneity.[2] We begin by assuming that $\boldsymbol{\Sigma_\beta}$ is a diagonal matrix and hence that the random variables $\beta_{i,p}$ are statistically independent of each other. Additionally, we make the assumption that $n_i = 1 \;\forall\; i = 1, \ldots, I$ (to assume a single observation for every individual in a data set) and $n_{I+1} = n_M = N$.

3.2.1 Bayesian Density Ratio Estimation

Let $G(\boldsymbol{x}) = GI + 1(\boldsymbol{x})$ be the reference CDF and define $p_{ij} = dG(\boldsymbol{x_{i,j}}) = dG_{I+1}(\boldsymbol{x_{i,j}})$. We can utilize the method of constrained empirical likelihood and estimate g_i and $\boldsymbol{\theta_i}$ as follows. We can write the empirical likelihood function, based on our pooled data $\boldsymbol{x_{ij}}$:

$$L(\boldsymbol{\theta}, G_M) = \prod_{i=1}^{M} \prod_{j=1}^{n_i} p_{ij} \prod_{i=1}^{I} \prod_{j=1}^{n_i} e^{\alpha_i + \boldsymbol{\beta_i' h(x_{i,j})}}, \tag{3.3}$$

where $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_I, \beta_{1,1}, \ldots, \beta_{I,P})$.[3] As stated above, we make the assumption that $n_i = 1 \;\forall\; i = 1, \ldots, I$ and $n_{I+1} = n_M = N = I$. As a result, the second product in the exponential terms does not contribute to multiplication of the distortion function. We marginalize the empirical likelihood function to generate a "marginalized empirical likelihood" by integrating the above empirical likelihood against the heterogeneity distributions:

[1] For example, one potential choice for $\boldsymbol{h}$ is $\boldsymbol{h(x) = x}$ as in Voulgaraki et al (2012) [160]. This definition can be generalized, however, and the dimension of $\boldsymbol{\beta_i'}$ would consequently also need to be altered.

[2] We assume the above parameterization for α_i (with a constant variance) to ensure statistical identifiability of the model after marginalization.

[3] Chapter 4 contains some theoretical discussion regarding empirical likelihood estimation involving density ratios.

$$
\begin{aligned}
\mathrm{ML}(\mu_\alpha, \boldsymbol{\mu_\beta}, \boldsymbol{\Sigma_\beta}, G_M) &= \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} L(\boldsymbol{\theta}, G_M) \prod_{i=1}^{I} \frac{1}{\sqrt{2\pi}} e^{(\frac{-(\alpha_i-\mu_\alpha)^2}{2})} d\alpha_i \\
&\cdot \prod_{p=1}^{P} \frac{1}{\sqrt{2\pi\sigma^2_{\beta_p}}} e^{(\frac{-(\beta_{i,p}-\mu_{\beta_p})^2}{2\sigma^2_{\beta_p}})} d\beta_{i,p} \\
&= \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \prod_{i=1}^{M}\prod_{j=1}^{n_i} p_{ij} \prod_{i=1}^{I}\prod_{j=1}^{n_i} e^{\alpha_i + \boldsymbol{\beta_i' h(x_{i,j})}} \frac{1}{\sqrt{2\pi}} e^{(\frac{-(\alpha_i-\mu_\alpha)^2}{2})} d\alpha_i \\
&\cdot \prod_{p=1}^{P} \frac{1}{\sqrt{2\pi\sigma^2_{\beta_p}}} e^{(\frac{-(\beta_{i,p}-\mu_{\beta_p})^2}{2\sigma^2_{\beta_p}})} d\beta_{i,p} \\
&= \prod_{i=1}^{M}\prod_{j=1}^{n_i} p_{ij} \prod_{i=1}^{I}\prod_{j=1}^{n_i} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{\alpha_i + \boldsymbol{\beta_i' h(x_{i,j})}} \frac{1}{\sqrt{2\pi}} e^{(\frac{-(\alpha_i-\mu_\alpha)^2}{2})} d\alpha_i \\
&\cdot \prod_{p=1}^{P} \frac{1}{\sqrt{2\pi\sigma^2_{\beta_p}}} e^{(\frac{-(\beta_{i,p}-\mu_{\beta_p})^2}{2\sigma^2_{\beta_p}})} d\beta_{i,p} \\
\mathrm{ML}(\mu_\alpha, \boldsymbol{\mu_\beta}, \boldsymbol{\Sigma_\beta}, G_M) &= \prod_{i=1}^{M}\prod_{j=1}^{n_i} p_{ij} \prod_{i=1}^{I}\prod_{j=1}^{n_i} e^{\mu_\alpha + \frac{1}{2}} e^{\boldsymbol{\mu_\beta}' \boldsymbol{h(x_{i,j})} + \frac{1}{2}\boldsymbol{h(x_{i,j})' \Sigma_\beta h(x_{i,j})}} \qquad (3.4)
\end{aligned}
$$

As a result, we have the following theorem:

Theorem 3.2.1 (Marginalized Empirical Likelihood for Density Ratio Model Assuming Normal Prior Distributions)**.** *The marginalized log-likelihood function of (3.3), assuming normal heterogeneity distributions, is provided by:*

$$
\begin{aligned}
LL(\mu_\alpha, \boldsymbol{\mu_\beta}, \boldsymbol{\Sigma_\beta}, G_M) &= \log ML(\mu_\alpha, \boldsymbol{\mu_\beta}, \boldsymbol{\Sigma_\beta}, G_M) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{n_i} \log p_{ij} + \sum_{i=1}^{I}\sum_{j=1}^{n_i} (\mu_\alpha + \frac{1}{2} + \boldsymbol{\mu_\beta}' \boldsymbol{h(x_{i,j})} + \frac{1}{2}\boldsymbol{h(x_{i,j})' \Sigma_\beta h(x_{i,j})}).
\end{aligned}
$$

This result is simply due to taking the logarithm of (3.4).

One maximizes the above marginalized likelihood subject to constraints analogous to those used in Voulgaraki et al 2012 [160]: $p_{ij} \geq 0$, $\sum_{i=1}^{M}\sum_{j=1}^{n_i} p_{ij} = 1$, and $\sum_{i=1}^{M}\sum_{j=1}^{n_i} p_{ij} e^{\mu_\alpha + \frac{1}{2} + \boldsymbol{\mu_\beta}' \boldsymbol{h(x_{i,j})} + \frac{1}{2}\boldsymbol{h(x_{i,j})' \Sigma_\beta h(x_{i,j})}} = 1$.[4] One can perform the opti-

[4] This constraint is easy to see after integration of both sides of the constraint imposed in

mization of the empirical likelihood function numerically to estimate μ_α, $\boldsymbol{\mu_\beta}$, and $\boldsymbol{\Sigma_\beta}$.

3.2.2 Advantages of Marginalization

The marginalization of the empirical likelihood function in equation (3.3) serves a few important purposes. Firstly, in data sets where n_i is small, each observation will contain very limited information regarding each α_i and $\boldsymbol{\beta_i}$. As a result, marginalizing the empirical likelihood function by integrating over this high-dimensional parameter space, enables us to markedly reduce the dimensionality of the problem, to a considerably less intricate model involving just μ_α, $\boldsymbol{\mu_\beta}$, and $\boldsymbol{\Sigma_\beta}$.

Note that this reduction in dimensionality occurs because the marginalization essentially transforms our model from one density ratio into another. In particular, after starting with I different samples, each of size 1, using a sample of size $N = I$ as a reference, and integrating over the parameter space, the density ratio becomes another exponential with a slightly different functional form. As a result, this new density ratio compares two different distributions, each of sample size N.

As the regularity conditions outlined in Fokianos (2004) clearly hold for our marginalized empirical likelihood, our empirical likelihood estimators are statistically unbiased and asymptotically normal [43]. A researcher can test the null hypothesis that each coefficient is equal to zero against the alternative that it is nonzero. As illustrated in Kedem et al (2009), the likelihood ratio of each null model to the full model asymptotically follows a chi-squared distribution [80].

3.2.3 Derivation of Distributions

We can use the results from our optimization to estimate the distributions. In particular, we can define $\gamma \equiv \lambda/2N$, where λ is a Lagrange multiplier. We can

Voulgaraki et al (2012): $\sum_{i=1}^{M}\sum_{j=1}^{n_i} p_{ij} e^{\alpha_k + \boldsymbol{\beta_k'} \boldsymbol{h}(\boldsymbol{x}_{i,j})} = 1$ over the heterogeneity distributions F, providing us with $\int \sum_{i=1}^{M}\sum_{j=1}^{n_i} p_{ij} e^{\alpha_k + \boldsymbol{\beta_k'} \boldsymbol{h}(\boldsymbol{x}_{i,j})}\, dF = \int 1\, dF \; \forall\, k = 1, \ldots, I$.

subsequently replace $\gamma, \mu_\alpha, \boldsymbol{\mu_\beta}$, and $\boldsymbol{\Sigma_\beta}$ by their estimators. As a result, following the derivations in Voulgaraki (2011) [159], estimators of $\hat{p}_{ij}$ and $\hat{G}(\boldsymbol{x})$ are provided by:

$$\hat{p}_{ij} = \frac{1}{2N} \frac{1}{1 + \hat{\gamma}[e^{\hat{\mu}_\alpha + \frac{1}{2}} e^{\boldsymbol{\hat{\mu}'_\beta h(x_{i,j})} + \frac{1}{2}\boldsymbol{h(x_{i,j})'\hat{\Sigma}_\beta h(x_{i,j})}} - 1]} \tag{3.5}$$

and:

$$\begin{aligned} \hat{G}(\boldsymbol{x}) &= \sum_{i=1}^{M}\sum_{j=1}^{n_i} \hat{p}_{ij} I(\boldsymbol{x}_{ij} \leq \boldsymbol{x}) \\ &= \frac{1}{2N}\sum_{i=1}^{M}\sum_{j=1}^{n_i} \frac{I(\boldsymbol{x}_{ij} \leq \boldsymbol{x})}{1 + \hat{\gamma}[e^{\hat{\mu}_\alpha + \frac{1}{2}} e^{\boldsymbol{\hat{\mu}'_\beta h(x_{i,j})} + \frac{1}{2}\boldsymbol{h(x_{i,j})'\hat{\Sigma}_\beta h(x_{i,j})}} - 1]}. \end{aligned} \tag{3.6}$$

Furthermore, for the "marginalized distribution," which we will hereafter refer to as $\hat{H}$, we have:

$$\begin{aligned} \hat{H}(\boldsymbol{x}) &= \sum_{i=1}^{M}\sum_{j=1}^{n_i} \hat{p}_{ij} e^{\hat{\mu}_\alpha + \frac{1}{2} + \boldsymbol{\hat{\mu}'_\beta h(x_{i,j})} + \frac{1}{2}\boldsymbol{h(x_{i,j})'\hat{\Sigma}_\beta h(x_{i,j})}} I(\boldsymbol{x}_{ij} \leq \boldsymbol{x}) \\ &= \frac{1}{2N}\sum_{i=1}^{M}\sum_{j=1}^{n_i} \frac{e^{\hat{\mu}_\alpha + \frac{1}{2} + \boldsymbol{\hat{\mu}'_\beta h(x_{i,j})} + \frac{1}{2}\boldsymbol{h(x_{i,j})'\hat{\Sigma}_\beta h(x_{i,j})}} I(\boldsymbol{x}_{ij} \leq \boldsymbol{x})}{1 + \hat{\gamma}[e^{\hat{\mu}_\alpha + \frac{1}{2}} e^{\boldsymbol{\hat{\mu}'_\beta h(x_{i,j})} + \frac{1}{2}\boldsymbol{h(x_{i,j})'\hat{\Sigma}_\beta h(x_{i,j})}} - 1]}. \end{aligned} \tag{3.7}$$

For researchers interested in estimating the probability density function of the sample, kernel density estimators can be constructed by smoothing the increments of $\hat{H}$ [43, 125, 160]. This research in fact illustrated that one can arrive at more efficient kernel density estimates as a result of combining data. As mentioned earlier in this dissertation, optimal bandwidth selection for the kernel density estimation is discussed in detail in Voulgaraki et al (2012) [160].

3.3 Numerical Simulations

To demonstrate the efficacy of our approach, we simulated data sets of size 100 from a variety of distributions (gamma, Weibull, and exponential). The gamma and Weibull distributions were parametrized with shape parameter α_i and scale parameter β_i while the exponential distribution was parameterized by α_i. These parameters were drawn from normal distributions, with $\alpha_i \sim N(\mu_\alpha, \sigma_\beta^2)$ and $\beta_i \sim N(\mu_\beta, \sigma_\beta^2)$. This stochastic nature of our distributions' parameters enabled us to simulate datasets that, as described earlier, are "heterogeneous" in nature.

Assuming the density ratio to follow (3.2) having linear exponential tilts $w(\alpha_i, \beta_i; x_i) = e^{\alpha_i + \beta_i x_i}$ and only one observation per sample (i.e. $n_i = 1 \ \forall \ i$), we fused this simulated data with a random draw of equal sample size from a specified distribution that served as a reference distribution. Specifically, we took our sample of size 100 and fused it with another sample (a reference distribution) of size 100 from a probability distribution (gamma, uniform, or normal) chosen a priori. When fusing with a gamma distribution, we fit our original sample to a gamma distribution, estimated the gamma distribution's parameters via standard maximum likelihood techniques, and used a sample of size 100 from this gamma distribution as a reference. When fusing with a uniform distribution, we fused our distribution with a sample over the range of our original sample's data and used a sample of size 100 from this uniform distribution as a reference. Finally, when fusing with a normal distribution, we fused our simulated data with a sample of size 100 from a normal distribution with sample mean and sample variance equal to that of the original sample. We then estimated the distribution of both samples from the combined data using the method of constrained empirical likelihood.

Given the limited information per sample in the simulated data set (as $n_i = 1$), the marginalization of the empirical likelihood discussed in Section 3.2 enables a significant reduction of the dimensionality of the model's parameter space (from 200 to 3) and is consequently quite useful for understanding distributional properties of the data. To quantify our model's ability to understand these distributional properties, we used a diagnostic statistic recommended in Voulgaraki et al (2012) [160]:

$$R^2_{\alpha,k} = 1 - \exp\left[-\left(\frac{x_\alpha}{m - x_\alpha}\right)^k\right]. \tag{3.8}$$

In the above equation, m is defined as the number of times the estimated semiparametric cumulative distribution function falls inside the estimated $1-\alpha$ confidence interval obtained from the corresponding empirical cumulative distribution function, both functions being evaluated at the sample points. Additionally, k is a prespecified constant, set by the statistician. Since we fused our data with a reference sample from a known distribution, we estimated (3.8) on our reference distribution. If our semiparametric density ratio model were inappropriate, then our estimates would ruin the integrity of this reference distribution, resulting in inaccurate estimates and consequently poor goodness of fit results. We estimated (3.8) for our simulations over choices of $k = 1$ and $k = 2$.

Following the notation pertaining to (3.2), we refer to our simulated dataset as $\boldsymbol{x_1} = (x_{1,1}, x_{2,1}, \ldots, x_{i,1}, \ldots, x_{I,1})$ (that is, $n_i = 1$ and $p = 1\ \forall\ i = 1, \ldots, I$ and $p = 1, \ldots, P$) and the sample with which we are fusing as $\boldsymbol{x_2} = (x_{1,2}, x_{2,2}, \ldots, x_{i,2}, \ldots, x_{I,2})$. In the following tables, $\hat{\alpha}_{MLE}$ and $\hat{\beta}_{MLE}$ denote the maximum likelihood estimators of $\boldsymbol{x_1}$ when the data are fused with a gamma distribution, $x_{(1,1)}$ and $x_{(I,1)}$ denote the minimum and maximum values of $\boldsymbol{x_1}$ respectively when the data are fused with a uniform distribution, and $\bar{\boldsymbol{x}}_{\mathbf{1}}$ and $s^2_{\boldsymbol{x}_{\mathbf{1}}}$ denote the mean and variance of $\boldsymbol{x}_{\mathbf{1}}$ when the data are fused with a normal distribution.

Our results are depicted in Tables 3.1-3.3. We notice that our Bayesian approach clearly fits better than, if not as well as, the existing DRE method as R^2 under our standard DRE approach is always less than or equal to R^2 under our new Bayesian DRE approach. Additionally, we also notice that the choice of $\boldsymbol{x}_{\mathbf{2}}$ with which we fuse our sample with, also affects the goodness of fit results. For example, at least in these simulations, we notice that fusing our data with a uniform distribution results in identical results when comparing the DRE approach to the Bayesian DRE approach. In some cases, such as for our exponentially distributed samples, we notice an identical goodness of fit (and hence the standard DRE approach is just as effective); however, for other cases, such as Weibull samples, fusing with a uniform distribution results in quite poor goodness of fit. These results suggest that when using out of sample fusion to make statistical inferences about a particular data set, the researcher should choose from a variety of different distributions from which to fuse her data with and compare goodness of fit diagnostics to optimally determine the distribution's best estimate.

These results demonstrate a number of important points. Firstly, the density ratios before the marginalization were slightly misspecified to begin with. For example, the proper ratio in equation (3.2) for a gamma distribution (fused with another gamma or uniform distribution) should be of the form $w(\alpha_i, \beta_{i,1}, \beta_{i,2}; x_{i,1}) = e^{\alpha_i + \beta_{i,1} x_{i,1} + \beta_{i,2} \log(x_{i,1})}$ whereas we used a density ratio have functional form $e^{\alpha_i + \beta_i x_i}$ instead. The proper ratios for the other distributions examined are also slightly different from what was examined here. In reality, however, when presented with a data set, a researcher generally does not have any detailed specifications regarding the distribution of the data. Making the simple specification as we have here, with an elementary linear tilt function and the marginalization of the resulting empirical

Tab. 3.1: First Simulation Set - Out of Sample Fusion with Gamma Distribution, Sample Size $I = 100$

						DRE		Bayesian DRE	
$x_{i,1}$	μ_α	σ^2_α	μ_β	σ^2_β	$x_{i,2}$	$R^2_{0.95,1}$	$R^2_{0.95,2}$	$R^2_{0.95,1}$	$R^2_{0.95,2}$
$\Gamma(\alpha_i, \beta_i)$	10	4	10	4	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.957	0.976	0.964	0.982
$\Gamma(\alpha_i, \beta_i)$	30	25	10	25	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.886	0.917	0.902	0.929
$\Gamma(\alpha_i, \beta_i)$	10	4	30	4	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.980	0.994	0.980	0.994
$\Gamma(\alpha_i, \beta_i)$	100	100	30	100	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.934	0.975	0.950	0.979
$\Gamma(\alpha_i, \beta_i)$	30	25	100	25	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.993	1.000	0.993	1.000
Weibull(α_i, β_i)	10	4	10	4	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.829	0.871	0.847	0.882
Weibull(α_i, β_i)	30	25	10	25	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.837	0.876	0.856	0.893
Weibull(α_i, β_i)	10	4	30	4	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.582	0.552	0.598	0.576
Weibull(α_i, β_i)	100	100	30	100	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.637	0.624	0.639	0.626
Weibull(α_i, β_i)	30	25	100	25	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.475	0.368	0.502	0.410
Exponential(α_i)	10	4	NA	NA	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.952	0.974	0.952	0.974
Exponential(α_i)	30	25	NA	NA	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.944	0.965	0.944	0.965
Exponential(α_i)	10	4	NA	NA	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.948	0.972	0.948	0.972
Exponential(α_i)	100	100	NA	NA	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.936	0.971	0.936	0.971
Exponential(α_i)	30	25	NA	NA	$\Gamma(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	0.965	0.987	0.965	0.987

Tab. 3.2: Second Simulation Set - Out of Sample Fusion with Normal Distribution, Sample Size $I = 100$

						DRE		Bayesian DRE	
$x_{i,1}$	μ_α	σ^2_α	μ_β	σ^2_β	$x_{i,2}$	$R^2_{0.95,1}$	$R^2_{0.95,2}$	$R^2_{0.95,1}$	$R^2_{0.95,2}$
$\Gamma(\alpha_i, \beta_i)$	10	4	10	4	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.870	0.905	0.876	0.906
$\Gamma(\alpha_i, \beta_i)$	30	25	10	25	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.730	0.738	0.743	0.754
$\Gamma(\alpha_i, \beta_i)$	10	4	30	4	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.952	0.984	0.955	0.987
$\Gamma(\alpha_i, \beta_i)$	100	100	30	100	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.808	0.827	0.822	0.844
$\Gamma(\alpha_i, \beta_i)$	30	25	100	25	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.988	0.999	0.988	0.999
Weibull(α_i, β_i)	10	4	10	4	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.881	0.922	0.897	0.936
Weibull(α_i, β_i)	30	25	10	25	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.898	0.947	0.922	0.966
Weibull(α_i, β_i)	10	4	30	4	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.589	0.550	0.604	0.571
Weibull(α_i, β_i)	100	100	30	100	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.664	0.666	0.704	0.723
Weibull(α_i, β_i)	30	25	100	25	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.511	0.431	0.514	0.439
Exponential(α_i)	10	4	NA	NA	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.782	0.824	0.786	0.830
Exponential(α_i)	30	25	NA	NA	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.853	0.913	0.853	0.913
Exponential(α_i)	10	4	NA	NA	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.797	0.850	0.801	0.855
Exponential(α_i)	100	100	NA	NA	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.859	0.910	0.859	0.910
Exponential(α_i)	30	25	NA	NA	$\mathrm{N}(\bar{\boldsymbol{x}}_{\mathbf{1}}, s^2_{\boldsymbol{x}_{\mathbf{1}}})$	0.832	0.891	0.832	0.891

Tab. 3.3: Third Simulation Set - Out of Sample Fusion with Uniform Distribution, Sample Size $I = 100$

						DRE		Bayesian DRE	
$x_{i,1}$	μ_α	σ^2_α	μ_β	σ^2_β	$x_{i,2}$	$R^2_{0.95,1}$	$R^2_{0.95,2}$	$R^2_{0.95,1}$	$R^2_{0.95,2}$
$\Gamma(\alpha_i, \beta_i)$	10	4	10	4	Unif$(x_{(1)}, x_{(N)})$	0.960	0.980	0.960	0.980
$\Gamma(\alpha_i, \beta_i)$	30	25	10	25	Unif$(x_{(1)}, x_{(N)})$	0.704	0.699	0.704	0.699
$\Gamma(\alpha_i, \beta_i)$	10	4	30	4	Unif$(x_{(1)}, x_{(N)})$	1.000	1.000	1.000	1.000
$\Gamma(\alpha_i, \beta_i)$	100	100	30	100	Unif$(x_{(1)}, x_{(N)})$	0.791	0.826	0.791	0.826
$\Gamma(\alpha_i, \beta_i)$	30	25	100	25	Unif$(x_{(1)}, x_{(N)})$	1.000	1.000	1.000	1.000
Weibull(α_i, β_i)	10	4	10	4	Unif$(x_{(1)}, x_{(N)})$	0.308	0.177	0.308	0.177
Weibull(α_i, β_i)	30	25	10	25	Unif$(x_{(1)}, x_{(N)})$	0.303	0.170	0.303	0.170
Weibull(α_i, β_i)	10	4	30	4	Unif$(x_{(1)}, x_{(N)})$	0.144	0.037	0.144	0.037
Weibull(α_i, β_i)	100	100	30	100	Unif$(x_{(1)}, x_{(N)})$	0.143	0.032	0.143	0.032
Weibull(α_i, β_i)	30	25	100	25	Unif$(x_{(1)}, x_{(N)})$	0.107	0.020	0.107	0.020
Exponential(α_i)	10	4	NA	NA	Unif$(x_{(1)}, x_{(N)})$	1.000	1.000	1.000	1.000
Exponential(α_i)	30	25	NA	NA	Unif$(x_{(1)}, x_{(N)})$	1.000	1.000	1.000	1.000
Exponential(α_i)	10	4	NA	NA	Unif$(x_{(1)}, x_{(N)})$	1.000	1.000	1.000	1.000
Exponential(α_i)	100	100	NA	NA	Unif$(x_{(1)}, x_{(N)})$	0.981	0.999	0.981	0.999
Exponential(α_i)	30	25	NA	NA	Unif$(x_{(1)}, x_{(N)})$	1.000	1.000	1.000	1.000

likelihood function, still enables us to reasonably estimate our distributions' data, as our goodness of fit results illustrate. The fact that misspecified tilt functions can still lead to reasonable distributional estimates has been illustrated in Katzoff et al (2013) [78].

In the following section, we discuss a few nice extensions of our theoretical results.

3.4 A Few Generalizations

3.4.1 Arbitrary Prior Distributions and Tilt Functions

Although we assumed normal heterogeneity distributions, there is no a priori reason to restrict ourselves to this parametric family. In particular, if we generalize our assumptions and posit an arbitrary prior distribution and tilt function, we notice that that the marginalized empirical likelihood function simply just involves the moment generating function of the prior measure. Mathematically, we can quite easily derive this generalization again under the assumption that $n_i = 1 \; \forall \; i = 1, \ldots, I$ and $n_{I+1} = n_M = N$. Following the density ratio in (3.2), and supposing that α_i follows a distribution with density $P(\alpha_i|\boldsymbol{\theta_\alpha})$ with respect to Lebesgue measure and that each $\beta_{i,p}$ follows a distribution, statistically independent of α_i and of each other, with density $P(\beta_{i,p}|\boldsymbol{\theta_{\beta_p}})$, we can generalize our marginalized empirical likelihood as follows:

$$\begin{aligned}
ML(\boldsymbol{\theta_\alpha}, \boldsymbol{\theta_{\beta_1}}, \ldots, \boldsymbol{\theta_{\beta_P}}, G_M) &= \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} L(\boldsymbol{\theta}, G_M) \prod_{i=1}^{I} P(\alpha_i|\boldsymbol{\theta_\alpha}) d\alpha_i \prod_{p=1}^{P} P(\beta_{i,p}|\boldsymbol{\theta_{\beta_p}}) d\beta_{i,p} \\
&= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{i=1}^{M} \prod_{j=1}^{n_i} p_{ij} \prod_{i=1}^{I} \prod_{j=1}^{n_i} e^{\alpha_i + \boldsymbol{\beta_i' h(x_{i,j})}} P(\alpha_i|\boldsymbol{\theta_\alpha}) d\alpha_i \\
&\quad \cdot \prod_{p=1}^{P} P(\beta_{i,p}|\boldsymbol{\theta_{\beta_p}}) d\beta_{i,p} \\
&= \prod_{i=1}^{M} \prod_{j=1}^{n_i} p_{ij} \prod_{i=1}^{I} \prod_{j=1}^{n_i} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{\alpha_i + \boldsymbol{\beta_i' h(x_{i,j})}} P(\alpha_i|\boldsymbol{\theta_\alpha}) d\alpha_i \\
&\quad \cdot \prod_{p=1}^{P} P(\beta_{i,p}|\boldsymbol{\theta_{\beta_p}}) d\beta_{i,p} \\
ML(\boldsymbol{\theta_\alpha}, \boldsymbol{\theta_{\beta_1}}, \ldots, \boldsymbol{\theta_{\beta_P}}, G_M) &= \prod_{i=1}^{M} \prod_{j=1}^{n_i} p_{ij} \prod_{i=1}^{I} \prod_{j=1}^{n_i} MGF(\boldsymbol{\theta_\alpha}; 1) \prod_{p=1}^{P} MGF(\boldsymbol{\theta_{\beta_p}}; \boldsymbol{h(x_{i,j})}), \qquad (3.9)
\end{aligned}$$

where $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_I, \beta_{1,1}, \ldots, \beta_{I,P})$ as in (3.3). The result is thus in terms of the moment generating functions of our prior distributions (MGF) and the tilt function $\boldsymbol{h(x_{i,j})}$. The log-likelihood, subject to the appropriate constraints (similar to those imposed on Theorem 3.2.1), can be optimized to estimate the model. Of course, prior distributions need to be chosen to ensure statistical identifiability.

3.4.2 Assuming homogeneity in the intercepts

To make estimation easier, the researcher may prefer to restrict assumptions regarding heterogeneity of coefficients corresponding to the explanatory variables only. We can do so by assuming that the intercepts follow a delta mass at a particular constant c_α such that $\alpha_i \sim \delta(c_\alpha)$:

$$\begin{aligned}
ML(c_\alpha, \boldsymbol{\theta_{\beta_1}}, \ldots, \boldsymbol{\theta_{\beta_P}}, G_M) &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} L(\boldsymbol{\theta}, G_M) \prod_{i=1}^{I} \delta(c_\alpha) d\alpha_i \prod_{p=1}^{P} P(\beta_{i,p}|\boldsymbol{\theta_{\beta_p}}) d\beta_{i,p} \\
&= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{i=1}^{M} \prod_{j=1}^{n_i} p_{ij} \prod_{i=1}^{I} \prod_{j=1}^{n_i} e^{\alpha_i + \boldsymbol{\beta_i' h(x_{i,j})}} \delta(c_\alpha) d\alpha_i \\
&\quad \cdot \prod_{p=1}^{P} P(\beta_{i,p}|\boldsymbol{\theta_{\beta_p}}) d\beta_{i,p} \\
&= \prod_{i=1}^{M} \prod_{j=1}^{n_i} p_{ij} \prod_{i=1}^{I} \prod_{j=1}^{n_i} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{\alpha_i + \boldsymbol{\beta_i' h(x_{i,j})}} \delta(c_\alpha) d\alpha_i \\
&\quad \cdot \prod_{p=1}^{P} P(\beta_{i,p}|\boldsymbol{\theta_{\beta_p}}) d\beta_{i,p} \\
ML(c_\alpha, \boldsymbol{\theta_{\beta_1}}, \ldots, \boldsymbol{\theta_{\beta_P}}, G_M) &= \prod_{i=1}^{M} \prod_{j=1}^{n_i} p_{ij} \prod_{i=1}^{I} \prod_{j=1}^{n_i} e^{c_\alpha} \prod_{p=1}^{P} MGF(\boldsymbol{\theta_{\beta_p}}; \boldsymbol{h(x_{i,j})}). \qquad (3.10)
\end{aligned}$$

In the following section, we apply our approach to a fundamental aspect of the American civil justice system with applications to health care reform.

3.5 An Application: Tort Reform

In this section, we analyze tort loss data across all fifty states. Recent research has found that tort reform has had a significant impact on reducing insurance premiums as well as on tort losses throughout the country. We utilize the Bayesian DRE approach in this study to quantify the probability of tort losses exceeding a specified threshold [19] for 2004 and 2006. Computation of these probabilities enables us to understand on how litigious the country indeed is, thereby shedding light on the efficacy of recently instituted tort reforms.

3.5.1 Data

Our dataset was the same data used in the Crain et al (2009) study and was provided to us by two of the paper's authors [19]. We examined per capita tort losses defined as the "payments by defendants (or their insurance companies) for judgments, settlements, attorney fees, and administrative expenses in tort lawsuits ...", in thousands of (real 2006) dollars per capita, of each of the fifty U.S. states in 2004 and 2006 [19]. For this analysis, we adhered to examining medical malpractice tort losses, although analysis of other aspects of the civil justice system (such as automobile insurance, product liability insurance, and homeowners insurance among others) are worthwhile endeavors for future research.

Our data is denominated in losses per capita to enable comparable analysis across the different states. The data set consisted of one observation for each of the fifty states (for a total sample size of 50 for each year), similar to what we assumed in the simulation above. As different states and localities throughout the country have the potential to be more litigious than others, it is important to be able to capture this heterogeneity in estimating the probability of tort losses exceeding a particular threshold. This fact, coupled with the fact that our data set contains only one observation per state, makes our semiparametric Bayesian approach particularly useful for reducing the dimensionality of the problem.

3.5.2 Estimation

We examined 2004 and 2006 separately, treating each datum within each year as an independent single-observation sample with its own unique parametrization. It is worthwhile to compare the distributions of per capita tort losses between 2004 and 2006 to understand the efficacy of tort reforms that had been recently instituted around that time period [19]. In particular, we used the following model specification:

$$\frac{g_i(x)}{g(x)} \quad = \quad e^{\alpha_i+\beta_i x}; \quad i = 1, \ldots, 50 \tag{3.11}$$

and let $\alpha_i \sim N(\mu_\alpha, 1)$ and $\beta_i \sim N(\mu_\beta, \sigma^2_\beta)$ as we did in Section 3.2. By allowing our model's coefficients to vary for each state, we enable our model to capture state-level heterogeneity. We estimated the marginalized distribution by fusing the sample with data regarding tort losses from 1996. After estimating our marginalized distribution, we then used the cumulative distribution outlined in (3.7) to compute the probabilities of extreme tort losses. These probabilities better enable us to understand the risks associated with litigation across the country.

Additionally, we used a bootstrap approach to develop 95% confidence intervals around these point estimates. Specifically, we resampled our dataset (with replacement) 1000 times and re-estimated our probabilities for each sample. We used the resulting set to generate our interval estimates. Our results are are outlined in Tables 3.4-3.7.

Tab. 3.4: Coefficient Estimates - 2004, Using Bayesian DRE Approach

Coefficient	Estimate	Lower 95% CI	Upper 95% CI
μ_α	-1.3514	-2.0632	-0.5848
μ_β	0.0419	-0.0300	0.0765
σ^2_β	0.0003	0.0000	0.0031

Tab. 3.5: Coefficient Estimates - 2006, Using Bayesian DRE Approach

Coefficient	Estimate	Lower 95% CI	Upper 95% CI
μ_α	0.4196	-0.2142	1.0331
μ_β	-0.1102	-0.1896	-0.0368
σ^2_β	0.0044	0.0066	0.0076

Tab. 3.6: Analysis of 2004 Tort Loss Data, Using Bayesian DRE Approach

Probability	Estimate	Lower 95% Limit	Upper 95% Limit
P(Tort Losses > 35000)	0.100	0.010	0.148
P(Tort Losses > 45000)	0.085	0.005	0.104
P(Tort Losses > 55000)	0.019	0.000	0.060

Tab. 3.7: Analysis of 2006 Tort Loss Data, Using Bayesian DRE Approach

Probability	Estimate	Lower 95% Limit	Upper 95% Limit
P(Tort Losses > 35000)	0.068	0.010	0.144
P(Tort Losses > 45000)	0.045	0.005	0.102
P(Tort Losses > 55000)	0.019	0.000	0.060

We also estimated the goodness of fit diagnostic statistics outlined in Table 3.8 for $k = 1$ and $k = 2$, comparing it to the existing DRE method. These results are outlined in Table 3.8. Additionally, Figures 3.1-3.4 depict plots comparing the estimated distributions of per capita tort losses $\hat{H}$ from (3.7) versus the corresponding empirical CDFs for 2004 and 2006. The plots suggest that the comparative fit between the two models is quite comparable in 2004 (as there is near perfect agreement between the estimated and empirical CDFs) but substantial improvement as a result of the Bayesian approach compared to the standard approach in 2006.

Tab. 3.8: Goodness of Fit Diagnostics

	DRE		Bayesian DRE	
	$R^2_{0.95,1}$	$R^2_{0.95,2}$	$R^2_{0.95,1}$	$R^2_{0.95,2}$
2004	0.915	0.998	0.911	0.997
2006	0.108	0.129	0.607	0.582

Our results illustrate that the Bayesian DRE approach substantially improved model fit in 2006 suggesting that there was unobserved heterogeneity across the country during that year that the standard DRE approach was unable to properly model. Although there are not really any noticeable differences between the two approaches in 2004 (in fact the standard DRE model performs slightly better), this

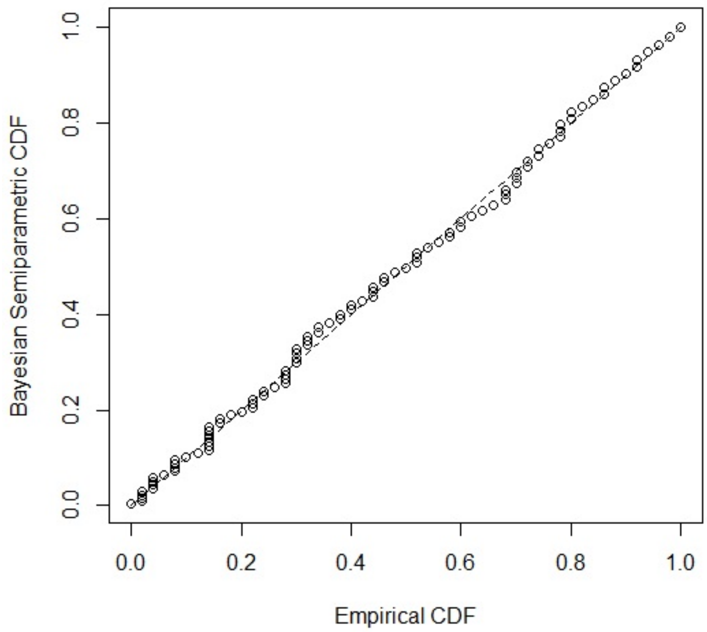

Fig. 3.1: Plot of $\hat{H}$ vs. $\tilde{H}$ - DRE, 2004 Per Capita Tort Loss Data

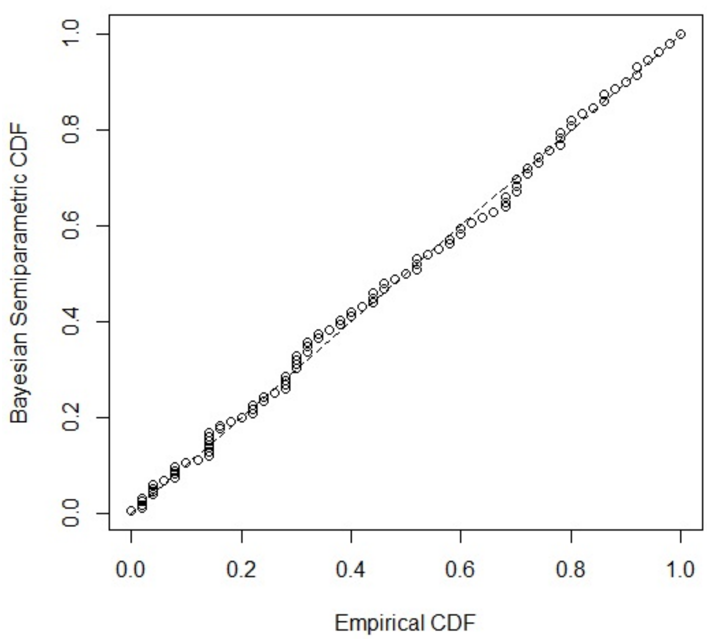

Fig. 3.2: Plot of $\hat{H}$ vs. $\tilde{H}$ - Bayesian DRE, 2004 Per Capita Tort Loss Data

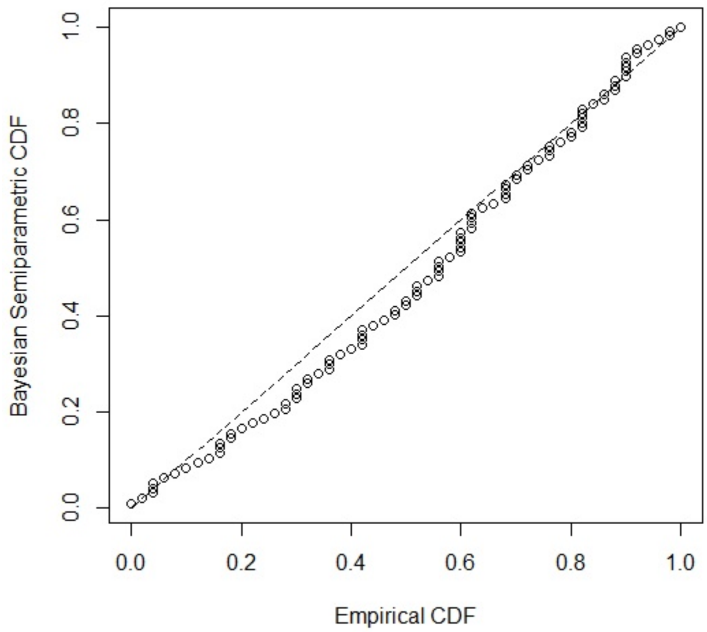

Fig. 3.3: Plot of $\hat{H}$ vs. $\tilde{H}$ - DRE, 2006 Per Capita Tort Loss Data

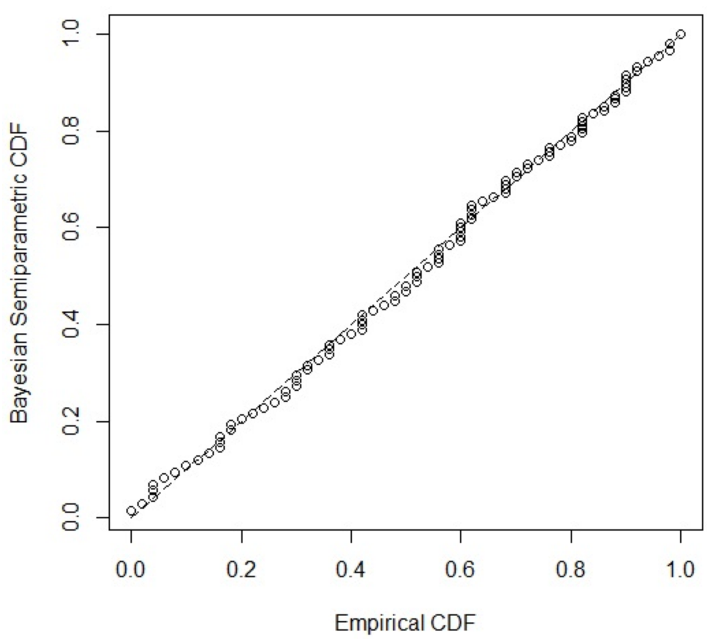

Fig. 3.4: Plot of $\hat{H}$ vs. $\tilde{H}$ - Bayesian DRE, 2006 Per Capita Tort Loss Data

similarity can be explained by the fact that in 2004, the 50 states may have been considerably more homogeneous in nature regarding litigation. In terms of per capita tort losses, the probabilities of extreme losses clearly declined between 2004 and 2006.

In terms of actual tort losses, we notice a reduction in the probabilities of extreme losses in 2006 compared to 2004. These results, in conjunction with the results from Crain et al (2009), illustrate the efficacy of the number of state-based medical malpractice reforms, many of which had been instituted around this time period [19]. Such reforms have included attorney fee limitations, requirements for pre-trial screening, standards regarding expert witnesses, imposition of strict statute of limitations for filing lawsuits, and economic damage caps. These results may also be explained by the fact that existing tort laws, not necessarily recently instituted, may have become more stringently enforced in 2006 compared to 2004. As a side note, although the coefficient estimates of σ^2_β appear to be low, they should not be taken to imply that there is no heterogeneity in the model. As the units of our data are in thousands of dollars per capita, seemingly small per capita differences between states imply substantial variation in litigiousness amongst the states.

3.5.3 Policy Implications

A 2004 study conducted by the non-partisan Congressional Budget Office found that state-based tort reform reduced the number of lawsuits filed, decreased the number of damage awards, and lowered insurance claims [17]. Our results above also illustrate that the risks associated with the civil justice system notably declined between 2004 and 2006. These reductions were primarily due to state-based tort reforms that were instituted throughout the country, including economic caps, appeal-bond caps, and standards regarding expert witnesses [19].

Although these results illustrate the efficacy of certain malpractice reforms instituted around this time period, considerably more can be done to reform the civil justice system including passing tort reforms in states where important laws are lacking and more stringently enforcing existing laws [19, 103]. Scholars at the Heritage Foundation, the Cato Institute, the American Enterprise Institute, and the Brookings Institution have discussed the benefits of malpractice reform, noting that such reforms have the potential to improve the environment for physicians to practice, benefiting patients in the process [18, 93, 139, 142].

These benefits are quite well-known. For example, a 2005 study published in the *Journal of the American Medical Association* found that malpractice reform increased physician supply, particularly with specialties associated with high malpractice costs [86]. The authors found that without appropriate reforms, the medical malpractice climate had been encouraging early retirements in and discouraging entry to particularly risky specialties such as obstetrics, anesthesiology, radiology, and surgery. Additionally, such an environment results in the practice of defensive medicine, causing physicians to order unneeded tests due to the fear of lawsuits, unnecessarily increasing health care costs [85, 151].

3.6 Conclusions and Future Research

Our study illustrates the efficacy of medical malpractice reform in reducing the probabilities of extreme tort losses. Medical malpractice reform is of course by no means a "silver bullet" for fixing our nation's broken health care system. Such reforms, however, are a very helpful component of the supply side of health care. On the demand side, policymakers should seek to instill free market competition in the industry, which will notably reduce costs and improve quality of care [13, 20, 21]

Methodologically, our study provides a notable contribution to both Bayesian estimation methodologies as well as to semiparametric modeling. In particular, we have applied empirical Bayesian methods to semiparametric density ratio modeling, allowing statisticians to incorporate individual-level heterogeneity in such models. An interesting aspect of this approach is that our marginalization yields a closed-form expression allowing us to make direct statistical inferences about the population without having to resort to the numerical approaches typically concomitant with Bayesian methods. As discussed in Chapter 2, these numerical approaches such as MCMC methods, can be quite computationally intensive, particularly for large data sets with high-dimensional parameterizations [50].

Additionally, although we focused our application on medical malpractice reform, there is no reason that this approach cannot be used in other settings where modeling individual-level heterogeneity is important without being forced to adhere to strict parametric assumptions. In particular, the approach outlined here can be useful in many application areas including medical research [160], economics [56], understanding athletic performance [24, 107], and other areas within public policy research [19, 136] among others. Regardless of the application, we believe that the Bayesian approach outlined here can be yet another useful tool for statisticians to use.

Chapter 4: Bayesian Inferences of Welfare Reform with Improved Credible Interval Estimation via Semiparametric Out of Sample Fusion

4.1 Introduction

4.1.1 "Ending Welfare as We Know it"

Welfare reform is a hot topic of debate amongst policymakers in America [61, 135,152]. The welfare state can be defined as a government system aimed at helping the disadvantaged by providing benefits such health care, pensions, and financial programs aimed to help those in need [14]. The first Chancellor of Germany, Otto von Bismarck, constructed the beginnings of the modern European welfare state in the late 1800s, creating programs such as old age pensions, accident insurance, and medical care. In the early 20th century, the welfare state came to Britain with the introduction of pension programs, free school meals, and unemployment and health benefits. Shortly afterwards, the welfare state began to grow throughout other European countries [14].

Although the United States does not have an overarching welfare state that many European countries do, the country does offer a variety welfare programs [14]. Initiated in the United States in the early 1900s, federal welfare programs were designed with the intention of operating as safety net programs to aid the disadvantaged. President Theodore Roosevelt, for example, proposed a number of such ideas as part of his ‘New Nationalism’ platform. During that time period, Roosevelt believed it was incumbent upon the federal government to institute programs such as a National Heath Service; social insurance for the unemployed, disabled, and elderly; and to provide workers’ compensation to those with certain work-related injuries [84].

As a component of his New Deal, President Franklin Delano Roosevelt instituted a number of federal cash assistance programs to assist the poor. These programs included Social Security, Aid to Dependent Children (ADC), and unemployment compensation programs among others. Several decades later, President Lyndon Johnson instituted the War on Poverty to further expand America’s welfare state with the goal of eradicating poverty. In particular, he expanded President Roosevelt’s ADC programs (which by this time were known as Aid to Families with Dependent Children or AFDC programs), began food stamp programs, public housing, Medicare, and Medicaid, all of which were intended to be “safety nets” for those in poverty [14].

In *Losing Ground*, Charles Murray of the American Enterprise Institute rigorously evaluated the effects of these types of welfare programs [111]. He concluded that without doubt many of these programs had the potential to foster dependency and cripple those they intend to help. Murray’s work shattered much of the political capital advocates had for open-endedly expanding America’s welfare state.

In 1992, Presidential Candidate Bill Clinton campaigned to “end welfare as we know it,” [16]. Two years later, Congressional Republicans made welfare reform a fundamental component of their Contract with America [53]. The goal of these reforms were to transform welfare away from a system that encouraged endless dependence on taxpayer dollars and instead transform it into a system providing temporary assistance to get people back to work and become active contributing members of society.

In 1996, Congressional Republicans worked with President Clinton to reform AFDC Programs [53]. Intended to operate as a safety net for those in poverty, AFDC programs were simply a government handout to qualifying single mothers. The government mailed checks to recipients who had virtually no responsibilities in return. In 1996, President Clinton signed into law the Personal Responsibility and Work Authorization Act (PROWA), restructuring AFDC programs into a more restrictive program known as Temporary Assistance to Needy Families (TANF) [118]. Under TANF, state governments were required to impose federal work standards on welfare recipients. In particular, TANF required recipients to work or study as a condition for receiving welfare. In addition, TANF also limited receipt of welfare benefits to a five year time frame. These requirements, although not particularly rigorous, were intended to transform the welfare program from a system that fostered dependency into a springboard that gave the disadvantaged temporary help to become contributing members of society [137].

It is important for policymakers to be able to rigorously examine such fundamental changes to government programs. Many reforms, including PROWA, have been constructed to provide states a certain degree of flexibility in implementing the law. As a result, different states have the potential to manifest different responses to the law. Although a number of policy studies have made some quantitative evaluations of PROWA, no studies, to our knowledge, have done so capturing this state-level heterogeneity in the associated statistical models [61, 132, 152].

We perform such an analysis in this study. In particular, we statistically evaluate the success of PROWA in helping to get people back to work. We do so by estimating hierarchical Bayesian linear models to rigorously compare AFDC programs (before welfare reform) to TANF programs (after welfare reform). We find that TANF was considerably more successful than AFDC in getting people back to work, illustrating the success of the PROWA.

In the process of our analysis, we improve on existing Bayesian statistical estimation techniques. In particular, standard Bayesian estimation techniques suffer from a number of significant limitations. Specifically, standard Markov Chain Monte Carlo (MCMC) methods used to estimate Bayesian statistical models can require a significant amount of time to adequately sample a posterior density, especially for large data sets involving high-dimensional parameter spaces [50]. Furthermore, despite diagnostic checks, it is difficult to truly know whether an MCMC sampler has adequately sampled the parameter space. As a result, posterior inferences based on MCMC sampling may or may not necessarily reflect reality. We provide a novel adaptation of a semiparametric estimation technique, used thus far only for frequentist statistical models, to more accurately summarize the posterior by providing more precise estimates of the density's credible intervals [160].

The contribution of this study is therefore twofold. Firstly, we statistically examine the efficacy of the welfare reform of the 1990s. Secondly, we improve on standard Bayesian estimation techniques for hierarchical linear models. As we discuss, our improvements to Bayesian computation are useful for a variety of models beyond just the hierarchical linear model structure studied here.

4.2 Statistically Modeling the Impact of Welfare Reform

4.2.1 A Hierarchical Bayesian Model Allowing for State-Level Heterogeneity

We are interested in understanding the efficacy of the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 [118]. This law imposed a five year time limit on welfare benefits and imposed some, albeit not particularly onerous requirements, on individuals to either work or study as a condition of receiving welfare. To understand the efficacy of this reform, we looked at the effect of the total number of welfare caseloads in each state on the number of employment related discontinuances of receiving welfare. Mathematically, this relationship can be modeled over $i \in \{1, \ldots, 53\}$ territories (50 states as well as Guam, Puerto Rico, and the District of Columbia) (units) over $t \in \{1, \ldots, T\}$ years.

Consider a data set consisting of $i \in \{1, \ldots, I\}$ observations (in our case state/territories) over $t \in \{1, \ldots, T\}$ occasions. We define our dependent vector valued variable to be $\boldsymbol{y_i} = (y_{i,t})_{t=1}^{T}$, which represents the welfare caseload discontinuances due to employment for state/territory i at time t. We define our explanatory vector valued variable, $\boldsymbol{x_i} = (x_{i,t'})_{t'=1}^{T'}$, as a vector of dimension equal to $\boldsymbol{y_i}$ representing the total number of welfare caseloads at time t' (for $t' < t$) as we wish to relate $y_{i,t}$ to past covariates. Thus, $\boldsymbol{x_i}$ represents a lagged number of welfare caseloads. As the PROWA imposed a new five year time limit for TANF recipients, which had previously not been present for AFDC programs, we conducted our analysis for lags of 1,2, 3, 4, and 5 years.

$$\boldsymbol{y_i} \sim \mathrm{MVN}(\beta_{i,0} + \beta_{i,1}\boldsymbol{x_{i,1}}, \Sigma) \tag{4.1}$$

$$\beta_{i,0} \sim N(\mu_0, \sigma_0^2) \tag{4.2}$$

$$\beta_{i,1} \sim N(\mu_1, \sigma_1^2) \tag{4.3}$$

$$\boldsymbol{\Sigma^{-1}} \sim \mathrm{Wishart}(\Phi_0, \nu_0) \tag{4.4}$$

$$\mu_0 \sim N(0, 10) \tag{4.5}$$

$$\mu_1 \sim N(0, 10) \tag{4.6}$$

$$\sigma_0^{-2} \sim \Gamma(10, 10) \tag{4.7}$$

$$\sigma_1^{-2} \sim \Gamma(10, 10) \tag{4.8}$$

$\boldsymbol{\Sigma}$ is a T-dimensional variance-covariance matrix that enables us to capture cross-year correlations (for example caseloads reductions in one particular state in one year may have a strong correlation with caseload reductions in that same state the following year). Another way of viewing this model would be to estimate a hierarchical Bayesian linear model defined as $y_{i,t} = \beta_{i,0} + \beta_{i,1}x_{i,t} + \epsilon_{i,t}$ for $i = 1, \ldots, I, t = 1, \ldots, t-1$ with the prior structure discussed above and with non-zero covariances between $\epsilon_{i,t'}$ and $\epsilon_{i,t}$ $\forall t, t' \leq T$.

To ensure statistical identifiability of our model, we requred $\beta_{1,0} \equiv 0$.[1] In the next section, we discuss estimation of this model, including improvements to estimating the posterior density's credible intervals.

4.2.2 The Limitations of Bayesian Computational Methods

Typically, these types of hierarchical Bayesian models lack an analytic form for their posterior functionals and are therefore estimated numerically using standard

[1] This coefficient pertains to the intercept regarding Alabama as our data was organized alphabetically.

Markov Chain Monte Carlo (MCMC) methods [50]. Researchers use the resulting MCMC samples to generate a variety of statistical estimators such as posterior means, posterior standard deviations, and endpoints depicting credible intervals.

There are a number of limitations associated with MCMC sampling, however. In particular, MCMC sampling can be computationally intensive, particularly for models involving high-dimensional parameter spaces. Recent research has developed alternative approaches for Bayesian estimation that attempt to address this issue. In particular, there are approaches using variational approximations to posterior densities, polynomial expansion approaches, and alternative sampling approaches among others [9, 10, 39, 57, 108].

Another limitation of MCMC methods is that although there are some evaluatory tools (e.g. Gelman and Rubin 1992) to assess convergence of the chain, these measures are simply just diagnostics, and it is difficult to determine with certainty if the MCMC chain has sufficiently navigated the posterior distribution's parameter space [51]. Although MCMC samplers theoretically do converge under reasonable conditions, this convergence is only guaranteed to occur asymptotically over an infinitely large number of draws [154]. In reality, however, the Bayesian MCMC samplers need to eventually be truncated. As a result, researchers are often rightfully concerned with whether their posterior sample based on truncation of the MCMC sampler truly represents the posterior density. A classic pathological example demonstrating this phenomenon involves MCMC sampling of the so-called "witch's hat," where the MCMC sampler can get stuck in a particular area of the posterior distribution [141]. Premature truncation of the MCMC sampler can consequently result in misleading statistical inferences regarding the posterior distribution.

We can mitigate this problem considerably by applying a semiparametric methodology, known as density ratio estimation, to posterior samples [123–125, 160].

This estimation technique can enable a researcher to more accurately estimate credible intervals. Although this semiparametric approach has been used in a variety of settings, no researchers, to our knowledge, have applied this method to Bayesian estimation.

4.2.3 Statistical Inferences Based on Density Ratio Estimation

Density ratio estimation (hereafter referred to as DRE) is a semiparametric statistical estimation technique. This technique has been applied to many settings including AIDS vaccine trials, the analysis of variance, mortality rate prediction, cluster detection, and cancer research among others [43,44,52,80,81,83]. In the one dimensional case, there are $i = I + 1$ random samples $\boldsymbol{x_i}$ with sample size n_i such that $\sum_{i=1}^{I+1} n_i = n$:

$$\boldsymbol{x_i} = (x_{i,1}, \ldots, x_{i,n_i}),$$

with probability density functions g_i, such that:

$$x_{i,j} \sim g_i, \qquad i = 1, \ldots, I, I + 1, j = 1, \ldots, n_i. \tag{4.9}$$

In utilizing this method, the statistician a priori assumes that $g_{I+1} \equiv g$ defines a reference probability density with a particular ratio between g_i and g $\forall$ $i = 1, \ldots, I$. In many cases, this ratio is an exponential involving a vector-valued function $\boldsymbol{h}(x)$, known as a tilt function:

$$\frac{g_i(x)}{g(x)} = e^{\alpha_i + \boldsymbol{\beta_i'} \boldsymbol{h}(x)}. \tag{4.10}$$

If both $g_i(x)$ and $g(x)$ are densities belonging to the exponential family having functional form:

$$
\begin{aligned}
g(x,\boldsymbol{\theta}) &= d(\boldsymbol{\theta})S(x)\exp\big[\sum_{j=1}^{J} c_j(\boldsymbol{\theta})T_j(x)\big] \\
&= \exp\Big[\sum_{j=1}^{J} c_j(\boldsymbol{\theta})T_j(x) + \log\big[d(\boldsymbol{\theta})S(x)\big]\Big], \qquad (4.11)
\end{aligned}
$$

where $\boldsymbol{\theta}$ is a parameter to be estimated, then the densities of any two such members will clearly have a ratio satisfying (4.15), as will the ratio of a single truncated such member coupled with the density of a uniformly distributed random variable. Some examples of special cases resulting in commonly used tilt functions are $\boldsymbol{h}(x) = (x, x^2)'$ (appropriate for normally distributed data):

$$\frac{g_i(x)}{g(x)} = e^{\alpha_i+\beta_i x+\gamma_i x^2}, \qquad (4.12)$$

$\boldsymbol{h}(x) = (x, \log(x))'$ (appropriate for data coming from a gamma distribution), and

$$\frac{g_i(x)}{g(x)} = e^{\alpha_i+\beta_i x+\gamma_i \log(x)}, \qquad (4.13)$$

$\boldsymbol{h}(x) = (x^\tau, \log(x))'$ (appropriate for potentially more heavily-skewed data coming from a generalized gamma distribution):

$$\frac{g_i(x)}{g(x)} = e^{\alpha_i+\beta_i x^\tau+\gamma_i \log(x)}. \qquad (4.14)$$

Parametric assumptions regarding densities, however, are sufficient but not necessary conditions for density ratios, such as those above, to be able to properly model real-world phenomena. Recent research, for example, has shown that the distributions described in (4.13) and (4.14) are useful in modeling situations where the data does not necessarily adhere to the strict parameterizations described by known parametric distributions [78, 82, 168].

4.2.4 Empirical Likelihood Estimation in Semiparametric Density Ratio Models

Suppose we assume the density ratio outlined in (4.15) with tilt function $\boldsymbol{h}(x)$,

$$\frac{g_i(x)}{g(x)} \;=\; e^{\alpha_i+\boldsymbol{\beta_i'}\boldsymbol{h}(x)}. \tag{4.15}$$

If we let $G(x)$ be the reference CDF and define probability masses $p_{ij} = dG(x_{i,j}) = dG_{I+1}(x_{i,j})$, then we can utilize the method of constrained empirical likelihood to estimate g_i and the associated parameters as follows. We can write our empirical likelihood function, parametrized by $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_I, \boldsymbol{\beta_1}, \ldots, \boldsymbol{\beta_I})$ (a vector of dimension Id), based on our pooled data:

$$L(\boldsymbol{\theta}, G) \;= \prod_{i=1}^{I+1}\prod_{j=1}^{n_i} p_{ij} \prod_{i=1}^{I}\prod_{j=1}^{n_i} e^{\alpha_i+\boldsymbol{\beta_i'}\boldsymbol{h}(x_{i,j})}. \tag{4.16}$$

The resulting log-likelihood is given by:

$$l \;= \log L = \sum_{i=1}^{I+1}\sum_{j=1}^{n_i} \log p_{ij} + \sum_{i=1}^{I}\sum_{j=1}^{n_i} \big(\alpha_i + \boldsymbol{\beta_i'}\boldsymbol{h}(x_{i,j})\big), \tag{4.17}$$

subject to the following constraints:

$$p_{ij} > 0, \;\; \sum_{i=1}^{I+1}\sum_{j=1}^{n_i} p_{ij} = 1, \;\; \sum_{i=1}^{I}\sum_{j=1}^{n_i} p_{ij} e^{\alpha_k+\boldsymbol{\beta_k'}\boldsymbol{h}(x_{i,j})} = 1, \text{ for } k = 1, \ldots, I. \tag{4.18}$$

The following conditions, as discussed in Fokianos (2004) and Qin and Lawless (1995) as cited by Voulgaraki (2011) ensure the existence of an empirical likelihood estimator [43, 123, 159]. Specifically, if we let $f(x, \hat{\boldsymbol{\theta}}) = (e^{\hat{\alpha}_1+\hat{\boldsymbol{\beta}}_{\boldsymbol{1}}\boldsymbol{h}(x)} - 1, \ldots, e^{\hat{\alpha}_I+\hat{\boldsymbol{\beta}}_{\boldsymbol{I}}\boldsymbol{h}(x)} - 1)'$, then given the following conditions, there exists, with probability approaching one, an extremum in a ball around the true parameter vector $\boldsymbol{\theta_0}$:

Lemma 4.2.1. *Under the assumptions that:*

(a) $E(f(x,\boldsymbol{\theta_0})f'(x,\boldsymbol{\theta_0}))$ *is positive definite,*

(b) $\frac{\partial f}{\partial \boldsymbol{\theta}}$ *is continuous in a ball around the true value of* $\boldsymbol{\theta_0}$

(c) $||\frac{\partial f}{\partial \boldsymbol{\theta}}||$ *and* $||f(x,\boldsymbol{\theta})||^3$ *are bounded by an integral function with respect to* G *in this ball,*

(d) $E\left(\frac{\partial f}{\partial \boldsymbol{\theta}}\right)$ *is of rank* Id*,*

where E is the expectation with respect to the probability measure corresponding to G, then the coefficients α_i and $\boldsymbol{\beta_i}$, as well as the discrete estimators of $G_i(x)$ $\forall\, i = 1, \ldots, I+1$ can be estimated by optimizing (4.17) subject to the constraints outlined in (4.18) via a two step estimation procedure utilizing the method of Lagrange multipliers. Specifically, if we can define $\mu_k \equiv \lambda_k/n$, where λ_k is the pertinent Lagrange multiplier we find the following estimators, $\hat{p}_{ij}$ and $\hat{G}_i(x)$ for $i = 1, \ldots, I+1$, as follows:

$$\hat{p}_{ij} = \frac{1}{n}\frac{1}{1+\sum_{k=1}^{I}\hat{\mu}_k[e^{\hat{\alpha}_i+\boldsymbol{\hat{\beta}_i'}\boldsymbol{h}(x_{ij})}-1]} \tag{4.19}$$

and:

$$\begin{aligned}\hat{G}_{I+1}(x) &= \sum_{i=1}^{I+1}\sum_{j=1}^{n_i}\hat{p}_{ij}I(x_{ij}\le x)\\ &= \frac{1}{n}\sum_{i=1}^{I+1}\sum_{j=1}^{n_i}\frac{I(x_{ij}\le x)}{1+\sum_{k=1}^{I}\hat{\mu}_k[e^{\hat{\alpha}_i+\boldsymbol{\hat{\beta}_i'}\boldsymbol{h}(x)}-1]}. \end{aligned} \tag{4.20}$$

We can generalize this result to distributions $\hat{G}_i$ for $i = 1, \ldots, I$ to generate the following estimators:

$$
\begin{aligned}
\hat{G}_k(x) &= \sum_{i=1}^{I+1}\sum_{j=1}^{n_i} \hat{p}_{ij} e^{\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_{\boldsymbol{k}}' \boldsymbol{h}(x)} I(x_{ij} \le x) \\
&= \frac{1}{n}\sum_{i=1}^{I+1}\sum_{j=1}^{n_i} \frac{e^{\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_{\boldsymbol{k}}' \boldsymbol{h}(x)} I(x_{ij} \le x)}{1 + \sum_{k=1}^{I} \hat{\mu}_k [e^{\hat{\alpha}_i + \hat{\boldsymbol{\beta}}_{\boldsymbol{i}}' \boldsymbol{h}(x)} - 1]}, \text{ for } k = 1, \ldots, I. \qquad (4.21)
\end{aligned}
$$

Furthermore, estimators of the probability densities $g_i(x)$ can be obtained via kernel density estimation applied to the jumps of $\hat{G}_i \ \forall\, i = 1, \ldots, I+1$. The interested reader should refer to Voulgaraki et al (2012), which provides a methodology for a complete discussion of the approach including techniques for determining the optimal bandwidth needed for accurate kernel density estimation [160].

Additionally, as long as the following properties hold, our estimators are statistically unbiased and asymptotically normal as illustrated in the following theorem. In particular, by defining the vector $\boldsymbol{\mu} \equiv (\mu_1, \ldots, \mu_I)$ as the Lagrange multipliers for the optimization of the empirical likelihood function subject to the constraints outlined in (4.18) (just as has been done above) and by letting $\boldsymbol{\mu_0}$ denote the true value of $\boldsymbol{\mu}$, then, we can state the following theorem:

Theorem 4.2.2. *Suppose the following conditions hold:*

(a) $\frac{\partial^2 f}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ *is continuous in a ball around the true value of* $\boldsymbol{\theta_0}$

(b) There exists an integrable function with respect to G *bounding* $||\frac{\partial^2 f}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}||$

(c) The four conditions outlined in Lemma 4.2.1 hold.

Then:

$$
\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0} \\ \hat{\boldsymbol{\mu}} - \boldsymbol{\mu_0} \end{pmatrix} \Rightarrow P, \qquad (4.22)
$$

where $\Rightarrow$ *denotes weak convergence to a probability measure* P*, induced by a multivariate normal random vector with mean* $\mathbf{0}$ *and a particular variance-covariance* *matrix* $\mathbf{\Sigma}$.

The details of the structure of $\mathbf{\Sigma}$ are complicated and derived in detail in Fokianos (2004). Lu (2007) contains a comprehensive proof of Theorem 4.2.2 [43,96].

One of the most appealing aspects about this semiparametric approach is that it can more accurately determine probability distributions from data compared to its more restrictive parametric counterparts. This advantage is due to the fact that the combined sample of larger size is used rather than any particular sample consisting of a smaller number of elements. In Section 4.2.2, we discussed many of the limitations of present day standard Bayesian computational methods. In the following section, we illustrate how this semiparametric density ratio technique can be used to ameliorate these limitations.

4.2.5 Numerical Simulations

In this section, we present a series of numerical simulations to illustrate the efficacy of using DRE to generate accurate estimates of percentiles from a variety of different samples. In particular, we applied the DRE method, assuming the ratio described in (4.12) through (4.14) to samples from known parametric distributions (normal, student-t, gamma, and Weibull) to determine the samples' 2.5th and 97.5th percentiles. Since the supports of these distributions are infinite, fusing a sample from any of these distributions with a uniformly distributed sample signifies that the density ratio model holds approximately.

Specifically, we applied the DRE approach outlined in Section 4.2.3 with $I = 1$ (i.e. 2 samples), taking a sample from one of our known distributions, given a particular choice of parameters, and fused this sample with a random sample drawn from a uniform distribution with support including this sample's entire range. This uniformly distributed sample served as our reference distribution. This methodology, known as "out of sample fusion," was first introduced by Zhou (2013) as a tool for estimating small probabilities [168].

We used the DRE approach to estimate the 2.5th and 97.5th percentiles ($DRE_{k,2.5}$ and $DRE_{k,97.5}$) of our original distribution over the course of $k = 1, \ldots, K$ simulations. We compared these estimates to the true (2.5th and 97.5th) percentiles of our original population, $Q_{k,2.5}$ and $Q_{k,97.5}$, by computing the squared difference between the two.

We performed $K = 100$ such simulations for each choice of parameters, summed these squared differences for each percentile, and computed averages across all simulations. We hereafter refer to these mean squared differences as $MSE_{DRE,2.5}$ and $MSE_{DRE,97.5}$:

$$MSE_{DRE,2.5} = \frac{\sum_{k=1}^{K} (DRE_{k,2.5} - Q_{k,2.5})^2}{K} \tag{4.23}$$

and:

$$MSE_{DRE,97.5} = \frac{\sum_{k=1}^{K} (DRE_{k,97.5} - Q_{k,97.5})^2}{K}. \tag{4.24}$$

The use of known parametric distributions enables us to understand the efficacy of using DRE for estimating low and high level quantiles of distributions. As a result, these simulations shed light on the usefulness of using the DRE approach for credible interval estimation of Bayesian regression coefficients.

Zhou (2013) also developed an approach for quantifying the efficacy of using out of sample fusion for estimating probabilistic thresholds [168]. Zhou's work found that out of sample fusion results in shorter confidence intervals compared to using empirical distributions solely based on within sample data. These confidence intervals, however, are inherently frequentist in nature, based on the assumption that each sample is the single realization of a long-run frequency of an asymptotically large number of samples. This approach, although not completely useless, is not philosophically compatible with the Bayesian philosophy of statistical estimation. As a result, although such an approach sheds light on the efficacy of out of sample fusion, the approach presented here provides further verification as well as a more legitimate basis for applying the DRE method to Bayesian estimation.

Our results are outlined in the Tables 4.1-4.10, which we conducted for samples of size 600, 1000, 5000, 10000, 15000, 20000, and 25000. $\overline{DRE_{2.5}}$ and $\overline{DRE_{97.5}}$ represent the estimates of the 2.5th and 97.5th percentiles using the DRE method, averaged over the 100 simulations, and $\overline{E_{2.5}}$ and $\overline{E_{97.5}}$ represent these same averages using the samples' empirical CDFs. For the Weibull quantile estimation, we estimated via maximum likelihood estimation an estimator for τ assuming the data were a random sample from a generalized gamma distribution. We then fixed this estimate as the value for τ in the density ratio and proceeded with the semiparametric estimation.

We notice that for all of the sample sizes and distributions examined, $MSE_{DRE_{2.5}}$ and $MSE_{DRE_{97.5}}$ are substantially lower than $MSE_{E_{2.5}}$ and $MSE_{E_{97.5}}$. Thus, by simply fusing samples from known distributions with uniformly distributed data and estimating the sample's distribution using the semiparametric DRE approach with the combined data based on a reasonable choice of a tilt function, we are able to more accurately determine the 2.5th and 97.5th percentiles. As our analysis illustrates, for data that appears to be normally distributed (i.e. symmetric and unimodal, the tilt function $\boldsymbol{h}(x) = (x, x^2)'$ is reasonable, whereas for skewed data a tilt function of the form $\boldsymbol{h}(x) = (x, \log(x))'$ or more generally $\boldsymbol{h}(x) = (x^\tau, \log(x))'$ is acceptable.

Tab. 4.1: Simulation Results based on Samples from a Normal Distribution with mean μ and standard deviation σ, fused with uniformly distributed sample over $x_{\min}$ and $x_{\max}$, assuming tilt function $\boldsymbol{h}(x) = (x, x^2)'$

N	μ	σ	$MSE_{DRE_{2.5}}$	$MSE_{DRE_{97.5}}$	$MSE_{E_{2.5}}$	$MSE_{E_{97.5}}$
600	0	1	0.006	0.007	0.012	0.011
600	10	2	0.030	0.018	0.041	0.042
600	5	3	0.076	0.084	0.143	0.111
600	-2	5	0.176	0.190	0.234	0.222
1000	0	1	0.004	0.005	0.008	0.008
1000	10	2	0.015	0.020	0.021	0.027
1000	5	3	0.046	0.040	0.070	0.066
1000	-2	5	0.101	0.119	0.164	0.204
5000	0	1	0.001	0.001	0.001	0.002
5000	10	2	0.018	0.015	0.027	0.031
5000	5	3	0.039	0.041	0.056	0.066
5000	-2	5	0.103	0.114	0.159	0.184
10000	0	1	< 0.001	0.001	0.001	0.001
10000	10	2	0.017	0.017	0.034	0.029
10000	5	3	0.038	0.044	0.060	0.068
10000	-2	5	0.118	0.118	0.164	0.184
15000	0	1	< 0.001	< 0.001	0.001	< 0.001
15000	10	2	0.001	0.001	0.002	0.002
15000	5	3	0.003	0.003	0.004	0.005
15000	-2	5	0.008	0.007	0.012	0.010
20000	0	1	< 0.001	< 0.001	< 0.001	< 0.001
20000	10	2	0.017	0.014	0.029	0.029
20000	5	3	0.041	0.031	0.075	0.070
20000	-2	5	0.090	0.116	0.195	0.161
25000	0	1	< 0.001	< 0.001	< 0.001	< 0.001
25000	10	2	0.017	0.017	0.035	0.024
25000	5	3	0.051	0.042	0.066	0.079
25000	-2	5	0.113	0.067	0.234	0.143

Tab. 4.2: Further simulations based on Samples from a student-t Distribution with varying degrees of freedom, fused with uniformly distributed sample over $x_{\min}$ and $x_{\max}$, assuming tilt function $\boldsymbol{h}(x) = (x, x^2)'$

N	df	$MSE_{DRE_{2.5}}$	$MSE_{DRE_{97.5}}$	$MSE_{E_{2.5}}$	$MSE_{E_{97.5}}$
600	10	0.0127	0.0139	0.0253	0.0234
600	15	0.0114	0.0094	0.0191	0.0190
600	20	0.0089	0.0110	0.0146	0.0165
600	25	0.0101	0.0107	0.0147	0.0157
600	30	0.0076	0.0113	0.0140	0.0200
600	35	0.0069	0.0070	0.0118	0.0140
600	40	0.0087	0.0079	0.0148	0.0128
1000	10	0.0127	0.0139	0.0253	0.0234
1000	15	0.0114	0.0094	0.0191	0.0190
1000	20	0.0089	0.0110	0.0146	0.0165
1000	25	0.0101	0.0107	0.0147	0.0157
1000	30	0.0076	0.0113	0.0140	0.0200
1000	35	0.0069	0.0070	0.0118	0.0140
1000	40	0.0087	0.0079	0.0148	0.0128
5000	10	0.0020	0.0017	0.0026	0.0027
5000	15	0.0014	0.0015	0.0024	0.0021
5000	20	0.0015	0.0013	0.0024	0.0023
5000	25	0.0012	0.0013	0.0023	0.0022
5000	30	0.0010	0.0010	0.0016	0.0016
5000	35	0.0011	0.0018	0.0019	0.0016
5000	40	0.0010	0.0010	0.0020	0.0017
10000	10	0.0018	0.0019	0.0012	0.0011
10000	15	0.0017	0.0016	0.0012	0.0010
10000	20	0.0016	0.0016	0.0019	0.0011
10000	25	0.0015	0.0014	0.0019	0.0017
10000	30	0.0016	0.0016	0.0017	0.0010
10000	35	0.0015	0.0014	0.0017	0.0017
10000	40	0.0015	0.0015	0.0018	0.0017

Tab. 4.3: Further simulations based on Samples from a student-t Distribution with varying degrees of freedom, fused with uniformly distributed sample over $x_{\min}$ and $x_{\max}$, assuming tilt function $\boldsymbol{h}(x) = (x, x^2)'$

N	df	$MSE_{DRE_{2.5}}$	$MSE_{DRE_{97.5}}$	$MSE_{E_{2.5}}$	$MSE_{E_{97.5}}$
15000	10	0.0017	0.0017	0.0019	0.0019
15000	15	0.0015	0.0015	0.0018	0.0017
15000	20	0.0014	0.0014	0.0017	0.0015
15000	25	0.0014	0.0014	0.0018	0.0016
15000	30	0.0014	0.0014	0.0015	0.0015
15000	35	0.0014	0.0013	0.0016	0.0015
15000	40	0.0014	0.0014	0.0016	0.0016
20000	10	0.0018	0.0016	0.0019	0.0017
20000	15	0.0014	0.0013	0.0015	0.0015
20000	20	0.0012	0.0013	0.0015	0.0015
20000	25	0.0012	0.0013	0.0013	0.0014
20000	30	0.0013	0.0013	0.0014	0.0015
20000	35	0.0012	0.0013	0.0013	0.0015
20000	40	0.0012	0.0012	0.0013	0.0013
25000	10	0.0016	0.0015	0.0016	0.0016
25000	15	0.0013	0.0013	0.0014	0.0014
25000	20	0.0012	0.0013	0.0013	0.0014
25000	25	0.0012	0.0012	0.0014	0.0014
25000	30	0.0012	0.0013	0.0013	0.0015
25000	35	0.0012	0.0012	0.0014	0.0014
25000	40	0.0012	0.0013	0.0014	0.0014

Tab. 4.4: Simulations based on Samples from a Gamma Distribution with parameters α and β, assuming tilt function $\boldsymbol{h}(x) = (x, \log(x))'$

N	α	β	$MSE_{DRE_{2.5}}$	$MSE_{DRE_{97.5}}$	$MSE_{E_{2.5}}$	$MSE_{E_{97.5}}$
600	1	2	0.001	1.367	0.001	2.040
600	3	2	0.092	2.170	0.116	2.655
600	5	2	0.235	2.898	0.279	3.644
600	1	4	< 0.001	0.297	< 0.001	0.366
600	3	4	0.030	0.495	0.032	0.601
600	5	4	0.065	0.642	0.078	0.979
600	7	4	0.104	0.898	0.122	1.073
600	1	8	< 0.001	0.083	< 0.001	0.095
600	3	8	0.005	0.149	0.005	0.167
600	5	8	0.021	0.240	0.024	0.254
600	7	8	0.030	0.255	0.037	0.304
1000	1	2	0.001	0.671	0.001	1.006
1000	3	2	0.053	0.930	0.058	1.287
1000	5	2	0.167	1.693	0.248	2.331
1000	1	4	< 0.001	0.174	< 0.001	0.232
1000	3	4	0.011	0.324	0.013	0.427
1000	5	4	0.046	0.339	0.049	0.467
1000	7	4	0.068	0.559	0.083	0.896
1000	1	8	< 0.001	0.066	< 0.001	0.070
1000	3	8	0.003	0.085	0.003	0.112
1000	5	8	0.013	0.116	0.015	0.138
1000	7	8	0.016	0.148	0.021	0.191

Tab. 4.5: Further simulations based on Samples from a Gamma Distribution with parameters α and β, assuming tilt function $\boldsymbol{h}(x) = (x, \log(x))'$

N	α	β	$MSE_{DRE_{2.5}}$	$MSE_{DRE_{97.5}}$	$MSE_{E_{2.5}}$	$MSE_{E_{97.5}}$
5000	1	2	< 0.001	0.133	< 0.001	0.211
5000	3	2	0.010	0.214	0.012	0.251
5000	5	2	0.035	0.251	0.041	0.423
5000	1	4	< 0.001	0.042	< 0.001	0.053
5000	3	4	0.003	0.070	0.003	0.087
5000	5	4	0.009	0.093	0.011	0.107
5000	7	4	0.014	0.113	0.018	0.148
5000	1	8	< 0.001	0.010	< 0.001	0.013
5000	3	8	0.001	0.021	0.001	0.022
5000	5	8	0.002	0.024	0.002	0.026
5000	7	8	0.004	0.021	0.004	0.034
10000	1	2	< 0.001	0.074	< 0.001	0.098
10000	3	2	0.004	0.110	0.005	0.161
10000	5	2	0.013	0.160	0.016	0.257
10000	1	4	< 0.001	0.019	< 0.001	0.025
10000	3	4	0.002	0.041	0.002	0.052
10000	5	4	0.004	0.037	0.005	0.047
10000	7	4	0.008	0.051	0.009	0.071
10000	1	8	< 0.001	0.005	< 0.001	0.006
10000	3	8	< 0.001	0.008	< 0.001	0.010
10000	5	8	0.001	0.012	0.001	0.014
10000	7	8	0.002	0.014	0.002	0.016
15000	1	2	< 0.001	0.047	< 0.001	0.063
15000	3	2	0.004	0.070	0.005	0.084
15000	5	2	0.011	0.107	0.013	0.161
15000	1	4	< 0.001	0.015	< 0.001	0.017
15000	3	4	0.001	0.019	0.001	0.023
15000	5	4	0.003	0.036	0.004	0.045
15000	7	4	0.005	0.034	0.006	0.048
15000	1	8	< 0.001	0.004	< 0.001	0.005
15000	3	8	< 0.001	0.006	< 0.001	0.007
15000	5	8	0.001	0.008	0.001	0.009
15000	7	8	0.001	0.010	0.001	0.011

Tab. 4.6: Further simulations based on Samples from a Gamma Distribution with parameters α and β, assuming tilt function $\boldsymbol{h}(x) = (x, \log(x))'$

N	α	β	$MSE_{DRE_{2.5}}$	$MSE_{DRE_{97.5}}$	$MSE_{E_{2.5}}$	$MSE_{E_{97.5}}$
20000	1	2	< 0.001	0.047	< 0.001	0.065
20000	3	2	0.003	0.055	0.003	0.079
20000	5	2	0.007	0.088	0.009	0.123
20000	1	4	< 0.001	0.010	< 0.001	0.011
20000	3	4	0.001	0.023	0.001	0.026
20000	5	4	0.001	0.020	0.002	0.024
20000	7	4	0.004	0.024	0.004	0.033
20000	1	8	< 0.001	0.002	< 0.001	0.002
20000	3	8	< 0.001	0.004	< 0.001	0.005
20000	5	8	0.001	0.006	0.001	0.007
20000	7	8	0.001	0.007	0.001	0.008
25000	1	2	< 0.001	0.027	< 0.001	0.035
25000	3	2	0.002	0.051	0.003	0.068
25000	5	2	0.007	0.082	0.009	0.133
25000	1	4	< 0.001	0.006	< 0.001	0.007
25000	3	4	0.001	0.013	0.001	0.018
25000	5	4	0.001	0.016	0.002	0.026
25000	7	4	0.002	0.018	0.003	0.024
25000	1	8	< 0.001	0.002	< 0.001	0.002
25000	3	8	< 0.001	0.003	< 0.001	0.004
25000	5	8	< 0.001	0.005	0.001	0.005
25000	7	8	0.001	0.007	0.001	0.009

Tab. 4.7: Simulations based on Samples from a Weibull Distribution with parameters α and β, assuming tilt function $\boldsymbol{h}(x) = (x^{\tau}, \log(x))'$

N	α	β	$MSE_{DRE_{2.5}}$	$MSE_{DRE_{97.5}}$	$MSE_{E_{2.5}}$	$MSE_{E_{97.5}}$
600	1	1	0.004	6.345	0.005	6.585
600	1.2	1	0.009	2.726	0.011	3.660
600	1.5	1	0.016	0.955	0.021	1.134
600	2.5	1	0.052	0.140	0.063	0.201
600	1	1.2	0.005	6.388	0.005	7.254
600	1.2	1.2	0.015	2.450	0.016	2.679
600	1.5	1.2	0.024	1.150	0.029	1.790
600	2.5	1.2	0.079	0.212	0.099	0.266
600	1	1.5	0.010	10.454	0.011	10.995
600	1.2	1.5	0.022	4.000	0.028	5.585
600	1.5	1.5	0.049	2.727	0.064	2.945
600	2.5	1.5	0.087	0.364	0.118	0.491
600	1	2.5	0.033	29.605	0.035	39.361
600	1.2	2.5	0.068	12.825	0.072	16.565
600	1.5	2.5	0.124	5.034	0.142	8.541
600	2.5	2.5	0.274	1.023	0.370	1.202
1000	1	1	0.003	2.524	0.003	3.272
1000	1.2	1	0.006	1.417	0.008	1.653
1000	1.5	1	0.010	0.551	0.013	0.722
1000	2.5	1	0.021	0.079	0.033	0.119
1000	1	1.2	0.004	3.900	0.005	5.232
1000	1.2	1.2	0.009	1.838	0.009	2.516
1000	1.5	1.2	0.015	0.727	0.019	1.008
1000	2.5	1.2	0.034	0.125	0.041	0.129
1000	1	1.5	0.006	7.696	0.006	9.494
1000	1.2	1.5	0.015	2.571	0.017	3.905
1000	1.5	1.5	0.025	1.546	0.033	1.871
1000	2.5	1.5	0.068	0.181	0.084	0.232
1000	1	2.5	0.012	15.829	0.012	22.411
1000	1.2	2.5	0.041	8.823	0.048	12.399
1000	1.5	2.5	0.081	3.573	0.101	4.808
1000	2.5	2.5	0.198	0.662	0.236	0.755

Tab. 4.8: Further Simulations based on Samples from a Weibull Distribution with parameters α and β, assuming tilt function $\boldsymbol{h}(x) = (x^{\tau}, \log(x))'$

N	α	β	$MSE_{DRE_{2.5}}$	$MSE_{DRE_{97.5}}$	$MSE_{E_{2.5}}$	$MSE_{E_{97.5}}$
5000	1	1	<0.001	0.459	<0.001	0.559
5000	1.2	1	0.001	0.256	0.001	0.353
5000	1.5	1	0.003	0.097	0.003	0.122
5000	2.5	1	0.005	0.021	0.007	0.025
5000	1	1.2	0.001	0.931	0.001	1.274
5000	1.2	1.2	0.002	0.452	0.002	0.722
5000	1.5	1.2	0.004	0.184	0.004	0.253
5000	2.5	1.2	0.008	0.029	0.009	0.036
5000	1	1.5	0.002	1.319	0.002	2.066
5000	1.2	1.5	0.002	0.452	0.002	0.722
5000	1.5	1.5	0.004	0.280	0.004	0.351
5000	2.5	1.5	0.011	0.035	0.016	0.037
5000	1	2.5	0.003	0.626	0.003	0.759
5000	1.2	2.5	0.007	1.608	0.008	1.826
5000	1.5	2.5	0.010	0.585	0.016	0.693
5000	2.5	2.5	0.045	0.116	0.054	0.169

4.3 The Impact of Welfare Reform

4.3.1 Data

We obtained data of AFDC (1982-1996) monthly caseloads and annual discontinuances in caseloads due to employment from the *Quarterly Public Assistance Statistics* [157]. The TANF data (1997-2009) was obtained from the HHS website as well as the each of program's annual reports to Congress [25–33]. Because the data consisted of thousands of caseloads and discontinuances for each each state, we changed the units of our data to thousands of caseloads to make our estimation easier. Additionally, we divided the annual discontinuance by 12 to represent average monthly discontinuances to have both covariates have the same units of measurement. Across all 53 territories, AFDC programs had slightly more than 75,000 caseloads with approximately 7,500 discontinuances due to employment over

Tab. 4.9: Further simulations based on Samples from a Weibull Distribution with parameters α and β, assuming tilt function $\boldsymbol{h}(x) = (x^{\tau}, \log(x))'$

N	α	β	$MSE_{DRE_{2.5}}$	$MSE_{DRE_{97.5}}$	$MSE_{E_{2.5}}$	$MSE_{E_{97.5}}$
10000	1	1	< 0.001	0.319	< 0.001	0.392
10000	1.2	1	< 0.001	0.124	0.001	0.154
10000	1.5	1	0.001	0.059	0.001	0.063
10000	2.5	1	0.003	0.008	0.003	0.009
10000	1	1.2	0.002	3.210	0.002	4.069
10000	1.2	1.2	0.001	0.186	0.001	0.267
10000	1.5	1.2	0.002	0.092	0.002	0.113
10000	2.5	1.2	0.004	0.014	0.005	0.020
10000	1	1.5	0.001	0.634	0.001	0.761
10000	1.2	1.5	0.001	0.339	0.001	0.456
10000	1.5	1.5	0.003	0.095	0.003	0.152
10000	2.5	1.5	0.005	0.024	0.006	0.029
10000	1	2.5	0.002	1.924	0.002	2.693
10000	1.2	2.5	0.003	0.733	0.004	1.172
10000	1.5	2.5	0.006	0.276	0.008	0.361
10000	2.5	2.5	0.016	0.071	0.018	0.090
15000	1	1	< 0.001	0.222	< 0.001	0.284
15000	1.2	1	< 0.001	0.106	< 0.001	0.134
15000	1.5	1	0.001	0.034	0.001	0.043
15000	2.5	1	0.002	0.009	0.003	0.010
15000	1	1.2	< 0.001	0.282	< 0.001	0.476
15000	1.2	1.2	< 0.001	0.108	0.001	0.154
15000	1.5	1.2	0.001	0.049	0.001	0.077
15000	2.5	1.2	0.003	0.009	0.003	0.012
15000	1	1.5	< 0.001	0.377	< 0.001	0.515
15000	1.2	1.5	0.001	0.196	0.001	0.265
15000	1.5	1.5	0.002	0.082	0.002	0.108
15000	2.5	1.5	0.004	0.014	0.005	0.020
15000	1	2.5	0.001	1.260	0.001	1.896
15000	1.2	2.5	0.002	0.432	0.002	0.633
15000	1.5	2.5	0.005	0.197	0.006	0.304
15000	2.5	2.5	0.009	0.039	0.011	0.075

Tab. 4.10: Further simulations based on Samples from a Weibull Distribution with parameters α and β, assuming tilt function $\boldsymbol{h}(x) = (x^\tau, \log(x))'$

N	α	β	$MSE_{DRE_{2.5}}$	$MSE_{DRE_{97.5}}$	$MSE_{E_{2.5}}$	$MSE_{E_{97.5}}$
20000	1	1	< 0.001	0.175	< 0.001	0.201
20000	1.2	1	< 0.001	0.059	< 0.001	0.075
20000	1.5	1	0.001	0.024	0.001	0.032
20000	2.5	1	0.001	0.006	0.001	0.007
20000	1	1.2	< 0.001	0.307	< 0.001	0.348
20000	1.2	1.2	0.001	0.100	0.001	0.124
20000	1.5	1.2	0.001	0.039	0.001	0.061
20000	2.5	1.2	0.002	0.009	0.003	0.011
20000	1	1.5	< 0.001	0.307	< 0.001	0.442
20000	1.2	1.5	0.001	0.129	0.001	0.186
20000	1.5	1.5	0.001	0.061	0.001	0.072
20000	2.5	1.5	0.003	0.014	0.003	0.015
20000	1	2.5	0.001	0.861	0.001	1.124
20000	1.2	2.5	0.002	0.377	0.002	0.505
20000	1.5	2.5	0.004	0.191	0.004	0.265
20000	2.5	2.5	0.008	0.030	0.009	0.041
25000	1	1	< 0.001	0.125	< 0.001	0.150
25000	1.2	1	< 0.001	0.045	< 0.001	0.065
25000	1.5	1	0.001	0.019	0.001	0.025
25000	2.5	1	0.001	0.004	0.001	0.005
25000	1	1.2	< 0.001	0.146	< 0.001	0.208
25000	1.2	1.2	< 0.001	0.089	< 0.001	0.118
25000	1.5	1.2	0.001	0.030	0.001	0.035
25000	2.5	1.2	0.001	0.007	0.002	0.008
25000	1	1.5	< 0.001	0.170	< 0.001	0.277
25000	1.2	1.5	< 0.001	0.103	< 0.001	0.137
25000	1.5	1.5	0.001	0.045	0.001	0.063
25000	2.5	1.5	0.003	0.009	0.003	0.010
25000	1	2.5	0.001	0.946	0.001	1.090
25000	1.2	2.5	0.001	0.403	0.001	0.482
25000	1.5	2.5	0.003	0.118	0.004	0.181
25000	2.5	2.5	0.007	0.027	0.009	0.044

a fifteen year time horizon. TANF, on the other hand, had over 40,000 caseloads with over 7,000 discontinuances due to employment over the course of thirteen years.

4.3.2 Analysis of the PROWA Using Bayesian Lagged Regression

Using our data, we regressed our dependent variables, y_t discontinuances in caseloads due to employment at year t, against our explanatory variable, total caseloads in a previous year t', $x_{t'}$. We estimated ten instances of the model defined in equation 3. In particular, we estimated separate regressions regarding the impact of each of the five previous year's caseloads ($t-1$, $t-2$, $t-3$, $t-4$, and $t-5$) on discontinuances due to employment for [1] AFDC programs and for [2] TANF programs as TANF had instituted a five year time limit.[2]

We used WinBUGS to run a Bayesian MCMC sampler over the parameter space of the hierarchical model outlined in equations (4.1) to (4.8) to sample the posterior distribution [50, 97]. We ran our MCMC sampler for 30,000 draws, using the first half for "burn-in" to dissipate initial conditions. In the above model, Φ_0 and ν_0 of equation 6 were randomly generated. We ran our simulations several times over different random choices of Φ_0 and ν_0 and found similar results. These results suggested that the random generation of Φ_0 and ν_0 had no impact on the results generated by our MCMC sampler. Additionally, our posterior samples overall had a low degree of autocorrelation, indicating that we had likely not oversampled particular regions within our density.

After completion of the MCMC sampler, we estimated posterior means, standard deviations, and 95% credible intervals of all posterior samples.[3] Since we had truncated the MCMC sampler after 30,000 draws, however, there is still some con-

[2] Concatenating the explanatory variables across multiple lags for a larger multiple regression analysis resulted in multiciollinearity issues due to correlations between lagged caseloads and hence was not performed as part of this analysis.

[3] A 95% credible interval for a univariate sample is a Bayesian interval estimator and denotes an interval bounded by the 2.5th and 97.5th percentiles of the posterior sample, serving as an estimate for the interval bounded by the 2.5th and 97.5th quantiles of the distribution.

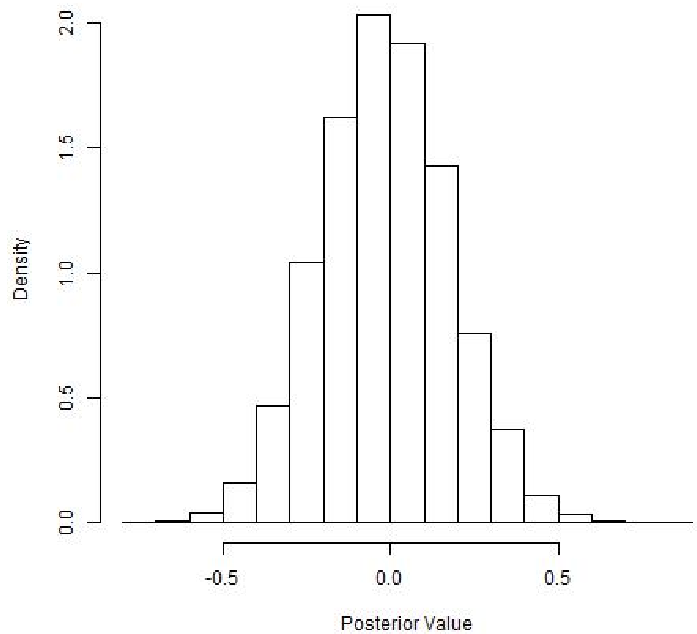

Fig. 4.1: Histogram of Posterior Sample for Posterior Intercept Mean Coefficient (μ_0), TANF regression, Lag 1

cern about missing information in the posterior sample. We therefore used the DRE method to improve upon our posterior sample. The posterior densities for μ_0 and μ_1 were symmetric and unimodal and were therefore appropriate for analysis using the tilt function $\boldsymbol{h}(x) = (x, x^2)'$ described in (4.12). The posterior densities of σ_0^2 and σ_1^2, however, appeared to be slightly skewed and we therefore more appropriate for analysis using the tilt function $\boldsymbol{h}(x) = (x, \log(x))'$ described in (4.13). The plots for a few selected coefficients are depicted in Figures 4.1-4.4 illustrating the reason for our choices of tilt functions:

For our posterior samples of μ_0 and μ_1, we took each sample of the marginal posterior (sample size 15,000), used as a reference density a random sample of equal size from a uniform distribution over the posterior sample's range, and applied the DRE method to estimate the cumulative distribution function (CDF) of the marginal posterior for this new sample of size 30,000. We then equated the CDF to

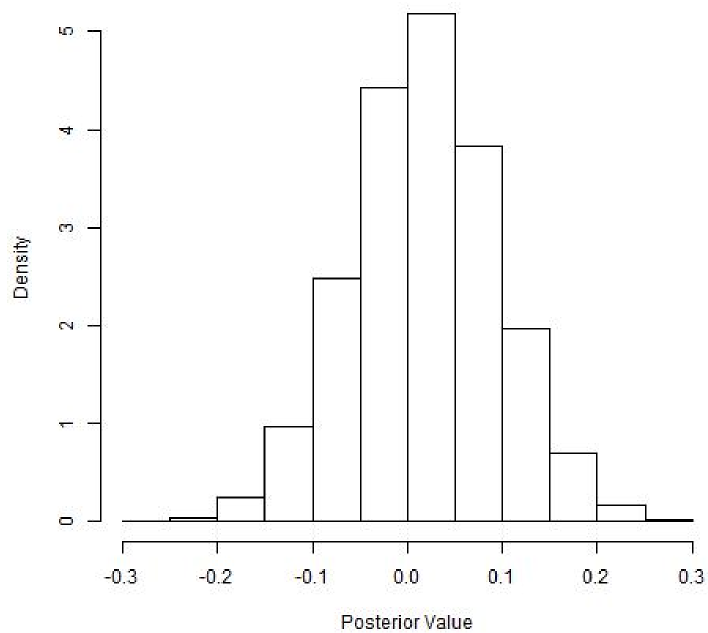

Fig. 4.2: Histogram of Posterior Slope Mean Coefficient (μ_1), TANF regression, Lag 1

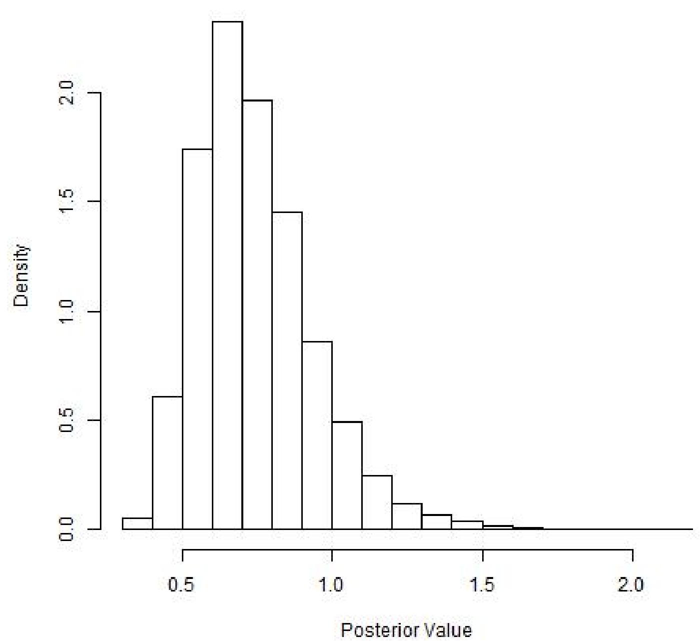

Fig. 4.3: Histogram of Posterior Intercept Variance Coefficient (σ_0^2), TANF regression, Lag 1

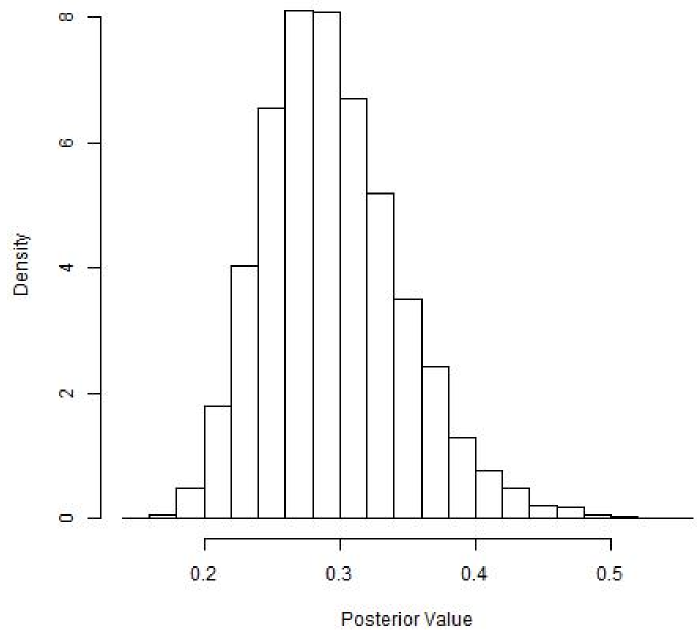

Fig. 4.4: Posterior Slope Variance Coefficient, TANF regression (σ_1^2), Lag 1

0.025 and solved the resulting equation to generate a lower limit for the 95% credible interval. We took a similar approach, equating the CDF to 0.975, to generate an upper estimate for the credible interval. We refer to this newly estimated region, based on semiparametric out of sample fusion, as the posterior density's 95% credible interval. Our results are outlined in Tables 4.11-4.30.

Tab. 4.11: AFDC Regression of Discontinuances as a Function of Caseloads - One Year Time Lag

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
μ_0	0.072	0.173	-0.263	0.417
μ_1	0.009	0.075	-0.136	0.155
σ_0^2	0.773	0.188	0.486	1.227
σ_1^2	0.290	0.049	0.208	0.400

Tab. 4.12: AFDC Regression of Discontinuances as a Function of Caseloads, Interval Estimates and Refinements - One Year Time Lag

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
μ_0	-0.263	0.417	-0.264	0.411
μ_1	-0.136	0.155	-0.136	0.156
σ_0^2	0.486	1.227	0.471	1.192
σ_1^2	0.208	0.400	0.204	0.396

Tab. 4.13: TANF Regression of Discontinuances as a Function of Caseloads - One Year Time Lag

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
μ_0	0.068	0.151	-0.228	0.365
μ_1	0.014	0.075	-0.132	0.163
σ_0^2	0.630	0.142	0.408	0.961
σ_1^2	0.292	0.050	0.209	0.406

Tab. 4.14: TANF Regression of Discontinuances as a Function of Caseloads, Interval Estimates and Refinements - One Year Time Lag

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
μ_0	-0.228	0.365	-0.228	0.364
μ_1	-0.132	0.163	-0.135	0.162
σ_0^2	0.408	0.961	0.398	0.947
σ_1^2	0.209	0.406	0.206	0.401

Tab. 4.15: AFDC Regression of Discontinuances as a Function of Caseloads - Two Year Time Lag

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
μ_0	0.272	0.183	-0.094	0.624
μ_1	0.001	0.074	-0.146	0.149
σ_0^2	0.761	0.186	0.479	1.196
σ_1^2	0.290	0.050	0.208	0.405

Tab. 4.16: AFDC Regression of Discontinuances as a Function of Caseloads, Interval Estimates and Refinements - Two Year Time Lag

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
μ_0	-0.094	0.624	-0.088	0.628
μ_1	-0.146	0.149	-0.145	0.148
σ_0^2	0.479	1.196	0.462	1.172
σ_1^2	0.208	0.405	0.204	0.399

Tab. 4.17: TANF Regression of Discontinuances as a Function of Caseloads - Two Year Time Lag

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
μ_0	0.240	0.151	-0.059	0.535
μ_1	0.000	0.076	-0.148	0.150
σ_0^2	0.578	0.125	0.380	0.870
σ_1^2	0.292	0.050	0.210	0.404

Tab. 4.18: TANF Regression of Discontinuances as a Function of Caseloads, Interval Estimates and Refinements - Two Year Time Lag

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
μ_0	-0.059	0.535	-0.059	0.534
μ_1	-0.148	0.150	-0.149	0.150
σ_0^2	0.380	0.870	0.372	0.854
σ_1^2	0.210	0.404	0.207	0.399

4.3.3 *Statistical Inferences Based on Hierarchical Bayesian Analysis of PROWA*

Although our overall coefficient estimates are small in magnitude, the reader should be reminded of the fact that the data set analyzed consisted of observations in units of thousands. Therefore, slight differences between the coefficients signify notable differences. For the first year time lag model, the marginal posterior mean for μ_1 is quite higher in Table 4.11 than in Table 4.13. A two sample Kolmogrov Smirnov test of the sample of the two posterior samples rejects the null hypothesis of distributional equality at $p < 0.001$, suggesting that there is a marked difference

Tab. 4.19: AFDC Regression of Discontinuances as a Function of Caseloads - Three Year Time Lag

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
μ_0	0.196	0.196	-0.184	0.582
μ_1	0.000	0.076	-0.147	0.150
σ_0^2	0.765	0.201	0.465	1.237
σ_1^2	0.292	0.050	0.210	0.406

Tab. 4.20: AFDC Regression of Discontinuances as a Function of Caseloads, Interval Estimates and Refinements - Three Year Time Lag

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
μ_0	-0.184	0.582	-0.189	0.583
μ_1	-0.147	0.150	-0.148	0.150
σ_0^2	0.465	1.237	0.448	1.212
σ_1^2	0.210	0.406	0.206	0.401

Tab. 4.21: TANF Regression of Discontinuances as a Function of Caseloads - Three Year Time Lag

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
μ_0	0.390	0.167	0.069	0.719
μ_1	-0.005	0.075	-0.153	0.143
σ_0^2	0.716	0.172	0.449	1.121
σ_1^2	0.291	0.050	0.208	0.405

Tab. 4.22: TANF Regression of Discontinuances as a Function of Caseloads, Interval Estimates and Refinements - Three Year Time Lag

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
μ_0	0.069	0.719	0.065	0.720
μ_1	-0.153	0.143	-0.154	0.143
σ_0^2	0.449	1.121	0.437	1.095
σ_1^2	0.208	0.405	0.204	0.399

Tab. 4.23: AFDC Regression of Discontinuances as a Function of Caseloads - Four Year Time Lag

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
μ_0	0.149	0.212	-0.26	0.563
μ_1	0.003	0.076	-0.147	0.153
σ_0^2	0.722	0.175	0.453	1.13
σ_1^2	0.292	0.05	0.211	0.407

Tab. 4.24: AFDC Regression of Discontinuances as a Function of Caseloads, Interval Estimates and Refinements - Four Year Time Lag

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
μ_0	-0.26	0.563	-0.265	0.562
μ_1	-0.147	0.153	-0.146	0.152
σ_0^2	0.453	1.13	0.441	1.108
σ_1^2	0.211	0.407	0.206	0.401

Tab. 4.25: TANF Regression of Discontinuances as a Function of Caseloads - Four Year Time Lag

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
μ_0	0.44	0.171	0.111	0.788
μ_1	-0.008	0.075	-0.156	0.141
σ_0^2	0.821	0.194	0.513	1.266
σ_1^2	0.293	0.051	0.21	0.407

Tab. 4.26: TANF Regression of Discontinuances as a Function of Caseloads, Interval Estimates and Refinements - Four Year Time Lag

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
μ_0	0.111	0.788	0.106	0.782
μ_1	-0.156	0.141	-0.156	0.141
σ_0^2	0.513	1.266	0.5	1.246
σ_1^2	0.21	0.407	0.206	0.403

Tab. 4.27: AFDC Regression of Discontinuances as a Function of Caseloads - Five Year Time Lag

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
μ_0	0.269	0.241	-0.203	0.742
μ_1	-0.002	0.076	-0.15	0.145
σ_0^2	0.822	0.213	0.501	1.325
σ_1^2	0.292	0.051	0.21	0.407

Tab. 4.28: AFDC Regression of Discontinuances as a Function of Caseloads, Interval Estimates and Refinements - Five Year Time Lag

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
μ_0	-0.203	0.742	-0.207	0.74
μ_1	-0.15	0.145	-0.15	0.145
σ_0^2	0.501	1.325	0.483	1.296
σ_1^2	0.21	0.407	0.206	0.402

Tab. 4.29: TANF Regression of Discontinuances as a Function of Caseloads - Five Year Time Lag

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
μ_0	0.541	0.175	0.205	0.889
μ_1	-0.013	0.075	-0.159	0.138
σ_0^2	0.744	0.191	0.458	1.19
σ_1^2	0.292	0.05	0.21	0.404

Tab. 4.30: TANF Regression of Discontinuances as a Function of Caseloads, Interval Estimates and Refinements - Five Year Time Lag

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
μ_0	0.205	0.889	0.198	0.886
μ_1	-0.159	0.138	-0.161	0.136
σ_0^2	0.458	1.19	0.444	1.171
σ_1^2	0.21	0.404	0.207	0.401

between the two posterior distributions.[4]

Overall, these result suggests that the previous year's caseloads for TANF had a notably stronger impact in getting people off the welfare rolls and back to work than the previous year's caseloads for AFDC programs did. The other models with the longer lagged covariates suggest that total caseloads had either a negligible or even a negative effect on employment-based discontinuances.

There are also significant differences in using DRE to refine the 95% credible intervals. For example, the interval estimates regarding μ_1 are altered by a factor of several thousand discontinuances per caseload, depending on the year. Additionally, the posterior distributions of the variances are also altered substantially. Regardless, however, the posterior estimates of σ_0^2 and σ_1^2 all remain relatively small, suggesting that there is not much uncertainty pertaining to the relationship between caseloads and discontinuances.

An additional fully frequentist statistical analysis is included in this chapter's Appendix. This analysis also finds similar results to those above.

4.4 Additional Analysis

In addition to the analysis above, we conducted further analysis to examine the efficacy of transforming AFDC to TANF. As our data of caseloads and discontinuances spanned from 1982-2009, we conducted another rigorous Bayesian analysis over the entire time horizon, with categorically-coded coefficients representing time in terms of year. Comparisons of these coefficient estimates from before the PROWA

[4] As mentioned earlier, autocorrelations of the sample were quite low, suggesting that independence of the sample, a sufficient condition for being able to apply the Kolmogorov Smirnov test, was not an unreasonable assumption.

to after the PROWA can enable us to understand the efficacy of the reforms themselves.

Again, following the notation corresponding to our previous model, we can again define this new hierarchical Bayesian model over $i \in \{1, \ldots, 53\}$ territories (50 states as well as Guam, Puerto Rico, and the District of Columbia) (units) over $t \in \{1982, \ldots, 2008\}$ years as follows:

$$\boldsymbol{y_i} \sim MVN(\beta_{i,0} + \beta_{i,1}\boldsymbol{x_{i,1}} + \beta_t, \Sigma) \tag{4.25}$$

$$\beta_{i,0} \sim N(\mu_0, \sigma_0^2) \tag{4.26}$$

$$\beta_{i,1} \sim N(\mu_1, \sigma_1^2) \tag{4.27}$$

$$\beta_t \sim N(0, 10) \tag{4.28}$$

$$\Sigma^{-1} \sim \text{Wishart}(\Phi_0, \nu_0) \tag{4.29}$$

$$\mu_0 \sim N(0, 10) \tag{4.30}$$

$$\mu_1 \sim N(0, 10) \tag{4.31}$$

$$\sigma_0^{-2} \sim \Gamma(10, 10) \tag{4.32}$$

$$\sigma_1^{-2} \sim \Gamma(10, 10) \tag{4.33}$$

This model is quite similar to the model we examined earlier, but runs across the data set's entire time horizon and also contains a term β_t that quantifies the impact of time t' on caseload reduction at time t. A posterior examination of this coefficient before and after the introduction of TANF can enable us to understand the impact of the 1996 reform. Just as we did earlier, we defined $\beta_{1,0} = 0$ as well as β_{1982}=0 to ensure statistical identifiability of our model. We ran this Bayesian MCMC sampler for 30,000 iterations, using first half to dissipate our initial conditions and the remaining half for statistical inference. Our Bayesian posterior estimates are outlined in Tables 4.31-4.32. As our posterior samples of μ_0, μ_1, and β_t appeared to be symmetric and unimodal we again utilized our DRE approach to improve the estimation of our Bayesian credible intervals.

Tab. 4.31: Additional Analysis

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
μ_0	0.093	0.167	-0.231	0.428
μ_1	0.006	0.074	-0.139	0.150
β_{1983}	0.042	0.067	-0.091	0.174
β_{1984}	-0.008	0.062	-0.131	0.113
β_{1985}	-0.028	0.073	-0.175	0.114
β_{1986}	0.010	0.080	-0.146	0.165
β_{1987}	0.011	0.077	-0.142	0.160
β_{1988}	0.032	0.082	-0.130	0.191
β_{1989}	0.015	0.081	-0.146	0.175
β_{1990}	0.048	0.082	-0.116	0.206
β_{1991}	0.086	0.087	-0.091	0.256
β_{1992}	0.076	0.082	-0.091	0.232
β_{1993}	0.064	0.091	-0.120	0.240
β_{1994}	0.199	0.100	0.005	0.394
β_{1995}	0.189	0.086	0.019	0.356
β_{1996}	0.573	0.183	0.201	0.919
β_{1997}	0.512	0.158	0.191	0.811
β_{1998}	0.290	0.123	0.044	0.531
β_{1999}	0.294	0.111	0.074	0.511
β_{2000}	0.249	0.104	0.040	0.451
β_{2002}	0.272	0.115	0.048	0.498
β_{2003}	0.305	0.116	0.079	0.534
β_{2004}	0.294	0.115	0.069	0.518
β_{2005}	0.344	0.134	0.084	0.613
β_{2006}	0.382	0.162	0.060	0.702
β_{2007}	0.298	0.151	-0.003	0.589
β_{2008}	0.256	0.157	-0.059	0.568
σ_0^2	0.778	0.164	0.508	1.151
σ_1^2	0.288	0.049	0.207	0.397

We see a noticeable increase in our coefficient estimates pertaining to time after 1997, once the law had become fully enacted. In fact, the small posterior estimates of β_t from the 1980s and early 1990s suggest that AFDC programs at the

Tab. 4.32: Additional Analysis, Interval Estimates and Refinements

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Lower Limit	Refined 95% Credible Upper Limit
μ_0	-0.231	0.428	-0.234	0.427
μ_1	-0.139	0.150	-0.138	0.151
β_{1983}	-0.091	0.174	-0.091	0.173
β_{1984}	-0.131	0.113	-0.130	0.113
β_{1985}	-0.175	0.114	-0.173	0.116
β_{1986}	-0.146	0.165	-0.148	0.166
β_{1987}	-0.142	0.160	-0.140	0.161
β_{1988}	-0.130	0.191	-0.129	0.193
β_{1989}	-0.146	0.175	-0.145	0.175
β_{1990}	-0.116	0.206	-0.114	0.208
β_{1991}	-0.091	0.256	-0.089	0.256
β_{1992}	-0.091	0.232	-0.087	0.235
β_{1993}	-0.120	0.240	-0.117	0.241
β_{1994}	0.005	0.394	0.004	0.393
β_{1995}	0.019	0.356	0.019	0.357
β_{1996}	0.201	0.919	0.210	0.926
β_{1997}	0.191	0.811	0.198	0.818
β_{1998}	0.044	0.531	0.046	0.532
β_{1999}	0.074	0.511	0.074	0.513
β_{2000}	0.040	0.451	0.042	0.453
β_{2002}	0.048	0.498	0.047	0.496
β_{2003}	0.079	0.534	0.076	0.535
β_{2004}	0.069	0.518	0.069	0.518
β_{2005}	0.084	0.613	0.082	0.608
β_{2006}	0.060	0.702	0.061	0.700
β_{2007}	-0.003	0.589	0.001	0.590
β_{2008}	-0.059	0.568	-0.053	0.567
σ_0^2	0.508	1.151	0.500	1.134
σ_1^2	0.207	0.397	0.203	0.392

time were essentially just a poverty trap, lulling people into poverty and hindering upward mobility. These posterior estimates increase notably in the 1990s, especially after the law was introduced in 1996.

In Table 32, we present our Bayesian 95% credible intervals and subsequent refinements. We find that all 95% credible intervals regarding time before the 1996 welfare reform all had negative probability mass, whereas after the law had become fully enacted, they had no negative mass until β_{2007} and β_{2008}. After our posterior refinements, however, all were positive except for β_{2008}. This negativity was due to our MCMC sampler not properly summarizing the posterior. The higher point and interval estimates of these "post-welfare reform" coefficients, indicate that the law had indeed been successful in reducing dependence on government assistance.

4.4.1 Policy Implications

Our analysis clearly illustrates the great success of welfare reform. Specifically, our Bayesian models suggest that the previous year's caseloads had a significantly more notable impact on getting people off of the welfare rolls and back to work under the TANF program compared to the AFDC program, particularly within one year. Subsequent year lags had a negligible or even a detrimental impact on caseload reductions due to employment.

These general findings are substantiated by research produced by The Heritage Foundation, the Cato Institute, and the Brookings Institution. In particular, research from these well-known think tanks has clearly illustrated that as a result of welfare reform, caseloads declined, earnings increased, and poverty rates dropped [61, 128, 145, 152]. More rigorous work standards and time limits would almost surely improve on these results.

4.5 Conclusions and Future Research

Our study offers a number of significant contributions. Firstly, our work also sheds light on the efficacy of welfare reform. Although safety nets are an important component of society, they should be constructed in a manner to help provide temporary assistance to get people back to work and become contributing members of society. If these programs are just an open-ended government handout, recipients can become overly dependent on them and can subsequently become incapable of realizing the American dream. As Charles Murray's *Losing Ground* illustrated, these programs have the potential to trap generations upon generations of people in poverty [111]. Although the PROWA is a case study in successful welfare reform, the law reformed only one of approximately seventy federal welfare programs [135]. Policymakers should consider similar reforms for these other programs. More fundamentally, from a policymaker's perspective, it is always important to understand the efficacy of policy reforms and make effective policy recommendations. Future research should similarly examine other programs including Social Security, Medicare, Medicaid, public housing, and the Affordable Care Act among others.

Methodologically, we have improved upon Bayesian credible interval estimation by applying semiparametric density ratio estimation to truncated posterior samples. Although this semiparametric modeling technique has had a number of applications in frequentist models, we are the first to apply the approach to Bayesian estimation. These methodological improvements of course need not only apply to Bayesian models pertaining to public policy research and are worthwhile tools to be applied in many other applied settings.

4.6 Appendix

4.6.1 An Elementary Frequentist Analysis

As an additional test of our work, we also performed a frequentist analysis examining the impact of AFDC/TANF caseloads on discontinuances due to employment. We estimated essentially the same model we estimated earlier, but instead made the (rather restrictive) assumption of homogeneity amongst the coefficients. Specifically, we estimated the simple linear model:

$$y_{i,t} = \beta_0 + \beta_1 x_{i,t} + \epsilon_{i,t}, \tag{4.34}$$

where $y_{i,t}$ is the number of employment related discontinuances for state/territory i at time t and $x_{i,t'}$ is the number of caseloads at time t', where $t' < t$ is a lagged time. We assumed $\epsilon_{i,t} \sim N(0, 1)$. Our results are presented in Table 4.33.

In all five lagged comparisons, we see that the slope coefficient for TANF programs is higher than the slope coefficient for AFDC programs, suggesting that TANF was more successful than AFDC programs in getting people back to work.

Tab. 4.33: Frequentist Statistical Analysis

	Coefficient	Estimate	Standard Error	t-stat	P-value	Adjusted R^2
AFDC Lag 1	$\hat{\beta}_0$	0.167	0.023	7.427	< 0.001	0.524
	$\hat{\beta}_1$	0.005	0.000	28.807	< 0.001	
TANF Lag 1	$\hat{\beta}_0$	0.184	0.033	5.653	< 0.001	0.561
	$\hat{\beta}_1$	0.011	0.000	28.722	< 0.001	
AFDC Lag 2	$\hat{\beta}_0$	0.172	0.024	7.141	< 0.001	0.513
	$\hat{\beta}_1$	0.005	0.000	27.147	< 0.001	
TANF Lag 2	$\hat{\beta}_0$	0.142	0.030	4.785	< 0.001	0.636
	$\hat{\beta}_1$	0.011	0.000	32.122	< 0.001	
AFDC Lag 3	$\hat{\beta}_0$	0.177	0.026	6.878	0.144	0.495
	$\hat{\beta}_1$	0.005	0.000	25.186	< 0.001	
TANF Lag 3	$\hat{\beta}_0$	0.098	0.028	3.492	0.002	0.690
	$\hat{\beta}_1$	0.011	0.000	34.544	< 0.001	
AFDC Lag 4	$\hat{\beta}_0$	0.185	0.028	6.605	< 0.001	0.478
	$\hat{\beta}_1$	0.005	0.000	23.380	< 0.001	
TANF Lag 4	$\hat{\beta}_0$	0.086	0.030	2.859	0.004	0.684
	$\hat{\beta}_1$	0.011	0.000	32.678	< 0.001	
AFDC Lag 5	$\hat{\beta}_0$	0.197	0.031	6.407	< 0.001	0.457
	$\hat{\beta}_1$	0.005	0.000	21.314	< 0.001	
TANF Lag 5	$\hat{\beta}_0$	0.079	0.033	2.373	0.018	0.670
	$\hat{\beta}_1$	0.011	0.000	29.507	< 0.001	

To ensure that these differences were not insignificant, we tested the null hypothesis H_0 for each lag that the slope coefficient for TANF programs, β_1, was equal to the slope coefficient for AFDC programs for that same time lag. For example, for TANF regression results for lag 1, we tested the null hypothesis H_0 that $\beta_0 = 0.005$ against the two-sided alternative H_a that $\beta_0 \neq 0.005$. Our results are outlined in Table 4.34 for all five lags:

Tab. 4.34: Statistical Significance of Slope Coefficients

Lag	t-stat	p-value
Lag 1	15.890	< 0.001
Lag 2	19.667	< 0.001
Lag 3	19.333	< 0.001
Lag 4	18.667	< 0.001
Lag 5	15.714	< 0.001

These results indicate a statistically significant difference between the slope coefficients of the lagged TANF models compared to the coefficients corresponding to the analogous AFDC models. These results lend further credence to the argument that TANF programs were more successful than AFDC programs in getting people off the welfare rolls and back to work.

Chapter 5: Generalized Bayesian Inferences for Counterterrorism Policy with Improved Credible Interval Estimation via Semi-parametric Out of Sample Fusion

5.1 Introduction

5.1.1 Combating Terrorism

From the Siccari during the first century siege of Jerusalem, to the Assassins during the 11th century in what is today's Middle East, to the Thugs in India during the 1400s to 1800s, to the Wall Street bombings of the early 1920s, terrorism has been an issue endangering civilians for generations [69, 126]. Today, myriads of questions abound regarding counterterrorsim policy throughout the world including questions pertaining to the U.S. military operations in Iraq and Afghanistan, the Troubles in Northern Ireland, the post-Suharto years in Indonesia, and the twenty-six year long civil war in Sri Lanka among others.

Amongst scholars in the field of international relations, terrorism is generally defined as "the deliberate use or threat of force against noncombatants by a non-state actor in pursuit of a political goal" [15]. Understanding how to fight terrorists has been an issue that policymakers throughout the world have debated and grappled with for years. Fighting terrorists is inherently different from many of the major conflicts of the twentieth century. During World War I and World War

II, for example, both sides had uniformed combatants representing nation-states at war. During the Cold War, both sides understood the concept of mutually assured destruction, and consequently did not want to be annihilated in a retaliatory strike.

Terrorists, however, are markedly different from traditional enemy combatants, as they generally blend in with civilians and deliberately target innocent men, women, and children, as well as military personnel. Many are typically brainwashed by the belief that a better life awaits them for slaughtering their enemies [38]. As we continue on into the twenty-first century, fighting terrorism remains an issue, and it is important to equip policymakers with the tools to be able to do so.

A Google Scholar search of the keyword "terrorism" yields nearly 893,000 studies on the topic. Many of these studies include research statistically examining the risk of terrorist attacks as well as looking at certain types of terrorist attacks in particular localities across the globe [40, 71, 77, 117, 120, 155]. In 1971, Hawkes conducted a cluster analysis of terrorism data [62]. Enders (2007) offers a comprehensive review of research on measuring terrorism, the efficacy of counterterroism policies, and the causes of terrorism and its various manifestations [35]. In the 1980s, Holden (1986, 1987) examined the "contagion effect" of American aircraft hijackings [65, 66]. Enders and Sandler (1995) examine terrorist behavior from game and choice theoretic perspectives [36]. Sandler and Arce (2003) discuss game theoretic analyses of terrorism and their various policy implications [144]. Li and Schaub (2005) conduct a time series analysis and examine the effect of economic globalization on terrorist attacks [95]. Kaplan et al (2005, 2006) look at the impact on different "counterterror tactics," on suicide bombings in Israel [75, 76]. Lewis et al (2012) statistically examine the changes over time in civilian deaths in Iraq as a result of terrorism [94]. In recent years, a number of researchers have utilized very sophisticated statistical modeling including negative binomial distribution models along with self-exiting hurdle models to examine the incidence as well as the probability of terrorist attacks [121, 162, 163]

Although these papers are interesting both mathematically and from a policy perspective, these studies have typically been restricted to simply just one conflict or one type of terrorist attack. Additionally, much of the heavily statistical research in this area provides limited advice to policymakers about how to fortify specific security measures to prevent various types of terrorist attacks. In this chapter, we rigorously analyze terrorist attack data from a number of major conflicts throughout the world. In particular, we utilize Bayesian logistic regression techniques to offer counterterrorism strategies in four major conflicts across the globe. Kyung et al (2011) also conducted an analysis of terrorist attack data; however, their statistical models were not constructed in a way to offer prescriptive policy advice [90]. We not only present a model capable of providing such advice but also improve on Bayesian credible interval estimation techniques in the process. Our study helps shed light on the factors that influence the success of terrorist attacks, providing policymakers with advice on how to more strategically target their security and resources to help them battle against this very dangerous enemy.

5.1.2 Dirichlet Process Priors and the Limitations Bayesian Parametric methods

With consistent improvements in statistical computing capabilities over the last several decades, the use of Bayesian methods has become increasingly common in applied research. Bayesian methods are an attractive approach for modeling real-world phenomena for a number of reasons. One reason is that the associated statistical inferences from such models are based conditionally on existing data rather than on the distributional properties of estimators or test statistics calculated over a long-run frequency of many imaginary unobserved samples. Additionally, Bayesian methods provide us with "exact-sample" results, rather than being rooted in the typical asymptotic theory that most frequentist statistical estimation methods generally assume. Thirdly, and perhaps most importantly, Bayesian methods enable us to tackle problems with high-dimensional parameter spaces that would generally be impossible to estimate from a purely frequentist perspective.

One of the most controversial aspects of Bayesian methods, however, is the formulation of a prior distribution. Typically, prior assumptions about a model's parameters are made subjectively by the researcher. Often these prior distributions belong to well-known parametric families. Normal and uniform distributions are for instance workhorse examples of priors in the applied Bayesian statistical literature.

A common criticism with these types of parametric prior distributions, however, is that since they are subjectively chosen by the researcher their distributional assumptions may therefore not necessarily model reality. The normal distribution, for example, has a limited degree of flexibility as it is unimodal, does not accommodate skewness, and has relatively thin tails. If a normally distributed prior is a misspecification to begin with, then misleading inferences can result. Other choices of parametric prior distributions have similar types of limitations.

Dirichlet Process priors (hereafter referred to as DP priors) do not have these restrictions as they allow researchers to weaken the restrictive assumptions generally concomitant with Bayesian statistical models [4, 41]. These prior distributions have been used in a number of settings including applied economics research, health policy research, and examining the incidence of illness among others [3, 72, 88]. DP priors avoid the typical strict parametric assumptions regarding heterogeneity mentioned above and instead utilize an unknown distribution G to model heterogeneity. As G is assumed to be random, a DP Prior can be placed on this distribution. Dirichlet Processes thus enable the researcher to place a probability distribution over a space of probability distributions.

Mathematically, we can describe a DP prior $G \sim DP(G|G_0, \alpha)$ consists of two "parameters:" G_0, a parametric baseline probability measure and a concentration parameter α. The baseline measure G_0 can be considered to be a prior assumption regarding the population distribution. The Dirichlet Process transforms this baseline measure into a discrete probability distribution with the concentration parameter α determining how close the non-parametric distribution G is to the baseline measure. Smaller values of α indicate greater departure from the baseline measure. In the limit, as $\alpha \to \infty$, $G \Rightarrow G_0$ (i.e. the Dirichlet Process converges in the measure to the baseline measure).

A commonly-used choice for the baseline measure is a normal distribution, which the Dirichlet Process discretizes. Since discrete probability measures have non-zero probabilities of observing identical values, they are often used for cluster analysis. For each particular sample from an MCMC simulation, identical posterior estimates of parameters believed to be a priori following a Dirichlet Process, are considered to belong to the same cluster of observations. Thus, a nice aspect of the Dirichlet Process is that it not only alleviates the strict parametric assumptions typically associated with parametric prior distributions, but it also alleviates the similarly restrictive assumptions associated with finite mixture models that a priori assume a particular number of segments [22, 74].

In this chapter, we employ DP priors to weaken the generally restrictive assumptions associated with commonly chosen parametric prior distributions for a Bayesian logistic regression model. In addition, we also offer some improvements to the credible interval estimates generated by our Bayesian estimation. We discuss the techniques for doing so in the next section.

5.1.3 A Brief Review of Density Ratio Estimation

Density ratio estimation (referred to as DRE throughout this dissertation) is a semiparametric modeling approach. As discussed in the previous chapter of this dissertation, this technique has seen myriads of applications ranging from AIDS research to mortality rate prediction to cluster detection among many others [43,52, 80,81,83]. In the univariate case, we typically assume that there are $I+1$ random samples $(x_{i1},\ldots,x_{in_i})$, having probability density functions g_i, such that:

$$x_{ij} \sim g_i, \qquad i=1,\ldots,I,I+1, j=1,\ldots,n_i. \tag{5.1}$$

This approach assumes that $g_{I+1} \equiv g$ defines a reference density having a known ratio between g_i and g. In many applications, this ratio is defined to be an exponential in terms of a vector-valued tilt function $\boldsymbol{h}(x)$:

$$\frac{g_i(x)}{g(x)} = e^{\alpha_i+\boldsymbol{\beta_i'}\boldsymbol{h}(x)}, \quad i=1,\ldots,I. \tag{5.2}$$

By assuming a particular ratio, the statistician can estimate the parameters α_i and $\boldsymbol{\beta_i}$ as well as the distributions of G_i and G (the CDFs of g_i and g) $\forall\, i=1,\ldots,I$ using the method of empirical likelihood. The methodology has been discussed in detail in the previous chapter of this dissertation.

5.1.4 Using DRE to Improving on Bayesian Estimation Results

Typically, one would estimate posterior functionals using standard Markov Chain Monte Carlo (MCMC) methods [50]. Researchers would use the resulting MCMC samples to generate a variety of statistical estimators such as posterior means, standard deviations, and credible interval estimates.

As discussed in Chapter 4, however, a common limitation of MCMC methods stems from computational feasibility. Most Bayesian MCMC samplers need to be truncated in "real-time." As a result, it is difficult to understand if we have properly navigated the posterior distribution. Although we have some diagnostic checks, such as autocorrelation of draws and convergence of chains [51], these tests are merely diagnostics that may shed light on the issue, but we never completely understand whether our MCMC sampler truly manifests our entire posterior distribution. Therefore, point and interval estimates based on MCMC samples may not necessarily be sufficiently accurate from which to make statistical inferences.

One can ameliorate this limitation by using the DRE Method [123–125, 160]. Although this semiparametric approach has been used in a variety of settings, this dissertation is the first to apply the methodology in Bayesian estimation. In the previous chapter of this dissertation, we applied the methodology to hierarchical Bayesian linear regression. In that chapter, we illustrated, via a series of numerical simulations, that the DRE method has the ability to provide more accurate quantiles of distributions. In this chapter, we extend our application of this methodology to a generalized lienar hierarchical Bayesian model. The interested reader is referred to the previous chapter for simulation results demonstrating the efficacy of this approach.

5.2 Problem Formulation

5.2.1 Model

Consider a data set of terrorist attacks recorded over the course of $t \in \{1, \ldots, T\}$ discrete time occasions. We define the binary outcome variable:

$$y_t = \begin{cases} 1 & \text{if the terrorist attack is successful at time } t \\ 0 & \text{otherwise,} \end{cases} \tag{5.3}$$

where $p_t = \text{Prob}(y_t = 1)$ is the probability that a terrorist attack is successful. We examined the success of terrorist attacks by estimating the following binary logistic regression model:

$$P(Y|\beta) = \prod_{t=1}^{T} \frac{e^{y_t(\beta_0+\beta_1 X_{t,1}+\boldsymbol{\beta_2 X_{t,2}}+\boldsymbol{\beta_{t,3} X_{t,3}})}}{1+e^{\beta_0+\beta_1 X_{t,1}+\boldsymbol{\beta_2 X_{t,2}}+\boldsymbol{\beta_{t,3} X_{t,3}}}}, \tag{5.4}$$

$$X_{t,1} = \begin{cases} 1 & \text{if the terrorist attack at time } t \text{ is a suicide attack} \\ 0 & \text{otherwise,} \end{cases} \tag{5.5}$$

$\boldsymbol{X_{t,2}} = (x_{t,2,1}, \ldots, x_{t,2,n_2})$ is a categorically coded explanatory variable denoting the type of terrorist attack (assassination, hijacking, bombing, etc.) Armed assaults, defined in the appendix, are used as a benchmark variable to ensure statistical identifiability of the model. Finally, $\boldsymbol{X_{t,3}} = (x_{t,3,1}, \ldots, x_{t,2,n_3})$ is a categorically coded explanatory variable denoting the target of the terrorist attack (airports/airlines, business, educational institution, food or water supply, government, etc.). For the analysis of the conflicts in Afghanistan, Iraq, Sri Lanka, and Northern Ireland after the 1998 Good Friday Agreement, airports/airlines were the benchmark variable. For analysis of the conflict of Northern Ireland before the 1998 Good Friday Agreement, abortion related targets were the benchmark variable. $\boldsymbol{\beta_2}$ and $\boldsymbol{\beta_{t,3}}$ are of course n_2 and Tn_3 dimensional vectors respectively.

We allow $\beta_0 \sim N(\mu_0, \sigma_0^2)$ and $\beta_1 \sim N(\mu_1, \sigma_1^2)$. Additionally, we allow $\boldsymbol{\beta_2} \sim N(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_{n_2 \times n_2}})$ and $\boldsymbol{\beta_{t,3}}$ to either follow a normally distributed prior or to follow a DP prior. Mathematically, our two choices for varying $\boldsymbol{\beta_{t,3}}$ are either $\boldsymbol{\beta_{t,3}} \sim N(\boldsymbol{\mu_3}, \boldsymbol{\Sigma_{Tn_3 \times Tn_3}})$ or $\boldsymbol{\beta_{t,3}} \sim G$, where $G \sim DP(\alpha, N(\boldsymbol{\mu_3}, \boldsymbol{\Sigma_{Tn_3 \times Tn_3}}))$. μ_0, μ_1, as well as each component of the vectors $\boldsymbol{\mu_2}$ and $\boldsymbol{\mu_3}$ all follow a normal distribution with mean 0 and variance 10. σ_0^2 and σ_1^2 are each specified to follow an inverse gamma distribution with shape and scale parameter each equal to 10. Additionally, for computational simplicity $\boldsymbol{\Sigma_{n_2 \times n_2}}$ and $\boldsymbol{\Sigma_{Tn_3 \times Tn_3}}$ are assumed to be diagonal matrices, with each component also drawn from an inverted gamma distribution with shape and scale parameter each equal to 10.

As discussed earlier, $\boldsymbol{\beta_{t,3}}$ to follow a DP prior enables us to weaken the strict parametric assumptions associated with a normal prior distributions as well as to cluster our analysis around the types of targets terrorists intend to attack.

5.2.2 Data

Our data sets were obtained from the START GTD Database, a database compiled by the University of Maryland, containing detailed information on over 113,000 terrorist attacks [54]. We performed a statistical analysis on four different conflicts - The War in Afghanistan, the War in Iraq, the Civil War in Sri Lanka, and the Civil War in Ireland. The dependent variable in our Bayesian logistic regression was whether or not the attack was successful, and our explanatory variables involved whether the attack was a suicide attack, the type of attack, and the targets of the attack. All of our explanatory variables were categorically coded. For each conflict, a few observations could not definitively provide information regarding the details of the attack (such as whether the attack was a suicide attack or whether the attack was successful) and hence was excluded from our analysis.

The START database defined the success of a terrorist attack as whether the type of attack actually took place. For example, a bombing/explosion is considered successful if the bomb involved actually detonated. More details are contained in the appendix to this chapter.

5.3 A Bayesian Analysis of Several Major Conflicts Across the Globe

5.3.1 Estimation

We rigorously examined four recent conflicts - The War in Afghanistan, the War in Iraq, the Civil War in Sri Lanka, and the Civil War in Northern Ireland. For each conflict, we estimated the two Bayesian logistic regression models outlined in Section 5.2, with one model assuming normally distributed priors for the coefficient corresponding to the target of the terrorist attack and other assuming the less restrictive DP priors. We estimated both models via MCMC methods over the course 30,000 iterations, using the first half for burn in and the remaining half for statistical inference. For the model assuming normal prior distributions, we ran our MCMC sampler in WinBUGS [97]. For the model assuming DP priors we ran our sampler using the R Package *DPPackage* [70] that used well-known algorithms for MCMC sampling from non-conjugate priors for DP prior models [37, 98, 112]. Autocorrelations of each marginal posterior sample were low, suggesting reasonable navigation around the posterior density. As mentioned earlier, however, autocorrelations are simply just a diagnostic check and cannot truly elicit whether the posterior density has been adequately sampled. As a result, in addition to our standard Bayesian analysis, we also present improvements to the Bayesian interval estimates of a few important posterior coefficients using DRE.

5.3.2 The War in Afghanistan

On September 11, 2001, foreign terrorists struck the United States in the most devastating attack on American soil since Pearl Harbor. A group of nineteen terrorists hijacked three U.S. passenger planes, crashing them into the World Trade

Center in New York and the Pentagon in Washington, D.C. A fourth hijacked plane, intended for the U.S. Capitol, crashed in rural Pennsylvania that same morning. A total of over 3,000 innocent Americans died in these attacks.

Shortly thereafter, overwhelming evidence made it quite apparent that the attacks were perpetrated by the terrorist group *al Qaeda*, headed at the time by Osama bin Laden. The Taliban in Afghanistan had been harboring bin Laden, was known for aiding and abetting *al Qaeda*, and was infamous for sponsoring terrorism [79]. After the Taliban refused to hand bin Laden over to American custody, the United States used military force to remove the Taliban from power, began efforts toward finding bin Laden, and helped install a democratic Afghan government that renounces terrorism [153]. Since military operations in Afghanistan began thirteen years ago, the United States military has maintained a consistent presence in Afghanistan, providing stability and support to help the relatively new Afghan government.

Unfortunately, however, terrorist attacks in Afghanistan have continued over the course of the last several years. Tables 5.1 and 5.2 present a Bayesian statistical analysis of these attacks, using the modeling approach outlined above. The data used spans from slightly after the September 11th attacks (when U.S. military operations in Afghanistan began) through December 2011 consisting of 2887 observations.

Our results are quite informative. In Afghanistan, both models indicate that suicide attacks, with posterior estimates around -1 in both models, were generally unsuccessful as were assassination attempts (with posterior mean coefficient estimate -3.147 in the normal model and -2.858 in the DP model), bombings (-0.541 in the normal model and -0.634 in the DP model), attacks on infrastructure (-0.176 in the normal model and -0.418 in the DP model), and hijackings (-1.931 in the normal model and -1.787 in the DP model.) These posterior estimates indicate the

Tab. 5.1: Afghanistan - Hierarchical Bayesian Model Using Normally Distributed Priors

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
Intercept	4.605	0.553	3.541	5.645
Suicide	-1.281	0.236	-1.741	-0.811
Assassination	-3.147	0.345	-3.841	-2.506
Bombing/Explosion	-0.541	0.327	-1.191	0.081
Facility/Infrastructure	-0.176	0.618	-1.321	1.110
Hijacking	-1.931	1.472	-4.556	1.276
Hostage Taking - Barricade Incident	2.035	2.199	-1.543	7.024
Hostage Taking - Kidnapping	1.075	0.815	-0.352	2.855
Unknown	1.787	2.238	-1.889	6.706
Mean - Business	0.917	0.798	-0.454	2.528
Mean - Educational Institution	-0.767	0.640	-2.011	0.452
Mean - Food or Water Supply	1.329	2.288	-2.595	6.176
Mean - Government (Diplomatic)	-0.454	0.695	-1.757	0.958
Mean - Government (General)	-0.610	0.493	-1.500	0.321
Mean - Journalists & Media	2.100	2.116	-1.270	6.734
Mean - Maritime	1.347	2.357	-2.746	6.468
Mean - Military	-0.318	0.564	-1.345	0.804
Mean - NGO	-0.530	0.777	-1.978	0.985
Mean - Other	0.077	1.204	-2.006	2.704
Mean - Police	-0.085	0.516	-1.016	0.957
Mean - Private Citizens & Property	1.674	0.679	0.465	2.989
Mean - Religious Figures/Institutions	1.323	0.974	-0.440	3.259
Mean - Telecommunication	2.743	2.274	-0.837	7.969
Mean - Terrorists	-0.381	1.089	-2.448	1.827
Mean - Tourists	0.640	2.851	-4.604	6.452
Mean - Transportation	-1.085	0.618	-2.256	0.162
Mean - Unknown	-2.469	0.696	-3.808	-1.111
Mean - Utilities	-1.660	1.017	-3.502	0.498
Variance - Business	1.122	0.399	0.585	2.136
Variance - Educational Institution	1.148	0.406	0.601	2.148
Variance - Food or Water Supply	1.097	0.381	0.583	2.038
Variance - Government (Diplomatic)	1.110	0.394	0.587	2.068
Variance - Government (General)	1.049	0.334	0.577	1.868
Variance - Journalists & Media	1.097	0.379	0.588	2.010
Variance - Maritime	1.101	0.382	0.587	2.070
Variance - Military	1.228	0.459	0.623	2.374
Variance - NGO	1.111	0.392	0.586	2.098
Variance - Other	1.134	0.409	0.594	2.177
Variance - Police	1.032	0.322	0.577	1.814
Variance - Private Citizens & Property	1.106	0.361	0.596	2.020
Variance - Religious Figures/Institutions	1.149	0.427	0.596	2.212
Variance - Telecommunication	1.094	0.383	0.580	2.021
Variance - Terrorists	1.157	0.430	0.597	2.214
Variance - Tourists	1.109	0.388	0.581	2.054
Variance - Transportation	1.095	0.379	0.588	2.043
Variance - Unknown	1.102	0.396	0.579	2.072
Variance - Utilities	1.152	0.430	0.594	2.233
Variance - Unknown	1.112	0.402	0.581	2.109
Variance - Utilities	1.144	0.417	0.594	2.183

Tab. 5.2: Afghanistan - Hierarchical Bayesian Model Using DP priors

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
(Intercept)	3.965	0.348	3.289	4.664
Suicide	-1.036	0.208	-1.429	-0.622
Assassination	-2.858	0.299	-3.445	-2.277
Bombing/Explosion	-0.634	0.314	-1.245	-0.013
Facility/Infrastructure	-0.418	0.562	-1.416	0.776
Hijacking	-1.787	1.225	-4.007	0.863
Hostage Taking (Barricade Incident)	1.976	2.220	-1.472	7.026
Hostage Taking (Kidnapping)	1.017	0.796	-0.363	2.820
Unknown	2.059	2.350	-1.635	7.147
μ_{Baseline}	3.853	0.635	2.569	5.086
Σ_{Baseline}	1.007	1.052	0.214	3.286
Number of clusters	4.691	1.552	2.000	8.000
α	1.054	0.566	0.249	2.419
Business	4.863	0.655	3.539	5.942
Educational Institution	3.644	0.346	2.910	4.312
Food or Water Supply	4.084	0.826	2.694	5.781
Government (Diplomatic)	3.721	0.371	3.028	4.543
Government (General)	3.636	0.287	3.067	4.204
Journalists & Media	4.274	0.838	3.048	5.898
Maritime	4.106	0.831	2.683	5.814
Military	3.734	0.316	3.139	4.384
NGO	3.715	0.435	2.907	4.812
Other	3.948	0.666	2.923	5.556
Police	3.835	0.322	3.228	4.523
Private Citizens & Property	5.255	0.418	4.460	6.120
Religious Figures/Institutions	4.878	0.683	3.475	5.985
Telecommunication	4.424	0.843	3.174	5.965
Terrorists	3.782	0.582	2.679	5.287
Tourists	3.989	0.854	2.301	5.742
Transportation	3.519	0.421	2.526	4.222
Unknown	2.634	0.727	1.321	3.956
Utilities	3.472	0.650	1.913	4.687

success of the U.S. military in preventing terrorist attacks of this nature. Attacks related to hostage taking, on the other hand, had positive coefficients and have therefore generally been much more successful. As a result, it could be useful for the U.S. military to work with Afghan security forces to determine where hostage taking is predominately occurring and improving security around such locations. Additionally, our model using DP priors suggests that there may be some clustering of attacks around potential targets not necessarily observable from the data directly. We will discuss this result in more detail later in this chapter.

5.3.3 The War in Iraq

For decades, Iraq has been a country housing three distinct ethnic and religious groups - Sunni Muslims, Shi'ite Muslims, and Kurds. Saddam Hussein, the President of Iraq from 1979 up until his fall in 2003, was infamous for being an egregious violator of human rights, providing preferential treatment for many of his fellow Sunni Muslims, while subjecting many Shi'ite Muslims and Kurds to severe and often deadly persecution. In 1982, for example, Saddam detained nearly 800 men, women, and children from the Shi'ite town of Dujail following a failed assassination attempt. In court, prosecutors and witnesses claimed that they witnessed torture and murder of many of these civilians. In 1988, during the final days of the Iran-Iraq war, Saddam ordered a poison gas attack in the northern Iraqi town of Halabja, killing thousands of innocent civilians and injuring scores of others. Many of the survivors are still suffering from the long term effects today [64, 101].

In addition to human rights violations, Saddam invaded two of his neighboring countries, supported terrorists, led the international community to believe that he had intentions to acquire illicit weapons, and attempted to assassinate a former U.S. President [34, 153]. In light of these violations of international law, President Bill Clinton in 1998 signed into law the Iraq Liberation Act, making regime change in Iraq the explicit policy of the United States government. Four years later, President George W. Bush signed into law the Iraq War Resolution, allowing the use of American ground troops to carry out the goals outlined in the Iraq Liberation Act [5, 68].

In 2003, a military coalition led by the United States and British governments removed Saddam from power and helped install a democratic regime that espouses the values of freedom and human rights. Although the initial campaign in Iraq to remove Saddam was astonishingly successful, maintaining stability in the immediate post-Saddam Iraq was quite difficult. Terrorist attacks occurred on a regular basis, jeopardizing the lives of coalition forces as well as civilians. Recent research has shown that most of these suicide attacks have been targeted towards civilians [148]. Tables 5.3-5.4 contain an analysis of these attacks from 2003 through 2011, which consisted of 7621 observations, using our Bayesian approach.

Our posterior coefficient estimates for suicide attacks illustrate that, just like Afghanistan, suicide attacks were generally unsuccessful (with posterior mean coef-

Tab. 5.3: Iraq - Hierarchical Bayesian Model Using Normally Distributed Priors

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
Intercept	4.838	0.502	4.011	5.910
Suicide	-0.522	0.199	-0.906	-0.130
Assassination	-4.261	0.290	-4.844	-3.721
Bombing/Explosion	-1.063	0.282	-1.628	-0.530
Facility/Infrastructure	-0.770	0.746	-2.128	0.791
Hijacking	0.522	2.822	-4.605	6.178
Hostage Taking - Barricade Incident	2.077	2.189	-1.384	6.985
Hostage Taking - Kidnapping	-0.682	0.559	-1.711	0.493
Unknown	1.265	2.475	-2.872	6.609
Mean - Business	0.677	0.573	-0.528	1.694
Mean - Educational Institution	-0.536	0.606	-1.768	0.599
Mean - Food or Water Supply	2.000	2.521	-2.309	7.333
Mean - Government (Diplomatic)	-0.420	0.590	-1.644	0.692
Mean - Government (General)	-0.076	0.488	-1.131	0.684
Mean - Journalists & Media	1.364	0.971	-0.373	3.532
Mean - Maritime	-1.709	1.514	-4.549	1.478
Mean - Military	0.112	0.510	-1.007	0.992
Mean - NGO	0.349	1.284	-1.981	2.995
Mean - Other	1.651	1.086	-0.433	3.839
Mean - Police	0.358	0.459	-0.660	1.071
Mean - Private Citizens & Property	0.998	0.493	-0.167	1.728
Mean - Religious Figures/Institutions	0.543	0.542	-0.664	1.532
Mean - Telecommunication	1.917	2.326	-1.943	7.024
Mean - Terrorists	0.027	0.594	-1.196	1.101
Mean - Tourists	1.214	2.109	-2.321	5.838
Mean - Transportation	0.280	0.614	-0.943	1.483
Mean - Unknown	-1.579	0.724	-3.007	-0.143
Mean - Utilities	-0.635	0.643	-1.911	0.615
Mean -Violent Political Party	3.536	1.590	0.875	7.189
Variance - Business	1.126	0.374	0.588	2.030
Variance - Educational Institution	1.176	0.419	0.601	2.201
Variance - Food or Water Supply	1.100	0.387	0.584	2.057
Variance - Government (Diplomatic)	1.226	0.459	0.618	2.349
Variance - Government (General)	1.392	0.460	0.707	2.452
Variance - Journalists & Media	1.185	0.469	0.598	2.364
Variance - Maritime	1.122	0.405	0.585	2.113
Variance - Military	1.080	0.348	0.586	1.969
Variance - NGO	1.107	0.383	0.581	2.059
Variance - Other	1.086	0.376	0.576	2.021
Variance - Police	0.831	0.229	0.479	1.323
Variance - Private Citizens & Property	1.163	0.394	0.615	2.172
Variance - Religious Figures/Institutions	1.280	0.519	0.616	2.560
Variance - Telecommunication	1.096	0.381	0.582	2.046
Variance - Terrorists	1.155	0.423	0.602	2.252
Variance - Tourists	1.101	0.384	0.584	2.055
Variance - Transportation	1.102	0.435	0.577	2.219
Variance - Unknown	1.118	0.396	0.595	2.116
Variance - Utilities	1.119	0.374	0.603	2.041
Variance -Violent Political Party	1.103	0.386	0.582	2.057
Variance - Utilities	1.135	0.411	0.587	2.150
Variance - Violent Political Party	1.107	0.393	0.586	2.064

Tab. 5.4: Iraq - Hierarchical Bayesian Model Using DP priors

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
(Intercept)	4.543	0.278	4.012	5.109
Suicide	-0.449	0.180	-0.785	-0.087
Assassination	-3.836	0.264	-4.361	-3.316
Bombing/Explosion	-1.049	0.267	-1.593	-0.540
Facility/Infrastructure	-1.022	0.639	-2.187	0.274
Hijacking	0.622	2.851	-4.434	6.538
Hostage Taking (Barricade Incident)	2.094	2.149	-1.361	6.976
Hostage Taking (Kidnapping)	-0.645	0.517	-1.590	0.404
Unknown	1.390	2.480	-2.778	6.753
μ_{Baseline}	4.546	0.434	3.669	5.387
Σ_{Baseline}	0.372	0.389	0.099	1.200
Number of clusters	4.685	1.566	2.000	8.000
α	1.046	0.572	0.242	2.440
Business	4.857	0.375	4.152	5.599
Educational Institution	4.251	0.368	3.473	4.961
Food or Water Supply	4.576	0.486	3.710	5.555
Government (Diplomatic)	4.257	0.347	3.548	4.939
Government (General)	4.241	0.269	3.725	4.785
Journalists & Media	4.823	0.441	3.992	5.685
Maritime	4.405	0.493	3.379	5.357
Military	4.425	0.317	3.822	5.064
NGO	4.506	0.462	3.639	5.467
Other	4.741	0.458	3.918	5.646
Police	4.633	0.281	4.097	5.199
Private Citizens & Property	5.143	0.294	4.579	5.740
Religious Figures/Institutions	4.673	0.363	3.993	5.419
Telecommunication	4.586	0.482	3.717	5.537
Terrorists	4.400	0.345	3.756	5.104
Tourists	4.563	0.482	3.689	5.528
Transportation	4.545	0.371	3.870	5.329
Unknown	4.112	0.491	2.898	4.930
Utilities	4.260	0.381	3.437	4.995
Mean -Violent Political Party	4.822	0.484	3.946	5.784

ficient estimates approximately -0.5 in both models), although not as unsuccessful as Afghanistan. Assassination attempts (-4.261 in the normal model and -3.736 in the DP model), bombings (-1.063 in the normal model and -1.049 in the DP model), and attacks on Facility/Infrastructure (-0.770 in the normal model and -1.022 in the DP model) were also generally unsuccessful. It is interesting to note, however, that both models suggest that hijacking (0.522 in the normal model and 0.622 in the DP model) and hostage taking via barricade incidents (2.077 in the normal model and 2.094 in the DP model), on the other hand, were successful. Hostage taking via kidnapping (-0.682 in the normal model and -0.645 in the DP model), however, was not. These results suggest that Iraqi security forces that have taken responsibility for the country after the departure of U.S. troops in 2011 should consider fortifying security to prevent hijackings and hostage takings via barricade incident. The Iraqi government could consider providing additional security personal or offer recommendations regarding private security to do so.

5.3.4 *The Sri Lankan Civil War*

The Sri Lankan Civil War was a 26 year conflict from 1983-2009. Fought by the Liberation Tigers of Tamil Eelam (LTTE), the war was waged against the government of Sri Lanka, who were seeking to create a separate Tamil state within the northeastern part of the country. The LTTE engaged in terrorist attacks ranging from suicide attacks against civilians, to coordinated attacks against religious figures and important facilities, to assassination attempts against members of the Sri Lankan government. These attacks included the 1993 assassination of President Ranasinghe Premadasa, the 1996 bombings of the Sri Lankan Central Bank, killing over 90 and injuring over 1400, and the 1998 bombing of the revered Temple of the Tooth among others [54]. Shortly after the killing of LTTE leader Velupillai Prabhakaran in 2009, the 26 year-long Civil War ended [161]. Tables 5.5 and 5.6 contain our Bayesian analysis of this data set from over the course of the conflict which consisted of 2924 observations.

The results from the civil war in Sri Lanka are quite similar to those in Iraq. In particular, suicide attacks (with posterior mean coefficient estimate -0.421 in the normal model and -0.371 in the DP model), assassination attempts (-1.487 in the normal model and -1.532 in the DP model), bombing attempts (-1.206 in the normal model and -1.169 in the DP model), attacks on facility/infrastructure (essentially zero in the normal model and -0.196 in the DP model), hostage taking via kidnapping (-0.833 in the normal model and -0.877 in the DP model), unarmed assaults (-1.206 in the normal model and -1.201 in the DP model), and terrorist attacks of an unknown nature (-1.454 in the normal model and -1.359 in the DP model) were largely unsuccessful. Hostage taking via barricade incident (1.405 in the normal model and 1.848 in the DP model) and hijackings (1.822 in the normal model and 1.848 in the DP model), on the other hand, were much more successful. Despite the fact that the war ended in 2009, understanding these issues could be

Tab. 5.5: Sri Lanka - Hierarchical Bayesian Model Using Normally Distributed Priors

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
Intercept	3.842	0.570	2.867	5.018
Suicide	-0.421	0.386	-1.153	0.351
Assassination	-1.487	0.330	-2.148	-0.844
Bombing/Explosion	-1.206	0.277	-1.752	-0.671
Facility/Infrastructure	0.006	0.628	-1.147	1.326
Hijacking	1.822	2.276	-1.933	6.908
Hostage Taking - Barricade Incident	1.405	2.448	-2.774	6.641
Hostage Taking - Kidnapping	-0.833	0.659	-2.039	0.547
Unarmed Assault	-1.206	1.455	-3.793	1.898
Unknown	-1.454	0.431	-2.282	-0.581
Mean - Business	0.448	0.671	-0.879	1.722
Mean - Educational Institution	3.426	1.919	0.363	7.843
Mean - Food or Water Supply	2.225	2.205	-1.553	6.978
Mean - Government (Diplomatic)	-1.061	0.962	-2.926	0.865
Mean - Government (General)	-0.021	0.580	-1.224	1.069
Mean - Journalists & Media	0.094	0.867	-1.507	1.883
Mean - Maritime	1.100	1.169	-0.941	3.678
Mean - Military	0.580	0.575	-0.509	1.579
Mean - NGO	3.277	1.884	0.096	7.414
Mean - Other	-0.701	0.920	-2.465	1.159
Mean - Police	0.416	0.602	-0.832	1.485
Mean - Private Citizens & Property	2.324	0.674	0.961	3.534
Mean - Religious Figures/Institutions	1.763	1.203	-0.340	4.471
Mean - Telecommunication	-0.960	0.939	-2.785	0.895
Mean - Terrorists	3.528	1.781	0.498	7.640
Mean - Tourists	1.671	2.359	-2.406	6.703
Mean - Transportation	0.418	0.631	-0.797	1.649
Mean - Unknown	-2.413	0.675	-3.776	-1.128
Mean - Utilities	0.232	0.850	-1.375	2.017
Mean - Violent Political Party	0.466	0.707	-0.876	1.857
Variance - Business	1.102	0.392	0.585	2.108
Variance - Educational Institution	1.092	0.385	0.571	2.028
Variance - Food or Water Supply	1.103	0.388	0.587	2.056
Variance - Government (Diplomatic)	1.132	0.403	0.592	2.145
Variance - Government (General)	1.141	0.430	0.591	2.157
Variance - Journalists & Media	1.117	0.389	0.585	2.099
Variance - Maritime	1.115	0.394	0.590	2.075
Variance - Military	1.149	0.452	0.591	2.313
Variance - NGO	1.096	0.385	0.582	2.050
Variance - Other	1.134	0.413	0.590	2.190
Variance - Police	1.146	0.439	0.600	2.213
Variance - Private Citizens & Property	1.083	0.359	0.574	1.997
Variance - Religious Figures/Institutions	1.094	0.382	0.577	2.070
Variance - Telecommunication	1.126	0.397	0.593	2.101
Variance - Terrorists	1.097	0.385	0.577	2.053
Variance - Tourists	1.109	0.393	0.581	2.090
Variance - Transportation	1.139	0.414	0.592	2.169
Variance - Unknown	1.180	0.437	0.602	2.271
Variance - Utilities	1.133	0.405	0.601	2.131
Variance - Violent Political Party	1.112	0.407	0.576	2.096

Tab. 5.6: Sri Lanka - Hierarchical Bayesian Model Using DP priors

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
(Intercept)	3.890	0.354	3.233	4.627
Suicide	-0.371	0.325	-0.982	0.282
Assassination	-1.532	0.300	-2.128	-0.950
Bombing/Explosion	-1.169	0.263	-1.699	-0.658
Facility/Infrastructure	-0.196	0.597	-1.262	1.044
Hijacking	1.848	2.151	-1.611	6.598
Hostage Taking (Barricade Incident)	1.345	2.439	-2.720	7.063
Hostage Taking (Kidnapping)	-0.870	0.598	-1.948	0.414
Unarmed Assault	-1.201	1.341	-3.520	1.832
Unknown	-1.359	0.407	-2.121	-0.517
μ_{Baseline}	3.506	0.775	1.896	4.967
Σ_{Baseline}	1.944	3.062	0.463	6.355
Number of clusters	4.513	1.349	3.000	8.000
α	0.999	0.529	0.263	2.306
Business	3.763	0.266	3.271	4.305
Educational Institution	4.698	0.979	3.371	6.455
Food or Water Supply	4.188	0.886	3.032	6.147
Government (Diplomatic)	3.262	0.830	1.368	4.230
Government (General)	3.669	0.284	3.053	4.191
Journalists & Media	3.742	0.369	3.027	4.473
Maritime	4.014	0.678	3.223	5.836
Military	3.783	0.241	3.333	4.284
NGO	4.572	0.973	3.336	6.374
Other	3.519	0.609	1.793	4.279
Police	3.759	0.242	3.302	4.253
Private Citizens & Property	5.452	0.465	4.598	6.410
Religious Figures/Institutions	4.456	0.875	3.357	6.159
Telecommunication	3.322	0.785	1.446	4.226
Terrorists	4.992	0.939	3.459	6.571
Tourists	4.060	0.930	2.080	6.074
Transportation	3.753	0.255	3.269	4.267
Unknown	1.604	0.454	0.716	2.499
Utilities	3.752	0.367	3.068	4.479
Violent Political Party	3.764	0.286	3.250	4.339

useful for the Sri Lankan government as post-war reconciliation processes continue.

5.3.5 The Troubles

Commonly known as "The Troubles," the Civil War in Ireland was an ethno-nationalist conflict over the constitutional status of Northern Ireland [102]. The Troubles bears many similarities to the Civil War in Sri Lanka. Irish nationalists, primarily Catholic, wanted an Independent State of Northern Ireland, while Unionists and loyalists, primarily Protestant, preferred Northern Ireland to remain part of the United Kingdom. The Unionists and loyalists were represented politically by the Ulster Volunteer Force (UVF) and the Ulster Defence Association (UDA), while the Irish nationalists were represented by the Irish Republican Army (IRA). Unlike many political groups, however, the UVF, UDA, and IRA were known for engaging in violent behavior, including engaging in attacks targeting civilians. In fact, these organizations have been deemed by the United States State Department as terrorist groups alongside organizations such as *al Qaeda*, Hezbollah, Hamas, and LTTE among others [158].

The Good Friday Agreement of 1998, a compromise that Northern Ireland would remain a component of the United Kingdom until the people of northern Ireland and the Republic of Ireland would determine otherwise, contained provisions for the creation of institutions to civilly discuss issues between Northern Ireland and the Republic of Ireland as well as between Britain and Ireland [114]. The Agreement, however, was not a panacea for the country's issues, and terrorist attacks continued after its passage. Tables 5.7-5.10 contain our analysis of the attacks in Northern Ireland from 1970 up through the Good Friday Agreement of 1998 (consisting of 3517 observations) and then from 1998 through the present (consisting of 457 observations). During these time periods, only one suicide attack occurred in our data set (on December 29, 1998 in Armagh, Northern Ireland) and was consequently excluded from our analysis.

Our posterior estimates indicate that hostage taking via kidnapping (posterior mean coefficient estimates 1.548 in the normal model and 1.429 in the DP model), unarmed assaults (2.585 in the normal model and 2.619 in the DP model), and terrorist attacks of an unknown nature (0.156 in the normal model and 2.619 in the DP model) were the most successful types of terrorist attacks in Ireland before the Good Friday Agreement of 1998. Although terrorist attacks continued after the Agreement, the most successful types of attacks were via hijacking (1.146 in the normal model and 2.869 in the DP model) and hostage taking via kidnapping (1.090 in the normal model and 1.042 in the DP model). The success of unarmed assaults declined after the Good Friday Agreement of 1998 as the posterior coefficient estimates became negative (-0.865 in the normal model -1.083 in the DP model).

Tab. 5.7: Ireland before the 1998 Good Friday Agreement - Hierarchical Bayesian Model Using Normally Distributed Priors

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
Intercept	2.346	0.639	1.087	3.644
Assassination	-0.421	0.236	-0.900	0.028
Bombing/Explosion	-0.972	0.233	-1.447	-0.524
Facility/Infrastructure	-0.831	0.319	-1.450	-0.196
Hijacking	-2.452	1.233	-4.837	0.056
Hostage Taking - Barricade Incident	-0.283	1.268	-2.476	2.485
Hostage Taking - Kidnapping	1.548	1.113	-0.313	4.033
Unarmed Assault	2.585	2.088	-0.842	7.188
Unknown	0.156	0.861	-1.349	2.012
Mean - Airports & Airlines	-1.152	1.132	-3.339	1.048
Mean - Business	0.774	0.620	-0.488	2.043
Mean - Educational Institution	-0.438	0.949	-2.235	1.471
Mean - Government (Diplomatic)	1.520	2.522	-3.090	6.930
Mean - Government (General)	-0.491	0.629	-1.783	0.751
Mean - Journalists & Media	-0.335	1.495	-3.178	2.703
Mean - Maritime	2.158	2.310	-1.853	7.131
Mean - Military	1.067	0.647	-0.291	2.447
Mean - NGO	1.917	2.223	-2.065	6.544
Mean - Other	1.935	1.165	-0.152	4.434
Mean - Police	0.216	0.623	-1.054	1.492
Mean - Private Citizens & Property	0.898	0.616	-0.415	2.122
Mean - Religious Figures/Institutions	-0.029	0.958	-1.781	1.971
Mean - Telecommunications	-0.950	1.723	-4.285	2.624
Mean - Terrorists	-0.279	0.669	-1.723	1.066
Mean - Tourists	1.464	2.532	-3.218	6.731
Mean - Transportation	0.543	0.756	-0.926	2.037
Mean - Unknown	1.774	0.984	0.043	3.871
Mean - Utilities	-2.775	2.285	-7.589	1.418
Variance - Airports & Airlines	1.128	0.403	0.591	2.161
Variance - Business	1.195	0.428	0.619	2.306
Variance - Educational Institution	1.119	0.400	0.585	2.119
Variance - Government (Diplomatic)	1.114	0.397	0.585	2.094
Variance - Government (General)	1.340	0.536	0.654	2.729
Variance - Journalists & Media	1.131	0.408	0.593	2.164
Variance - Maritime	1.103	0.381	0.585	2.041
Variance - Military	1.162	0.444	0.586	2.326
Variance - NGO	1.106	0.391	0.579	2.082
Variance - Other	1.116	0.394	0.589	2.113
Variance - Police	1.210	0.428	0.631	2.324
Variance - Private Citizens & Property	1.001	0.335	0.547	1.866
Variance - Religious Figures/Institutions	1.124	0.406	0.590	2.144
Variance - Telecommunications	1.120	0.405	0.591	2.121
Variance - Terrorists	1.070	0.365	0.576	1.987
Variance - Tourists	1.104	0.386	0.584	2.084
Variance - Transportation	1.134	0.410	0.593	2.148
Variance - Unknown	1.113	0.389	0.582	2.103
Variance - Utilities	1.111	0.394	0.587	2.092

Tab. 5.8: Ireland before the 1998 Good Friday Agreement - Hierarchical Bayesian Model Using DP priors

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
(Intercept)	2.336	0.240	1.854	2.802
Assassination	-0.393	0.195	-0.768	-0.001
Bombing/Explosion	-0.897	0.191	-1.259	-0.530
Facility/Infrastructure	-0.847	0.253	-1.314	-0.360
Hijacking	-2.282	1.075	-4.222	-0.064
Hostage Taking (Barricade Incident)	-0.012	1.174	-2.052	2.263
Hostage Taking (Kidnapping)	1.429	0.997	-0.308	3.603
Unarmed Assault	2.619	2.193	-0.859	7.858
Unknown	0.222	0.817	-1.192	1.955
μ_{Baseline}	2.231	0.396	1.453	2.998
Σ_{Baseline}	0.391	0.350	0.113	1.187
Number of clusters	4.453	1.441	2.000	8.000
α	1.006	0.546	0.234	2.320
Airports & Airlines	1.977	0.491	1.162	3.007
Business	2.783	0.203	2.359	3.167
Educational Institution	2.021	0.472	1.262	3.002
Government (Diplomatic)	2.379	0.536	1.365	3.174
Government (General)	1.703	0.250	1.228	2.213
Journalists & Media	2.236	0.542	1.299	3.118
Maritime	2.419	0.526	1.399	3.179
Military	2.831	0.185	2.476	3.195
NGO	2.420	0.530	1.392	3.183
Other	2.689	0.375	1.766	3.240
Police	2.152	0.201	1.757	2.545
Private Citizens & Property	2.821	0.184	2.467	3.182
Religious Figures/Institutions	2.191	0.510	1.330	3.086
Telecommunications	2.194	0.547	1.260	3.099
Terrorists	1.810	0.279	1.291	2.362
Tourists	2.366	0.543	1.359	3.168
Transportation	2.555	0.391	1.728	3.145
Unknown	2.782	0.291	2.088	3.273
Utilities	2.144	0.554	1.214	3.090

Tab. 5.9: Ireland after the 1998 Good Friday Agreement - Hierarchical Bayesian Model Using Normally Distributed Priors

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
Intercept	2.819	1.058	0.910	4.946
Assassination	0.462	1.114	-1.734	2.530
Bombing Explosion	-2.260	1.049	-4.355	-0.379
Facility/Infrastructure	3.759	2.077	-0.005	8.227
Hijacking	1.146	2.613	-3.549	6.608
Hostage Taking - Kidnapping	1.090	1.203	-1.318	3.368
Unarmed Assault	-0.865	1.648	-3.965	2.529
Mean - Business	-0.226	0.591	-1.361	0.962
Mean - Educational Institution	0.223	0.482	-0.716	1.162
Mean - Government (General)	2.191	1.224	0.158	5.149
Mean - Journalists & Media	-2.353	2.281	-7.232	1.800
Mean - Military	-0.680	1.573	-3.670	2.601
Mean - NGO	1.068	2.654	-3.878	6.628
Mean - Other	-1.390	1.829	-4.945	2.298
Mean - Police	-0.889	1.135	-3.037	1.455
Mean - Private Citizens & Property	-0.209	0.388	-1.013	0.512
Mean - Religious Figures/Institutions	0.668	1.322	-1.743	3.554
Mean - Terrorists	3.080	1.941	-0.166	7.442
Mean - Tourists	-1.516	1.224	-3.942	0.885
Mean - Transportation	-1.288	0.908	-3.060	0.476
Mean - Unknown	-3.902	1.071	-6.157	-1.986
Mean - Utilities	-2.514	2.379	-7.595	1.825
Mean - Violent Political Party	2.808	2.037	-0.659	7.306
Variance - Business	1.148	0.402	0.602	2.147
Variance - Educational Institution	1.085	0.367	0.580	1.998
Variance - Government (General)	1.109	0.392	0.581	2.082
Variance - Journalists & Media	1.112	0.398	0.581	2.101
Variance - Military	1.106	0.391	0.585	2.078
Variance - NGO	1.108	0.390	0.584	2.077
Variance - Other	1.114	0.398	0.590	2.084
Variance - Police	1.143	0.406	0.594	2.149
Variance - Private Citizens & Property	1.136	0.410	0.599	2.169
Variance - Religious Figures/Institutions	1.097	0.380	0.585	2.037
Variance - Terrorists	1.100	0.387	0.584	2.077
Variance - Tourists	1.110	0.392	0.590	2.067
Variance - Transportation	1.111	0.391	0.589	2.102
Variance - Unknown	1.087	0.374	0.577	2.021
Variance - Utilities	1.109	0.392	0.585	2.073
Variance - Violent Political Party	1.106	0.387	0.588	2.069

Tab. 5.10: Ireland after the 1998 Good Friday Agreement - Hierarchical Bayesian Model Using DP priors

	Post. Mean	Post. Std Dev	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit
(Intercept)	2.417	0.474	1.514	3.372
Assassination	-0.210	1.268	-2.336	2.821
Bombing/Explosion	-2.288	0.414	-3.159	-1.542
Facility/Infrastructure	0.475	0.742	-0.863	2.117
Hijacking	2.869	1.964	-0.496	7.233
Hostage Taking (Kidnapping)	1.042	2.514	-3.330	6.289
Unarmed Assault	-1.083	1.368	-3.460	2.115
μ_{Baseline}	1.796	0.885	-0.109	3.435
Σ_{Baseline}	1.212	1.564	0.181	4.792
Number of clusters	3.272	1.291	2.000	6.000
α	0.798	0.491	0.154	2.031
Business	2.737	0.422	1.995	3.633
Educational Institution	2.453	0.776	0.423	3.618
Government (General)	2.685	0.426	1.891	3.547
Journalists & Media	2.543	0.806	0.377	3.857
Military	2.909	0.646	2.024	4.669
NGO	2.284	0.945	-0.108	3.580
Other	2.470	0.724	0.546	3.568
Police	2.738	0.417	2.000	3.621
Private Citizens & Property	2.685	0.396	1.981	3.514
Religious Figures/Institutions	2.723	0.521	1.780	3.807
Terrorists	2.839	0.640	1.955	4.504
Tourists	2.238	0.911	0.007	3.507
Transportation	2.278	0.808	0.357	3.468
Unknown	0.529	0.901	-1.337	2.226
Utilities	2.155	1.077	-0.507	3.568
Violent Political Party	2.798	0.624	1.872	4.262

5.3.6 The Use of Dirichlet Process Priors

We applied either a normal prior or a DP Prior to the coefficients regarding the targets of terrorist attacks. To compare the DP model for each conflict with its normal counterpart, we estimated pseudo-Bayes factors (PsBF) [47, 48]. The PsBF is based on a model's cross-validation predictive density. As standard Bayes factors are often not computationally feasible to compute, the PsBF is considered a useful surrogate. If we let $\boldsymbol{y}$ be our observed data, y_t be the t^{th} terrorist attack at time t across $t = 1, \ldots, T$ observations, and $\boldsymbol{y_{(t)}}$ be the data of attacks with observation t deleted, the cross validative predictive density $\pi(y_t|\boldsymbol{y_{(t)}})$ is:

$$\begin{aligned} \pi(y_t|\boldsymbol{y_{(t)}}) &= \int \pi(y_t|\boldsymbol{\beta}, \boldsymbol{y_{(t)}})\pi(\boldsymbol{\beta}, |\boldsymbol{y_{(t)}})\mathrm{d}\boldsymbol{\beta} \\ &\approx \left[\frac{1}{R}\sum_{r=1}^{R}\frac{1}{f(y_t, \boldsymbol{\beta}^{(r)})}\right]^{-1}, \end{aligned} \tag{5.6}$$

where f is the likelihood function and $\boldsymbol{\beta}^{(r)}$ is the vector of parameter values obtained during the r^{th} MCMC iteration [115]. The relation in (5.6) is based on a truncated series approximation of the harmonic mean of the logistic regression function evaluated at each posterior draw, averaged across all R post burn-in MCMC iterations [115]. As a result, the PsBF comparing the normal model (M=1) to the DP model (M=2) can be written as follows:

$$\text{PsBF} = \prod_{t=1}^{T} \frac{\pi(y_t|\boldsymbol{y}_{(t)}, M=2)}{\pi(y_t|\boldsymbol{y}_{(t)}, M=1)}. \tag{5.7}$$

We can take advantage of the additive properties of logarithms and utilize (5.6) to estimate the logarithms of the numerator as well as the denominator in (5.11). These quantities, approximations to the log-marginal data likelihoods for each model, enable us to estimate PsBFs for each conflict examined. Table 5.11 contains our results.

Tab. 5.11: PsBF Computation comparing DP Model to Normal Model

	Normal Model	DP Model	Pseudo Bayes Factor
Afghanistan	-518.066	-510.610	1730.213
Iraq	-1183.940	-1183.310	1.878
Sri Lanka	-563.147	-556.940	496.310
N. Ireland Before 1998 GF Agreement	-1278.738	-1272.590	467.781
N. Ireland After 1998 GF Agreement	-230.315	-231.140	0.438

PsBF estimates greater than one indicate stronger support for the DP Model, whereas estimates less than one provide support for the normal model. Furthermore, the larger the value, the stronger the indication of support [47, 48].[1] Therefore, these results indicate substantial support for the DP Model for understanding the determinants of successful attacks in Afghanistan, Sri Lanka, and Northern Ireland before the 1998 Good Friday Agreement. The PsBF for Iraq is 1.878, indicated some support for the DP Model over the normal model for that conflict. Thus, for four

[1] Pseudo-Bayes Factors have the potential to legitimately take on extremely large values (as large as exp(50)) that would indicate overwhelming evidence in favor of one model over another, as illustrated in Ansari and Mela's (2003) hierarchical Bayesian probit analysis of online customer clickstream behavior [3].

of the five conflicts examined, the DP model effectively captures clustering around potential targets for these conflicts compared to the normal model. The normal model, on the other hand, is considerably more adequate for modeling the conflict in Northern Ireland after the 1998 Good Friday Agreement. The adequacy of the normal model compared to the DP Model in this instance may be due the lack of clustering of attacks around targets in Northern Ireland in recent years.

Table 5.12 contains our average cluster sizes for each conflict, averaged over our MCMC iterations:

Tab. 5.12: Average Cluster Size - DP prior Model

	Average Cluster Size	Standard Deviation
Afghanistan	4.691	1.552
Iraq	4.685	1.566
Sri Lanka	4.513	1.349
Northern Ireland (before 1998 GF Agreement)	4.453	1.441
Northern Ireland (after 1998 GF Agreement)	3.272	1.291

These results suggest an interesting phenomenon illustrated by our DP prior model. In particular, our models cluster around a small number of targets. This clustering manifests the terrorists' strategy, improving the model's explanatory power for three of the conflicts described above. Future research could look into determining the exact composition of these clusters and provide military advice accordingly. Recent research has performed similar cluster analysis of other real-world phenomena, including the topics discussed by Barack Obama, John McCain, and Mitt Romney in the last two presidential elections [113].

5.3.7 Improving Bayesian Credible Interval Estimates

In the previous chapter of this dissertation, we discussed utilizing the semi-parametric DRE method to improve the credible interval estimation of a hierarchical linear model regarding welfare reform. In this section, we extend this technique to refine the credible interval estimates of the hierarchical generalized linear model

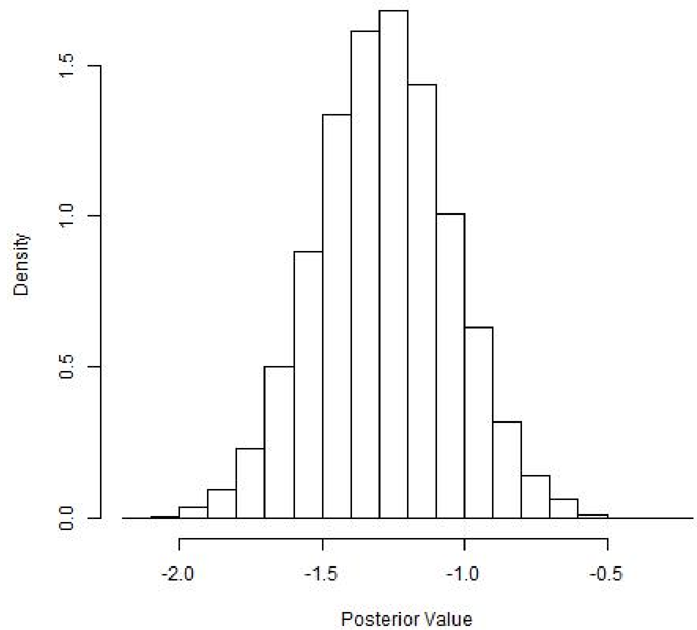

Fig. 5.1: Histogram of Posterior Sample for Suicide Attack Coefficient in Afghanistan, Assuming a Normally Distributed Prior

of this chapter. Specifically, we took posterior samples of several coefficients that warranted a more rigorous examination of their distributional properties. Upon looking at the histograms of these posterior samples, they all seemed to be reasonably "normal-like" in nature, appearing to be unimodal and reasonably symmetric with limited skewness. Several of these histograms are in Figures 5.1-5.10

As a result of these marginal posterior samples' distributional behavior, we used the approach in the previous chapter to estimate the 95% credible interval, again assuming a tilt function of the form $\boldsymbol{h}(x) = (x, x^2)'$. As the simulations of the previous chapter illustrated, applying the DRE method to posterior samples can provide more accurate quantile estimation. Our results are presented in Tables 5.13-5.17:

As we saw with our rigorous examination of welfare reform, the semiparametric DRE method enables us to better understand our posterior interval estimates re-

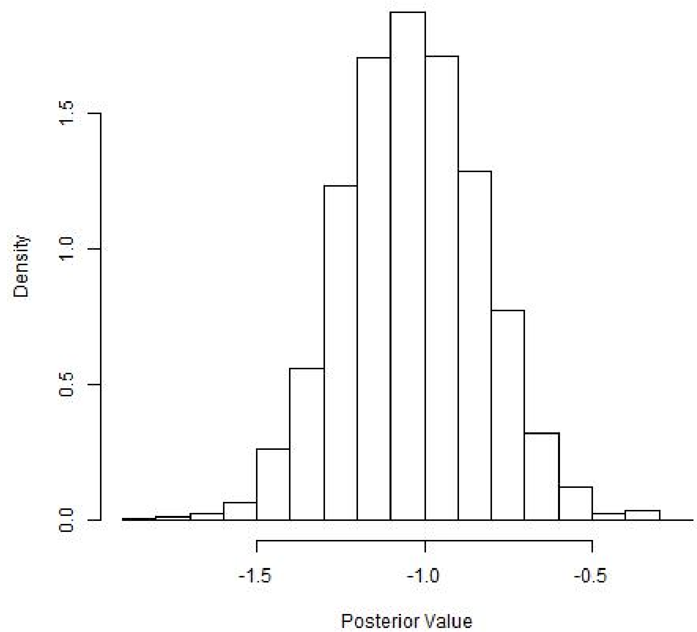

Fig. 5.2: Histogram of Posterior Sample for Suicide Attack Coefficient in Afghanistan, Assuming a DP prior

Tab. 5.13: Bayesian Credible Interval Refinements for Suicide Attack Posterior Coefficient, Afghanistan

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
Normal Prior	-1.741	-0.811	-1.748	-0.817
DP Prior	-1.429	-0.622	-1.439	-0.626

Tab. 5.14: Bayesian Credible Interval Refinements for Bombing/Explosion Posterior Coefficient, Iraq

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
Normal Prior	-1.628	-0.530	-1.622	-0.521
DP Prior	-1.593	-0.540	-1.571	-0.528

Tab. 5.15: Bayesian Credible Interval Refinements for Facility/Infrastructure Posterior Coefficient, Sri Lanka

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
Normal Prior	-1.147	1.326	-1.206	1.275
DP Prior	-1.262	1.044	-1.339	1.010

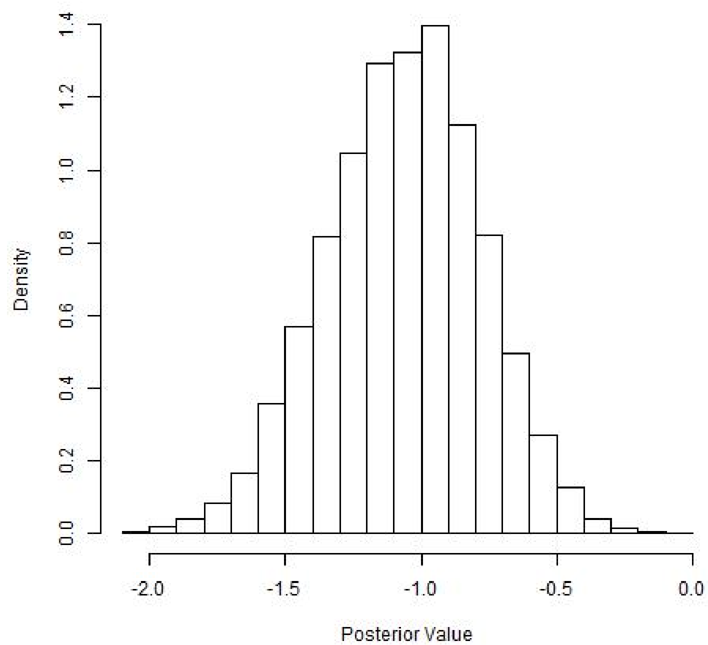

Fig. 5.3: Histogram of Posterior Sample for Bombing/Explosion Attack Coefficient in Iraq, Assuming a Normally Distributed Prior

Tab. 5.16: Bayesian Credible Interval Refinements for Hijacking Posterior Coefficient, Northern Ireland before 1998 Good Friday Agreement

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
Normal Prior	-4.837	0.056	-4.878	0.028
DP Prior	-4.222	-0.064	-4.342	-0.132

Tab. 5.17: Bayesian Credible Interval Refinements for Unarmed Assault Posterior Coefficient, Northern Ireland bafterefore 1998 Good Friday Agreement

	95% Credible Interval Lower Limit	95% Credible Interval Upper Limit	Refined 95% Credible Interval Lower Limit	Refined 95% Credible Interval Upper Limit
Normal Prior	-3.965	2.529	-4.072	2.431
DP Prior	-3.460	2.115	-3.672	1.737

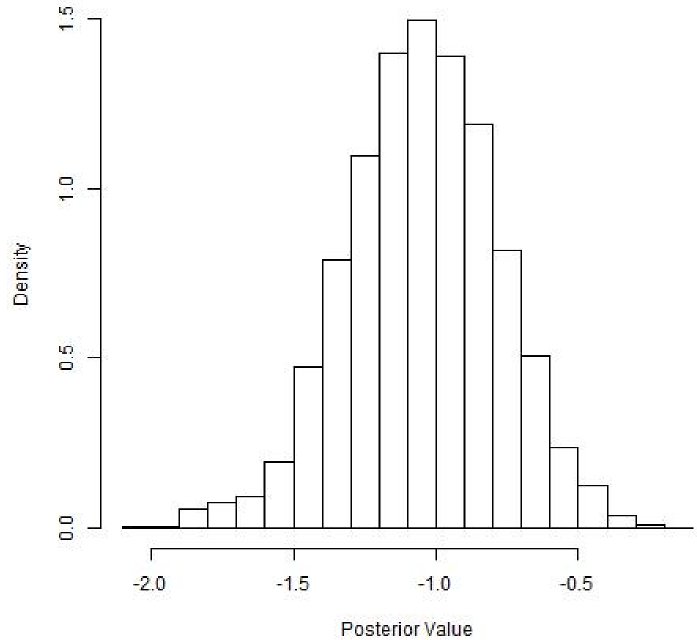

Fig. 5.4: Histogram of Posterior Sample for Bombing/Explosion Attack Coefficient in Iraq, Assuming a DP prior

garding terrorist attacks. This information is useful for policymakers to understand upper and lower limits of response coefficients so they can make more accurate statistical inferences regarding fortification of appropriate security measures. For example, in Northern Ireland after the 1998 Good Friday Agreement, although the success of unarmed assaults remains relatively low, the upper limit of the pertinent marginal posterior distribution still remains positive. This result suggests that this positive upper interval estimate was unlikely due to truncation of the MCMC sampler. We can use these results to advise associated policymakers that regardless of the negative Bayesian point estimator generated pertaining to unarmed assaults, they should not neglect security issues regarding attacks of this nature.

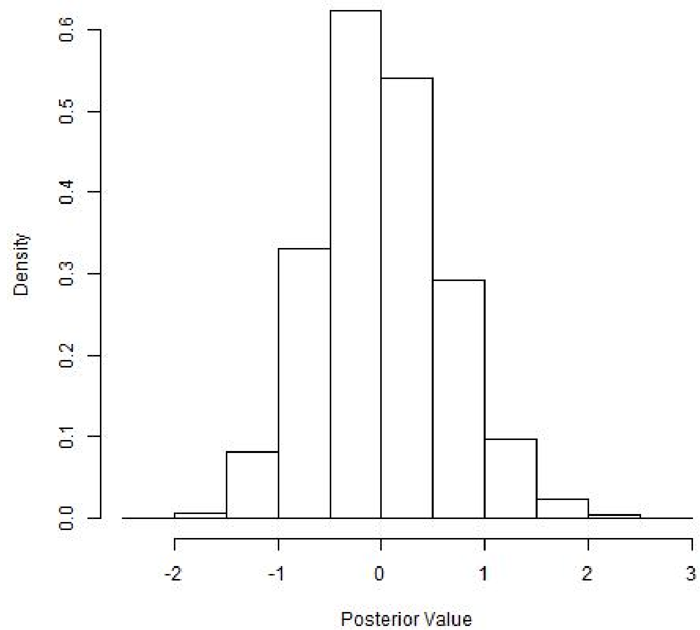

Fig. 5.5: Histogram of Posterior Sample for Facility/Infrastructure Coefficient in Sri Lanka, Assuming a Normally Distributed Prior

5.4 Conclusions and Future Research

Our results provide policymakers with a model offering useful counterterrorism strategies shedding light on which types of attacks are sufficiently defended against and which others warrant further fortification. As a result, policymakers can use our model in conjunction with other statistical models aimed at understanding terrorist networks as well as theoretical research in political science examining the goals, causes, and onset of terrorism [2, 15, 89, 120, 143, 166]. It is interesting to note that across the conflicts studied here, suicide attacks are generally unsuccessful, despite the tremendous attention they garner from the mainstream media. This result is consistent with the findings of Kyung et al (2011)'s study of the Middle East and Northern Africa [90]. Other considerably more coordinated non-suicide attacks, however, such as hostage taking, are often much more successful. This phenomenon

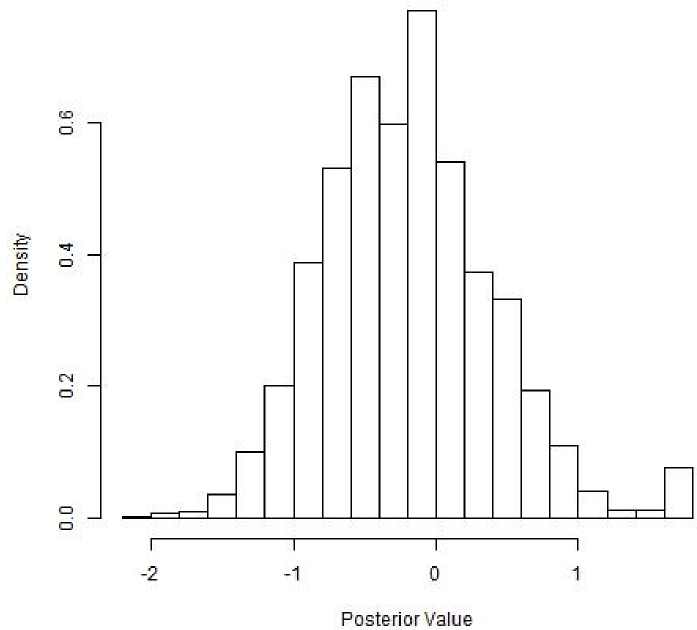

Fig. 5.6: Histogram of Posterior Sample for Facility/Infrastructure Coefficient in Sri Lanka, Assuming a DP prior

can be explained by the fact that security forces in these countries may be successful at deterring terrorist attacks, although further measures need to be taken to deter other types of attacks.

There are many extensions of this research that we hope will aid policymakers. For example, policy researchers could potentially use this type of model on a considerably more local level as different regions within a country will almost surely require different security measures. The Iraq troop surge of 2007, for example, was intended to quell violence in Baghdad as well as the Al-Anbar province [12]. A more micro analysis, perhaps at a provincial level, can help policymakers understand what exactly such security measures may be necessary. Another potential avenue for future research could be to estimate a similar Bayesian statistical model presented here and instead examine the impact of different types of weapons on the success of terrorist attacks. This information is available in the rich START database and

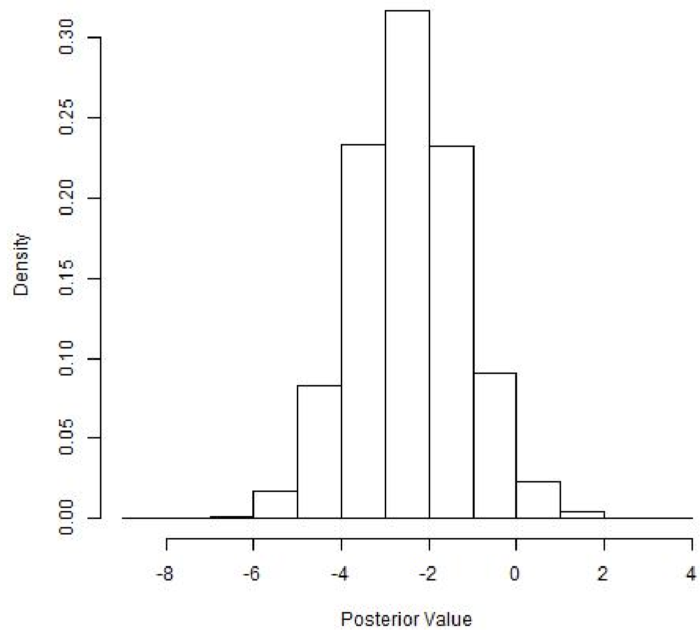

Fig. 5.7: Histogram of Posterior Sample for Hijacking Coefficient in Northern Ireland (before 1998 GF Agreement), Assuming a Normally Distributed Prior

could provide valuable information to policymakers.

Methodologically, there is also a significant amount of future research that we hope that this study will encourage. In particular, four of the eight models in this study applied standard DP priors to weaken the typically restrictive parametric assumptions associated with Bayesian statistical models. An interesting avenue of future research is to apply distance-dependent DP priors, where the mechanism for clustering is guided based on distance between terrorist attacks [8]. Bayesian inferences from models of this nature could provide useful military advice for combating terrorist attacks in the future.

There is also of course no a priori reason to restrict this type of modeling to counterterrorism policy as there are myriads of other applications of Bayesian models ranging in applied economics, business, and professional sports among others [1,3,72]. In this study, we also utilized the semiparametric DRE method to pro-

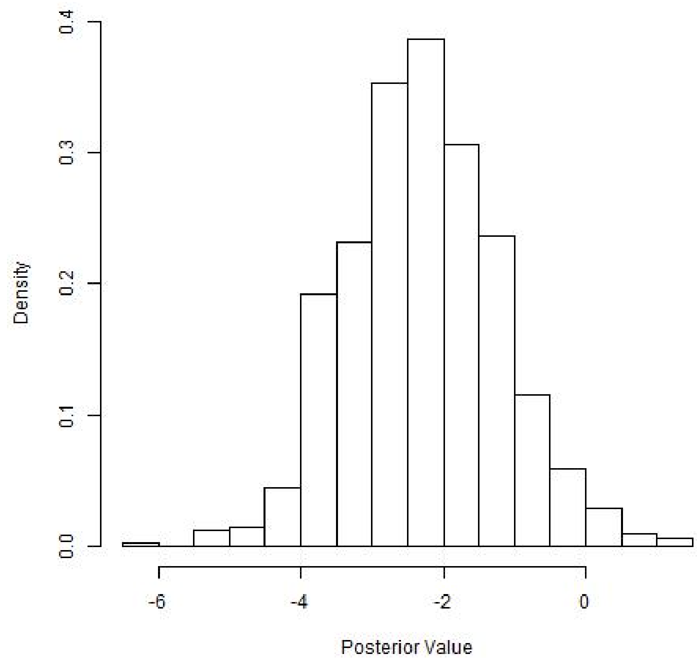

Fig. 5.8: Histogram of Posterior Sample for Hijacking Coefficient in Northern Ireland (before 1998 GF Agreement), Assuming a DP prior

vide more accurate estimation of several Bayesian credible intervals for our posterior coefficients. An interesting avenue of future research could be to use this approach to provide detailed estimates for posterior densities corresponding to other parametric and non-parametric models. Additionally, in this study, we only applied the DRE method to marginal posterior distributions. Another potentially interesting avenue for future research could be to apply this approach to develop Bayesian credible regions for multivariate posterior densities. Finally, we applied this approach to posterior samples, based on significantly larger samples generated by MCMC methods. For high-dimensional models involving large data sets, adequately sampling a posterior density can be quite time consuming and computationally burdensome. Generalized direct sampling is a faster alternative to MCMC that generates independent samples and can more quickly navigate the posterior [10]. Although these independent samples are quite small compared to MCMC samples, associated sta-

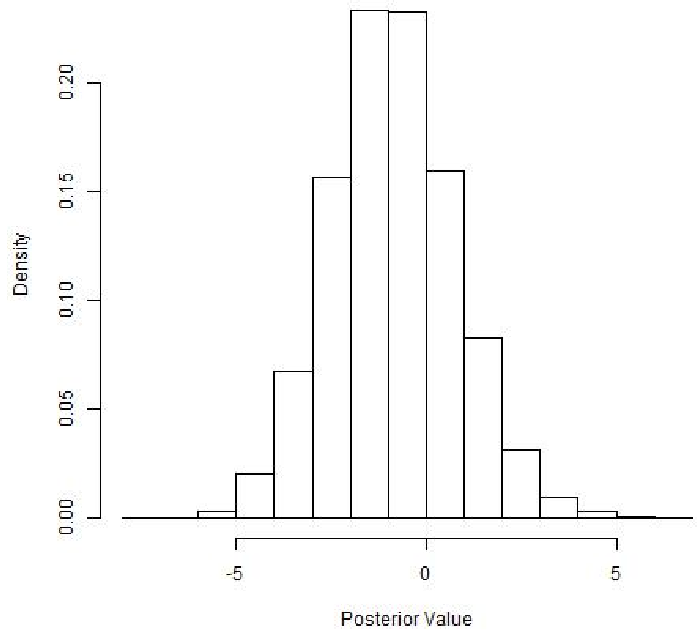

Fig. 5.9: Histogram of Posterior Sample for Unarmed Assault Coefficient in Northern Ireland (after 1998 GF Agreement), Assuming a Normally Distributed Prior

tistical inferences based on such samples can be significantly improved upon by utilizing the semiparametric DRE method.

5.5 Appendix

5.5.1 Variable Definition

Below are descriptions of the variables used in this chapter, as defined in the GTD Codebook [55]. Our dependent variable was whether the terrorist attack was successful.

Dependent Variable Used in Model - Successful attack

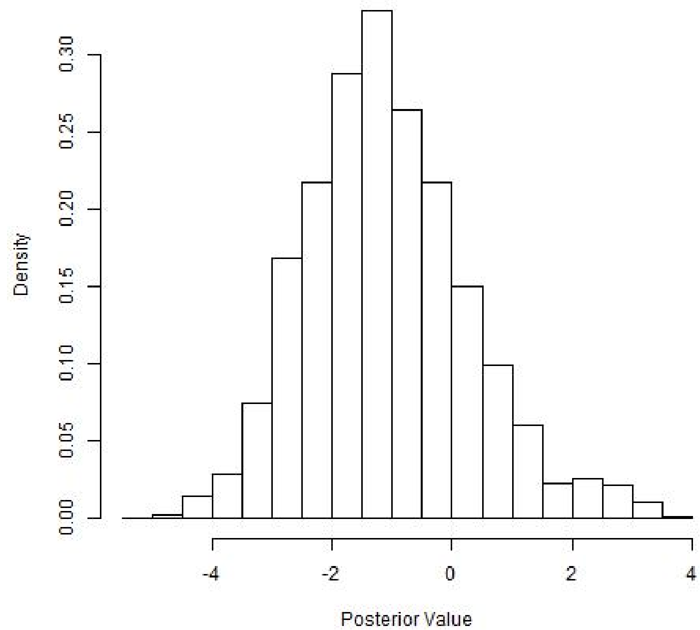

Fig. 5.10: Histogram of Posterior Sample for Unarmed Assault Coefficient in Northern Ireland (after 1998 GF Agreement), Assuming a DP prior

Success - "Success of a terrorist strike is defined according to the tangible effects of the attack. Success is not judged in terms of the larger goals of the perpetrators. For example, a bomb that exploded in a building would be counted as a success even if it did not succeed in bringing the building down or inducing government repression." Please see descriptions under each of the individual explanatory variables below about how this definition is applied to each type of terrorist attack.

Explanatory Variable Used in Model - Suicide Attack

Suicide - "This variable is coded 'Yes' in those cases where there is evidence that the perpetrator did not intend to escape from the attack alive."

Explanatory Variable Used in Model - Types of Terrorist Attacks

Assassination - “An act whose primary objective is to kill one or more specific, prominent individuals. Usually carried out on persons of some note, such as high-ranking military officers, government officials, celebrities, etc. Not to include attacks on non-specific members of a targeted group. The killing of a police officer would be an armed assault unless there is reason to believe the attackers singled out a particularly prominent officer for assassination ... In order for an assassination to be successful, the target of the assassination must be killed. For example, even if an attack kills numerous people but not the target, it is an unsuccessful assassination.”

Armed Assault - “An attack whose primary objective is to cause physical harm or death directly to human beings by use of a firearm, incendiary, or sharp instrument (knife, etc.). Also included under this attack type would be CBRN (chemical, biological, radiological, nuclear) weapons. Not to include acts of a purely personal or criminal nature, or acts in which people are incidentally harmed in pursuit of another primary objective. Not to include attacks involving the use of fists, rocks, sticks, or other handheld (less-than-lethal) weapons ... An armed assault is determined to be successful if the assault takes place and if a target is hit. Unsuccessful armed assaults are those in which the perpetrators attack and do not hit the target. An armed assault is also unsuccessful if the perpetrators are apprehended on their way to commit the assault. To make this determination, however, there must be information to indicate that an actual assault was imminent.”

Bombing/Explosion - “An attack where the primary effects are caused by an energetically unstable material undergoing rapid decomposition and releasing a pressure wave that causes physical damage to the surrounding environment.

Can include either high or low explosives (including a dirty bomb) but does not include a nuclear explosive device that releases energy from fission and/or fusion, or an incendiary device where decomposition takes place at a much slower rate ... A bombing is successful if the bomb or explosive device detonates. Bombings are considered unsuccessful if they do not detonate. The success or failure of the bombing is not based on whether it hit the intended target."

Hijacking - "An act whose primary objective is to take control of a vehicle such as an aircraft, boat, bus, etc. for the purpose of diverting it to an unprogrammed destination, force the release of prisoners, or some other political objective. Obtaining payment of a ransom should not the sole purpose of a Hijacking, but can be one element of the incident so long as additional objectives have also been stated. Hijackings are distinct from Hostage Taking because the target is a vehicle, regardless of whether there are people/passengers in the vehicle ... A hijacking is successful if the hijackers assume control of the vehicle at any point, whereas a hijacking is unsuccessful if the hijackers fail to assume control of the vehicle. The success or failure of the hijacking is not based on whether the vehicle reached the intended destination of the hijackers."

Hostage Taking (Barricade Incident) - "An act whose primary objective is to take control of hostages for the purpose of achieving a political objective through concessions or through disruption of normal operations. Such attacks are distinguished from kidnapping since the incident occurs and usually plays out at the target location with little or no intention to hold the hostages for an extended period in a separate clandestine location ... A barricade incident is successful if the hostage takers assume control of the individuals at any point, whereas a barricade incident is unsuccessful if the hostage takers fail to assume control of the individuals."

Hostage Taking (Kidnapping) - "An act whose primary objective is to take control of hostages for the purpose of achieving a political objective through concessions or through disruption of normal operations. Kidnappings are distinguished from Barricade Incidents (above) in that they involve moving and holding the hostages in another location ... A kidnapping is successful if the kidnappers assume control of the individuals at any point, whereas a kidnapping is unsuccessful if the kidnappers fail to assume control of the individuals."

Facility/Infrastructure - "An act, excluding the use of an explosive, whose primary objective is to cause damage to a non-human target, such as a building, monument, train, pipeline, etc. Such attacks include arson and various forms of sabotage (e.g., sabotaging a train track is a Facility/Infrastructure, even if passengers are killed). Facility/Infrastructures can include acts which aim to harm an installation, yet also cause harm to people incidentally (e.g. an arson attack primarily aimed at damaging a building, but causes injuries or fatalities) ... A facility attack is determined to be successful if the facility is damaged. If the facility has not been damaged, then the attack is unsuccessful."

Unarmed Assault - "An attack whose primary objective is to cause physical harm or death directly to human beings by any means other than explosive, firearm, incendiary, or sharp instrument (knife, etc.) ... An unarmed assault is determined to be successful there is a victim that who has been injured. Unarmed assaults that are unsuccessful are those in which the perpetrators do not injure anyone ... An unarmed assault is also unsuccessful if the perpetrators are apprehended when on their way to commit the assault. To make this determination, however, there must be information to indicate that an assault was imminent."

Explanatory Variable Used in Model - Targets of Terrorist Attacks

Business - "Businesses are defined as individuals or organizations engaged in commercial or mercantile activity as a means of livelihood. Any attack on a business or private citizens patronizing a business such as a restaurant, gas station, music store, bar, caf, etc. This includes attacks carried out against corporate offices or employees of firms like mining companies, or oil corporations. Furthermore, includes attacks conducted on business people or corporate officers. Included in this value as well are hospitals and chambers of commerce and cooperatives. Does not include attacks carried out in public or quasi-public areas such as business district or commercial area, (these attacks are captured under Private Citizens and Property, see below.)"

Government (General) - "Any attack on a government building; government member, former members, including members of political parties in official capacities, their convoys, or events sponsored by political parties; political movements; or a government sponsored institution where the attack is expressly carried out to harm the government. This value includes attacks on judges, public attorneys (e.g., prosecutors), courts and court systems, politicians, royalty, head of state, government employees (unless police or military), election-related attacks, intelligence agencies and spies, or family members of government officials when the relationship is relevant to the motive of the attack."

Police - "This value includes attacks on members of the police force or police installations; this includes police boxes, patrols headquarters, academies, cars, checkpoints, etc. Includes attacks against jails or prison facilities, or jail or prison staff or guards."

Military - "Includes attacks against army units, patrols, barracks, and convoys, jeeps, etc. Also includes attacks on recruiting sites, and soldiers engaged in internal policing functions such as at checkpoints and in anti-narcotics activities. Excludes attacks against non-state militias and guerrillas, these types of attacks are coded as Terrorist/Non-state Militias see below."

Abortion Related - "Attacks on abortion clinics, employees, patrons, or security personnel stationed at clinics."

Airports and Airlines - "An attack that was carried out either against an airplane or against an airport. Attacks against airline employees while on board are also included in this value. Includes attacks conducted against airport business offices and executives. Attacks where airplanes were used to carry out the attack (such as three of the four 9/11 attacks) are not included."

Government (Diplomatic) - "Attacks carried out against foreign missions, including embassies, consulates, etc. This value includes cultural centers that have diplomatic functions, and attacks against diplomatic staff and their families (when the relationship is relevant to the motive of the attack) and property. The United Nations is a diplomatic target."

Educational Institution - "Attacks against schools, teachers, or guards protecting school sites. Includes attacks against university professors, teaching staff and school buses. Moreover, includes attacks against religious schools in this value. As noted below in the Private Citizens and Property value, the GTD has several attacks against students. If attacks involving students are not expressly against a school, university or other educational institution or are carried out in an educational setting, they are coded as private citizens and property. Excludes attacks against military schools (attacks on military schools are coded as Military, see below)."

Food or Water Supply - “Attacks on food or water supplies or reserves are included in this value. This generally includes attacks aimed at the infrastructure related to food and water for human consumption.”

Journalists and Media - “Includes, attacks on reporters, news assistants, photographers, publishers, as well as attacks on media headquarters and offices. Attacks on transmission facilities such as antennae or transmission towers, or broadcast infrastructure are coded as Telecommunications, see below.”

Maritime - “Implies civilian maritime. Includes attacks against fishing ships, oil tankers, ferries, yachts, etc. (Attacks on fishermen are coded as Private Citizens and Property, see below).”

NGO - “Includes attacks on offices and employees of non-governmental organizations (NGOs). NGOs here include large multinational non-governmental organizations such as the Red Cross and Doctors without Borders. Does not include labor unions, social clubs, student groups, and other non-NGO (such cases are coded as Other, see immediately below).”

Other - “This value includes acts of terrorism committed against targets which do not fit into other categories.”

Private Citizens and Property - “This value includes attacks on individuals, the public in general or attacks in public areas including markets, commercial streets, busy intersections and pedestrian malls. Also includes ambiguous cases where the target/victim was a named individual, or where the target/victim of an attack could be identified by name, age, occupation, gender or nationality. This value also includes ceremonial events, such as weddings and funerals.

The GTD contains a number of attacks against students. If these attacks are not expressly against a school, university or other educational institution or are not carried out in an educational setting, these attacks are coded using this value. Also, includes incidents involving political supporters as private citizens and property, provided that these supporters are not part of a government-sponsored event. Finally, this value includes police informers. Does not include attacks causing civilian casualties in businesses such as restaurants, cafes or movie theaters (these categories are coded as Business see above)."

Religious Figures and Institutions - "This value includes attacks on religious leaders, (Imams, priests, bishops, etc.), religious institutions (mosques, churches), religious places or objects (shrines, relics, etc.). This value also includes attacks on organizations that are affiliated with religious entities that are not NGOs, businesses or schools. Attacks on religious pilgrims are considered Private Citizens and Property; attacks on missionaries are considered religious figures."

Telecommunication - "This includes attacks on facilities and infrastructure for the transmission of information. More specifically this value includes things like cell phone towers, telephone booths, television transmitters, radio, and microwave towers."

Terrorists/Non-State Militias - "Terrorists or members of identified terrorist groups within the GTD are included in this value. Membership is broadly defined and includes informants for terrorist groups, but excludes former or surrendered terrorists. This value also includes cases involving the targeting of militias and guerillas."

Tourists - "This value includes the targeting of tour buses, tourists, or 'tours.' Tourists are persons who travel primarily for the purposes of leisure or amusement. Government tourist offices are included in this value. The attack must clearly target tourists, not just an assault on a business or transportation system used by tourists. Travel agencies are coded as business targets.

Transportation (Other than aviation) - "Attacks on public transportation systems are included in this value. This can include efforts to assault public buses, minibuses, trains, metro/subways, highways (if the highway itself is the target of the attack), bridges, roads, etc. The GTD contains a number of attacks on generic terms such as cars or vehicles. These attacks are assumed to be against Private Citizens and Property unless shown to be against public transportation systems. In this regard, buses are assumed to be public transportation unless otherwise noted."

Unknown - "The target type cannot be determined from the available information."

Utilities - "This value pertains to facilities for the transmission or generation of energy. For example, power lines, oil pipelines, electrical transformers, high tension lines, gas and electric substations, are all included in this value. This value also includes lampposts or street lights. Attacks on officers, employees or facilities of utility companies excluding the type of faculties above are coded as business."

Violent Political Parties - "This value pertains to entities that are both political parties (and thus, coded as 'government' in this coding scheme) and terrorists. It is operationally defined as groups that engage in electoral politics and appear as Perpetrators in the GTD."

Chapter 6: Conclusions and Future Research - Where do we go from here?

This dissertation has offered a number of improvements to Bayesian statistical methodologies alongside a number of contributions to public policy research. There are still, however, many other important questions future research should look at. In particular, the tilt functions used in the semiparametric density ratio models called upon in this thesis as well as in other research are chosen a priori by the researcher. Future research could look into a systematic manner, perhaps via variational calculus, of determining the optimal tilt functions. There are also many other statistical methodologies that warrant future research including improvements to other non-MCMC based Bayesian estimation techniques, such as variational Bayesian methods, particle filtering, and generalized direct sampling [6,10,87]. Fast alternatives to MCMC methods are of utmost importance to researchers as the need for analyzing large data sets has grown tremendously in recent years in a number of applied fields.

Additionally, there are myriads of other questions beyond the ones looked at here that could benefit from rigorous analysis, including questions regarding education policy, energy policy, immigration policy, health policy, and macroeconomic modeling. Methodologies similar to those utilized in this dissertation, as well as many of the techniques mentioned above, could be particularly useful in looking at these questions and others.

We emphasize, however, that the ideas presented in this dissertation are not applicable just to public policy research. Regardless of the application, it is always important to be able to properly model real-world phenomena. We hope that this dissertation provides a number of useful tools for applied statisticians to use.

Bibliography

[1] Albright, S. C. A statistical analysis of hitting streaks in baseball. *Journal of the American Statistical Association 88*, 424 (1993), 1175–1183.

[2] Allanach, J., Tu, H., Singh, S., Willett, P., and Pattipati, K. Detecting, tracking, and counteracting terrorist networks via hidden Markov models. In *Aerospace Conference, 2004. Proceedings. 2004 IEEE* (2004), vol. 5, IEEE.

[3] Ansari, A., and Mela, C. F. E-customization. *Journal of Marketing Research 40*, 2 (2003), 131–145.

[4] Antoniak, C. E., et al. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics 2*, 6 (1974), 1152–1174.

[5] Authorization for Use of Military Force Against Iraq Resolution of 2002. Public Law 107-243, 116 Stat. 1498, 2002.

[6] Beal, M. J. *Variational algorithms for approximate Bayesian inference*. Ph.D. Dissertation, University of London, 2003.

[7] Bender, R., and Grouven, U. Using binary logistic regression models for ordinal data with non-proportional odds. *Journal of Clinical Epidemiology 51*, 10 (1998), 809–816.

[8] Blei, D. M., and Frazier, P. I. Distance dependent Chinese restaurant processes. *The Journal of Machine Learning Research 12* (2011), 2461–2488.

[9] Bradlow, E. T., Hardie, B. G., and Fader, P. S. Bayesian inference for the negative binomial distribution via polynomial expansions. *Journal of Computational and Graphical Statistics 11*, 1 (2002).

[10] Braun, M., and Damien, P. Generalized direct sampling for hierarchical Bayesian models. *arXiv preprint arXiv:1108.2245* (2011).

[11] Bucklin, R. E., and Gupta, S. Brand choice, purchase incidence, and segmentation: An integrated modeling approach. *Journal of Marketing Research 29*, 2 (1992).

[12] Bush, G. W. President's address to the nation. `http://georgewbush-whitehouse.archives.gov/news/releases/2007/01/20070110-7.html`, January 2007. [Online; accessed April 17, 2014].

[13] Capretta, J., and Dayaratna, K. D. Compelling evidence makes the case for a market-driven health care system. *Heritage Foundation Backgrounder*, 2867 (2013).

[14] Castles, F. G., and Leibfried, S. *The Oxford Handbook of the Welfare State*. Oxford Handbooks Online, 2010.

[15] Chenoweth, E. Terrorism and democracy. *Annual Review of Political Science 16* (2013), 355–378.

[16] Clinton, B. *My Life*. Random House LLC, 2005.

[17] Congressional Budget Office. The effects of tort reform: Evidence from the states. *A CBO Paper* (June 2004).

[18] Cornyn, J., and Meese, E. Health care and medical malpractice reform: The necessity of reform in the current debate. *Heritage Foundation Lecture 1142*.

[19] Crain, N., Crain, M., McQuillan, L. J., and Abramyan, H. Tort law tally: How state tort reforms affect tort losses and tort insurance premiums. *Pacific Research Institute, San Francisco, CA* (2009).

[20] Dayaratna, K. D. Studies show: Medicaid patients have worse access and outcomes than the privately insured. *Heritage Foundation Backgrounder*, 2740 (2012).

[21] Dayaratna, K. D. Competitive markets in health care: The next revolution. *Heritage Foundation Backgrounder*, 2833 (2013).

[22] Dayaratna, K. D., and Kannan, P. K. A mathematical reformulation of the reference price. *Marketing Letters 23*, 3 (2012), 839–849.

[23] Dayaratna, K. D., and Miller, S. J. First order approximations of the Pythagorean won-loss formula for predicting MLB teams' winning percentages. *By The Numbers: The Newsletter of the SABR Statistical Analysis Committee 22*, 1 (2012), 15–19.

[24] Dayaratna, K. D., and Miller, S. J. The Pythagorean won-loss formula and hockey: A statistical justification for using the classic baseball formula as an evaluative tool in hockey. *The Hockey Research Journal: A Publication of the Society for International Hockey Research* (2013), 193–209.

[25] Department of Health and Human Services, Administration for Children and Families, Office of Family Assistance. Temporary Assistance for Needy Families Program: Eighth annual report to Congress.

[26] Department of Health and Human Services, Administration for Children and Families, Office of Family Assistance. Temporary Assistance for Needy Families Program: Fifth annual report to Congress.

[27] Department of Health and Human Services, Administration for Children and Families, Office of Family Assistance. Temporary Assistance for Needy Families Program: Fourth annual report to Congress.

[28] Department of Health and Human Services, Administration for Children and Families, Office of Family Assistance. Temporary Assistance for Needy Families Program: Ninth annual report to Congress.

[29] Department of Health and Human Services, Administration for Children and Families, Office of Family Assistance. Temporary Assistance for Needy Families Program: Second annual report to Congress.

[30] Department of Health and Human Services, Administration for Children and Families, Office of Family Assistance. Temporary Assistance for Needy Families Program: Seventh annual report to Congress.

[31] Department of Health and Human Services, Administration for Children and Families, Office of Family Assistance. Temporary Assistance for Needy Families Program: Sixth annual report to Congress.

[32] Department of Health and Human Services, Administration for Children and Families, Office of Family Assistance. Temporary Assistance for Needy Families Program: Third annual report to Congress.

[33] Department of Health and Human Services, Administration for Children and Families, Office of Family Assistance. Acf archives. `http://archive.acf.hhs.gov/programs/ofa/data-reports/afdc/rpt3800/3800index.htm`, 1994-1997. [Online; accessed April 17, 2014].

[34] Drehle, D. V., and Smith, J. R. U.S. strikes Iraq for plot to kill Bush. *The Washington Post* (June 1993).

[35] Enders, W. Terrorism: An empirical analysis. *Handbook of Defense Economics 2* (2007), 815–866.

[36] Enders, W., and Sandler, T. Terrorism: Theory and applications. *Handbook of Defense Economics 1* (1995), 213–249.

[37] Escobar, M. D., and West, M. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association 90*, 430 (1995), 577–588.

[38] Esposito, J. L. *Unholy War: Terror in the Name of Islam.* Oxford University Press, 2003.

[39] Everson, P. J., and Bradlow, E. T. Bayesian inference for the beta-binomial distribution via polynomial expansions. *Journal of Computational and Graphical Statistics 11*, 1 (2002).

[40] Ezell, B. C., Bennett, S. P., Von Winterfeldt, D., Sokolowski, J., and Collins, A. J. Probabilistic risk analysis and terrorism risk. *Risk Analysis 30*, 4 (2010), 575–589.

[41] Ferguson, T. S. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* (1973), 209–230.

[42] Fienberg, S. E. Bayesian models and methods in public policy and government settings. *Statistical Science 26*, 2 (2011), 212–226.

[43] Fokianos, K. Merging information for semiparametric density estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66*, 4 (2004), 941–958.

[44] Fokianos, K., Kedem, B., Qin, J., and Short, D. A. A semiparametric approach to the one-way layout. *Technometrics 43*, 1 (2001).

[45] Fokianos, K., and Qin, J. A note on Monte Carlo maximization by the density ratio model. *Journal of Statistical Theory and Practice 2*, 3 (2008), 355–367.

[46] Garner, B. A., and Black, H. C. *Black's Law Dictionary*. St. Paul, MN: Thomson/West, 2004.

[47] Geisser, S., and Eddy, W. F. A predictive approach to model selection. *Journal of the American Statistical Association 74*, 365 (1979), 153–160.

[48] Gelfand, A. E. Model determination using sampling-based methods. In *Markov chain Monte Carlo in practice*. Springer, 1996, pp. 145–161.

[49] Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association 85*, 412 (1990), 972–985.

[50] Gelfand, A. E., and Smith, A. F. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 410 (1990), 398–409.

[51] Gelman, A., and Rubin, D. B. Inference from iterative simulation using multiple sequences. *Statistical Science* (1992), 457–472.

[52] Gilbert, P. B., Lele, S. R., and Vardi, Y. Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika 86*, 1 (1999), 27–43.

[53] Gingrich, N. *Real Change: From the World that Fails to the World that Works*. Regnery Publishing, 2008.

[54] Global Terrorism Database [Data file]. *National Consortium for the Study of Terrorism and Responses to Terrorism (START)*. 2011.

[55] Global Terrorism Database [Data file]. START GTD Database Codebook: Inclusion Criteria and Variables, October 2012.

[56] Greenspan, A. *The Map and the Territory: Risk, Human Nature, and the Future of Forecasting*. Penguin, 2013.

[57] Grimmer, J. An introduction to Bayesian inference via variational approximations. *Political Analysis 19*, 1 (2011), 32–47.

[58] Guadagni, P. M., and Little, J. D. A logit model of brand choice calibrated on scanner data. *Marketing Science 2*, 3 (1983), 203–238.

[59] Guadagni, P. M., and Little, J. D. When and what to buy: A nested logit model of coffee purchase. *Journal of Forecasting 17*, 3-4 (1998), 303–326.

[60] Gupta, S., and Chintagunta, P. K. On using demographic variables to determine segment membership in logit mixture models. *Journal of Marketing Research* (1994), 128–136.

[61] Haskins, R., and Greenberg, M. Welfare Reform: Success or Failure? *Policy and Practice 64*, 1 (2006), 10.

[62] Hawkes, A. G. Spectra of some self-exciting and mutually exciting point processes. *Biometrika 58*, 1 (1971), 83–90.

[63] Hederman, R. S., and Rector, R. The Role of Parental Work in Child Poverty. *Heritage Foundation Center for Data Analysis Report 03*, 01 (2003).

[64] Hiltermann, J. R. *A Poisonous Affair: America, Iraq, and the Gassing of Halabja*. Cambridge University Press, 2007.

[65] Holden, R. T. The contagiousness of aircraft hijacking. *American Journal of Sociology* (1986), 874–904.

[66] Holden, R. T. Time series analysis of a contagious process. *Journal of the American Statistical Association 82*, 400 (1987), 1019–1026.

[67] Hymowitz, K., Carroll, J., Wilcox, W., and Kaye, K. Knot Yet: The Benefits and Costs of Delayed Marriage in America. `http://twentysomethingmarriage.org/`, 2013. [Online; accessed April 17, 2014].

[68] Iraq Liberation Act of 1998. Public Law 105–338, 112 Stat. 3178, 1998.

[69] Jackson, O. A. The impact of the 9/11 terrorist attacks on the US economy. *Journal of 9/11 Studies 20* (2008), 1–27.

[70] Jara, A., Hanson, T. E., Quintana, F. A., Müller, P., and Rosner, G. L. DPpackage: Bayesian semi and nonparametric modeling in R. *Journal of Statistical Software 40*, 5 (2011), 1.

[71] Jha, M. K. Dynamic Bayesian network for predicting the likelihood of a terrorist attack at critical transportation infrastructure facilities. *Journal of Infrastructure Systems 15*, 1 (2009), 31–39.

[72] Jochmann, M., and Leon-Gonzalez, R. Estimating the demand for health care with panel data: A semiparametric Bayesian approach. *Health Economics 13*, 10 (2004), 1003–1014.

[73] Jones, A. Should tort reform be part of health care reform? A "liberal" thinks so. *The Wall Street Journal* (May 2010).

[74] Kamakura, W. A., and Russell, G. J. A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research* (1989), 379–390.

[75] Kaplan, E. H., Mintz, A., and Mishal, S. Tactical prevention of suicide bombings in Israel. *Interfaces 36*, 6 (2006), 553–561.

[76] Kaplan, E. H., Mintz, A., Mishal, S., and Samban, C. What happened to suicide bombings in Israel? Insights from a terror stock model. *Studies in Conflict & Terrorism 28*, 3 (2005), 225–235.

[77] Kaplan, S. Applying the general theory of quantitative risk assessment (QRA) to terrorism risk. *Risk-Based Decision Making* (2002), 77–81.

[78] Katzoff, M., Zhou, W., Khan, D., Lu, G., and Kedem, B. Out of sample fusion in risk prediction. *Journal of Statistical Theory and Practice Forthcoming* (2013).

[79] Kean, T. *The 9/11 commission report: Final report of the national commission on terrorist attacks upon the United States*. Government Printing Office, 2011.

[80] Kedem, B., Kim, E.-y., Voulgaraki, A., and Graubard, B. I. Two-dimensional semiparametric density ratio modeling of testicular germ cell data. *Statistics in Medicine 28*, 16 (2009), 2147–2159.

[81] Kedem, B., Lu, G., Wei, R., and Williams, P. D. Forecasting mortality rates via density ratio modeling. *Canadian Journal of Statistics 36*, 2 (2008), 193–206.

[82] Kedem, B., Pan, L., and Zhou, W. Out of sample fusion: A Monte Carlo method. *Proceedings of the 2014 IEEE Conference on Computational Science and Computational Intelligence* (2014), 364–367.

[83] Kedem, B., and Wen, S. Semi-parametric cluster detection. *Journal of Statistical Theory and Practice 1*, 1 (2007), 49–72.

[84] Kennedy, D. M., Cohen, L., and Bailey, T. A. *The American Pageant*. Houghton Mifflin Harcourt, 2005.

[85] Kessler, D., and McClellan, M. Do doctors practice defensive medicine? *The Quarterly Journal of Economics 111*, 2 (1996), 353–390.

[86] Kessler, D. P., Sage, W. M., and Becker, D. J. Impact of malpractice reforms on the supply of physician services. *Journal of the American Medical Association 293*, 21 (2005), 2618–2625.

[87] Khan, Z., Balch, T., and Dellaert, F. MCMC-based particle filtering for tracking a variable number of interacting targets. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 27*, 11 (2005), 1805–1819.

[88] Kottas, A., Duan, J. A., and Gelfand, A. E. Modeling disease incidence data with spatial and spatio temporal Dirichlet process mixtures. *Biometrical Journal 50*, 1 (2008), 29–42.

[89] Kydd, A. H., and Walter, B. F. The strategies of terrorism. *International Security 31*, 1 (2006), 49–80.

[90] Kyung, M., Gill, J., and Casella, G. New findings from terrorism data: Dirichlet process random-effects models for latent groups. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 60*, 5 (2011), 701–721.

[91] Lazar, N. A. Bayesian empirical likelihood. *Biometrika 90*, 2 (2003), 319–326.

[92] Lenk, P. J., and DeSarbo, W. S. Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika 65*, 1 (2000), 93–119.

[93] Levy, R. Do's and dont's of tort reform. *Cato Institute Commentary* (May 2005).

[94] Lewis, E., Mohler, G., Brantingham, P. J., and Bertozzi, A. L. Self-exciting point process models of civilian deaths in Iraq. *Security Journal 25*, 3 (2012), 244–264.

[95] Li, Q., and Schaub, D. Economic globalization and transnational terrorism: A pooled time-series analysis. *Journal of Conflict Resolution 48*, 2 (2004), 230–258.

[96] Lu, G. *Asymptotic theory for multiple-sample semiparametric density ratio models and its application to mortality forecasting.* Ph.D. Dissertation, University of Maryland, 2007.

[97] Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. WinBUGS - A Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing 10*, 4 (2000), 325–337.

[98] MacEachern, S. N., and Müller, P. Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics 7*, 2 (1998), 223–238.

[99] McFadden, D. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics* (1973), 105–142.

[100] McFadden, D., and Train, K. Mixed MNL models for discrete response. *Journal of Applied Econometrics 15*, 5 (2000), 447–470.

[101] McGeough, P. Witness won't let Saddam intimidate him. *The Sydney Morning Herald* (December 2005).

[102] McKittrick, D., and McVea, D. *Making Sense of the Troubles: The Story of the Conflict in Northern Ireland.* New Amsterdam Books, 2002.

[103] McQuillan, L. J. *Jackpot Justice: The True Cost of America's Tort System.* Pacific Research Institute, 2007.

[104] McQuillan, L. J., and Abramyan, H. U.S. tort liability index: 2010 report. *Pacific Research Institute, San Francisco, CA* (2010).

[105] McShane, B., Adrian, M., Bradlow, E. T., and Fader, P. S. Count models based on Weibull interarrival times. *Journal of Business & Economic Statistics 26*, 3 (2008).

[106] Mengersen, K. L., Pudlo, P., and Robert, C. P. Bayesian computation via empirical likelihood. *Proceedings of the National Academy of Sciences 110*, 4 (2013), 1321–1326.

[107] Miller, S. J. A derivation of the Pythagorean Won-Loss Formula in baseball. *Chance 20*, 1 (2007), 40–48.

[108] Miller, S. J., Bradlow, E. T., and Dayaratna, K. Closed-form Bayesian inferences for the logit model via polynomial expansions. *Quantitative Marketing and Economics 4*, 2 (2006), 173–206.

[109] Miller, S. J., and Takloo-Bighash, R. *An Invitation to Modern Number Theory*, vol. 13. Princeton University Press, 2006.

[110] Morris, C. N. Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association 78*, 381 (1983), 47–55.

[111] Murray, C. *Losing Ground: American Social Policy, 1950-1980.* Basic books, 2008.

[112] Neal, R. M. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics 9*, 2 (2000), 249–265.

[113] Nguyen, V.-A., Boyd-Graber, J., Resnik, P., Cai, D. A., Midberry, J. E., and Wang, Y. Modeling topic control to detect influence in conversations using nonparametric topic models. *Machine Learning* (2013), 1–41.

[114] Northern Ireland Peace Agreement (The Good Friday Agreement). `http://peacemaker.un.org/sites/peacemaker.un.org/files/IE%20GB_980410_Northern%20Ireland%20Agreement.pdf`, April 1998. [Online; accessed April 19, 2014].

[115] Ntzoufras, I. *Bayesian Modeling using WinBUGS*, vol. 698. John Wiley & Sons, 2011.

[116] Owen, A. B. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*, 2 (1988), 237–249.

[117] Paté-Cornell, E. Fusion of intelligence information: A Bayesian approach. *Risk Analysis 22*, 3 (2002), 445–454.

[118] Personal Responsibility and Work Opportunity Reconciliation Act of 1996. Public Law 104–193, 110 Stat. 2105, 1996.

[119] Phue, J.-N., Kedem, B., Jaluria, P., and Shiloach, J. Evaluating microarrays using a semiparametric approach: Application to the central carbon metabolism of *Escherichia coli* BL21 and JM109. *Genomics 89*, 2 (2007), 300–305.

[120] Popp, R. L., and Yen, J. *Emergent Information Technologies and Enabling Policies for Counter-terrorism*, vol. 6. John Wiley & Sons, 2006.

[121] Porter, M. D., White, G., et al. Self-exciting hurdle models for terrorist activity. *The Annals of Applied Statistics 6*, 1 (2012), 106–124.

[122] Prentice, R. L., and Pyke, R. Logistic disease incidence models and case-control studies. *Biometrika 66*, 3 (1979), 403–411.

[123] Qin, J., and Lawless, J. Estimating equations, empirical likelihood and constraints on parameters. *Canadian Journal of Statistics 23*, 2 (1995), 145–159.

[124] Qin, J., and Zhang, B. A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika 84*, 3 (1997), 609–618.

[125] Qin, J., and Zhang, B. Density estimation under a two-sample semiparametric model. *Nonparametric Statistics 17*, 6 (2005), 665–683.

[126] Rapoport, D. C. Fear and trembling: Terrorism in three religious traditions. *The American Political Science Review* (1984), 658–677.

[127] Reagan, R. *A Time for Choosing: The Speeches of Ronald Reagan, 1961-1982*. Gateway Books, 1983.

[128] Rector, R. Undermining true welfare reform. *Heritage Foundation Commentary 14*, 3, 283–298.

[129] Rector, R. Welfare reform, dependency reduction, and labor market entry. *Journal of Labor Research 14*, 3 (1993), 283–297.

[130] Rector, R. Marriage: America's greatest weapon against child poverty. *Heritage Foundation Backgrounder*, 2465 (2010).

[131] Rector, R., Bradley, K., and Sheffield, R. Obama to spent $10.3 trillion on welfare. *Heritage Foundation Special Report SR*, 67 (2009).

[132] Rector, R., and Fagan, P. The continuing good news about welfare reform. *Heritage Foundation Backgrounder*, 1620 (2003).

[133] Rector, R., and Johnson, K. Roles of couples' relationship skills and fathers' employment in encouraging marriage. *Heritage Foundation Center for Data Analysis Report 04*, 14 (2004).

[134] Rector, R., Johnson, K., and Noyes, L. R. Teenage sexual abstinence and academic achievement. *Heritage Foundation Center for Data Analysis Report 03*, 04 (2003).

[135] Rector, R., and Marshall, J. The unfinished work of welfare reform. *National Affairs* (2013).

[136] Rector, R., and Sheffield, R. Understanding poverty in the United States: Surprising facts about America's poor. *Heritage Foundation Backgrounder*, 2607 (2011).

[137] Rector, R., and Sheffield, R. Ending work for welfare: An overview. *Heritage Foundation Issue Brief*, 3712 (2012).

[138] Revelt, D., and Train, K. Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics 80*, 4 (1998), 647–657.

[139] Rivlin, A. Curing health care: the next president should complete, not abandon, Obama's reform. *Brookings Institution Campaign 2012 Policy Brief* (May 2012).

[140] Rossi, P. E., and Allenby, G. M. Bayesian Statistics and Marketing. *Marketing Science 22*, 3 (2003), 304–328.

[141] Rossi, P. E., Allenby, G. M., and McCulloch, R. E. *Bayesian Statistics and Marketing.* J. Wiley & Sons, 2005.

[142] Rubin, P. H., and Shepherd, J. M. Tort reform and accidental deaths. *Journal of Law and Economics 50*, 2 (2007), 221–238.

[143] Sanchez-Cuenca, I., and De la Calle, L. Domestic terrorism: The hidden side of political violence. *Annual Review of Political Science 12* (2009), 31–49.

[144] Sandler, T., et al. Terrorism & game theory. *Simulation & Gaming 34*, 3 (2003), 319–337.

[145] Sawhill, I. V., and Haskins, R. *Work and Marriage: The Way to End Poverty and Welfare.* Welfare Reform & Beyond Initiative, Brookings Institution, 2003.

[146] Schennach, S. M. Bayesian exponentially tilted empirical likelihood. *Biometrika 92*, 1 (2005), 31–46.

[147] Schmittlein, D. C., Morrison, D. G., and Colombo, R. Counting your customers: Who-are they and what will they do next? *Management Science 33*, 1 (1987), 1–24.

[148] Seifert, K. R., and McCauley, C. Suicide bombers in Iraq, 2003–2010: disaggregating targets can reveal insurgent motives and priorities. *Terrorism and Political Violence Forthcoming*.

[149] Shiryaev, A. N. *Probability.* Springer-Verlag, New York, 1996.

[150] Smith, K. R., and Zick, C. D. The incidence of poverty among the recently widowed: Mediating factors in the life course. *Journal of Marriage and the Family* (1986), 619–630.

[151] Studdert, D. M., Mello, M. M., Sage, W. M., DesRoches, C. M., Peugh, J., Zapert, K., and Brennan, T. A. Defensive medicine among high-risk specialist physicians in a volatile malpractice environment. *Journal of the American Medical Association 293*, 21 (2005), 2609–2617.

[152] Tanner, M., and DeHaven, T. TANF and federal welfare. *Downsizing the Federal Government* (September 2010).

[153] Tenet, G., and Harlow, B. *At the Center of the Storm: My Years at the CIA.* HarperPress, 2007.

[154] Tierney, L. Markov chains for exploring posterior distributions. *The Annals of Statistics* (1994), 1701–1728.

[155] Tranchita, C., Hadjsaid, N., and Torres, A. Ranking contingency resulting from terrorism by utilization of Bayesian networks. In *Electrotechnical Conference, 2006. MELECON 2006. IEEE Mediterranean* (2006), IEEE, pp. 964–967.

[156] Unicon Research Corporation [producer and distributor of CPS Utilities]. Current Population Surveys, March 1962-2009 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor]. Los Angeles, CA., 2009.

[157] US Department of Health and Human Services and others. Quarterly Public Assistance Statistics. *Washington DC: HHS, Administration on Children and Families* (1982-1993).

[158] US Department of State Office of the Coordinator for Counterterrorism. Country Reports on Terrorism 2004. `http://www.state.gov/documents/organization/45313.pdf`, April 2005. [Online; accessed April 17, 2014].

[159] Voulgaraki, A. *Semiparametric Regression and Mortality Rate Prediction.* Ph.D. Dissertation, University of Maryland, 2011.

[160] Voulgaraki, A., Kedem, B., Graubard, B. I., et al. Semiparametric regression in testicular germ cell data. *The Annals of Applied Statistics 6*, 3 (2012), 1185–1208.

[161] Weaver, M., and Chamberlain, G. Sri Lanka declares end to war with Tamil Tigers. *The Guardian 19* (2009).

[162] White, G., and Porter, M. D. GPU accelerated MCMC for modeling terrorist activity. *Computational Statistics & Data Analysis 71* (2014), 643–651.

[163] White, G., Porter, M. D., and Mazerolle, L. Terrorism Risk, Resilience and Volatility: A Comparison of Terrorism Patterns in Three Southeast Asian Countries. *Journal of Quantitative Criminology 29*, 2 (2013), 295–320.

[164] Wisner, D. H. A stepwise logistic regression analysis of factors affecting morbidity and mortality after thoracic trauma: effect of epidural analgesia. *Journal of Trauma-Injury, Infection, and Critical Care 30*, 7 (1990), 799–805.

[165] Yang, Y., He, X., et al. Bayesian empirical likelihood for quantile regression. *The Annals of Statistics 40*, 2 (2012), 1102–1131.

[166] Young, J. K., and Findley, M. G. Promise and pitfalls of terrorism research. *International Studies Review 13*, 3 (2011), 411–431.

[167] Zhang, J., and Kai, F. Y. What's the relative risk?: A method of correcting the odds ratio in cohort studies of common outcomes. *Journal of the American Medical Association 280*, 19 (1998), 1690–1691.

[168] Zhou, W. *Out of Sample Fusion*. Ph.D. Dissertation, University of Maryland, 2012.