UMBC

DISSERTATION APPROVAL SHEET

Title of Dissertation: Statistical Modeling using Conditionally Specified Joint Distributions with Applications

Name of Candidate: Nadeesri Wijekoon

Doctor of Philosophy, 2021

Graduate Program: Statistics

Dissertation and Abstract Approved:

Nagaraj Neerchal

Nagaraj Neerchal

Professor

Mathematics and Statistics

11/19/2021 | 10:37:52 PM EST

NOTE: *The Approval Sheet with the original signature must accompany the thesis or dissertation. No terminal punctuation is to be used.

ABSTRACT

Title of dissertation: Statistical Modeling using Conditionally Specified Joint Distributions with Applications

Nadeesri Wijekoon, Doctor of Philosophy, 2021

Dissertation directed by: Professor Nagaraj K. Neerchal
Department of Mathematics and Statistics

Often in practice, conditional distributions are easier to model and interpret while the joint distribution itself is either intractable or not available in closed form. When the observed response consists of both continuous and discrete components, specifying conditionals is more convenient. There are many real-world applications where the conditional specification approach is intuitively appealing, and knowing the conditional distributions makes it easier to understand and visualize the joint distribution. Furthermore, the researcher can obtain a better insight by investigating and interpreting the conditional distributions. In this thesis, we propose a joint distribution that can be specified using its respective conditionals and which can handle both continuous data and discrete data together. In literature, such models are referred to as conditionally specified models. We explored the theoretical aspects of conditionally specified models, where conditionals are from the exponential family of distributions, including parameter estimation, data generation, and uniqueness of the joint distributions.

The Maximum Likelihood (ML) method, which is the preferred estimation

method of parametric models, turns out to be difficult to implement for estimating the parameters of conditionally specified joint distributions because it contains an awkward normalizing constant. Thus, Composite Likelihood (CL) was used as an alternative method of estimation. We used numerical methods to obtain the estimates of parameters since closed-form expressions for estimates using the proposed density are not feasible. Simulation studies were conducted for different sample sizes to investigate the properties of ML estimates and CL. It showed that the ML method has less bias (and nearly zero in some cases) than the CL method, however CL method involves relatively less computational burden. In both methods, the variances of the estimates decrease as the sample size increases. Further, joint asymptotic relative efficiency (JARE) between the ML method and CL method were calculated for different sample sizes using the Godambe Information matrix. In addition, we conducted a performance analysis utilizing the two methods. The results showed that for a larger sample size, the computational advantage of the CL method surpasses that of the ML method quickly. Thus, choosing the CL method over the ML method is a trade-off between efficiency and computational cost. The proposed normal-logistic joint density was applied to the stock prices (continuous data) and expert recommendations (categorical data) for buying/selling specific stocks. Parameters of the model were estimated using both ML and CL methods.

STATISTICAL MODELING USING CONDITIONALLY SPECIFIED JOINT DISTRIBUTIONS WITH APPLICATIONS

by

Nadeesri Wijekoon

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland Baltimore County in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor Nagaraj K. Neerchal, Chair/Advisor
Professor Bimal Sinha
Professor Thomas Matthew
Professor Anindya Roy
Dr. Mathangi Gopalakrishnan
Dr. Andrew Raim

Dedicated to my parents

Acknowledgments

First I'd like to thank my advisor, Professor Nagaraj K. Neerchal for giving me the opportunity to work on challenging and interesting projects over the past years. His advice and guidance tremendously helped me to complete this dissertation. I would also like to thank Professor Bimal Sinha, Professor Thomas Matthew, Professor Anindya Roy, Dr. Mathangi Gopalakrishnan, and Dr. Andrew Raim for agreeing to serve on my thesis committee. I'm thankful not only for their invaluable feedback on improving my research work but also for Special thanks goes to Dr. Mathangi Gopalakrishnan and Dr. Andrew Raim for sparing their invaluable time reviewing the dissertation and giving extensive comments and feedback on my thesis.

I would also like to thank my other teachers and staff members of the statistics faculty who shaped my graduate education. I'm forever grateful for Dr. Kofi Adragni, Dr. Yi Huang, Dr. DoHwan Park, Dr. Junyong Park, Dr. Yaakov Malinovsky, Dr. Yehenew Kifle, Dr. Nagaraj Neerchal, Dr. Thu Nguyen, and Dr. Elizabeth Stanwyck for being such amazing educators. Also, my deepest gratitude goes towards Ms. Raji Baradwaj; you are one of the amazing people I met in my life. I will never forget how you took me under your wing and trained me and guided me to be a better educator.

I'm forever grateful to my teachers at Mahamaya Girls' College, Kandy, and the University of Peradeniya, Sri Lanka for giving me a solid foundation in the field of Statistics and also to make me a better person.

The past six years of my life at UMBC gave me not only knowledge but also friendships that I will cherish for a lifetime. The graduate school will be difficult for me to survive without the help of fellow students Neha Agarwala, Sumaya Alzuhairy, Vahid Andalib, Aryana Arsham, Rowena Bastero, Lillian Chow, Rabab Elnaiem, Zhou Feng, Iris Ivy Gauran, Nivya George, Abhishek Guin, Gaurab Hore, Mina Hosseini, Yewon Kim, Eswar Kumar, Reetam Majumder, Alain Moluh, Luan Chip Nguyen, Janita Patwardhan, Sai Kumar Popuri, Mark Ramos, Michael Retzlaf, Jing Wang, Mingkai Yu, and John Zylstra.

I would also like to give a special thank you to the department staff members Janet Burgee, Maggie Kennedy, Marshall Turner, and Boris Alemi for being concerned about the needs of graduate students.

I owe my deepest thanks to my family - I'm indebted to my mother and father who have always stood by me and guided me through my career, and have pulled me through against impossible odds at times. Without you, I won't be able to reach my goal. My brother and my sister-in-law never failed to show their love and support throughout my Ph.D. training. I also like to thank Nipuni Adhikari, Achala Dengamage, Neranga Hannadige, and Prabashana Rathnayake for their love and support throughout these past six years.

Finally, I thank my husband Chamara and baby daughter Avini for loving me and supporting me throughout this endeavor. I am truly blessed to have you two in my life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CS	Conditionally Specified
CEF	Conditionally Specified models in Exponential Families
UMR	Uniform marginal representation
MGF	Moment Generating Function
ML	Maximum Likelihood
PL	Pseudolikelihood
CL	Composite Likelihood
LS	Least Square
MOM	Method of Moments
JARE	Joint Asymptotic Relative Efficiency
GIM	Godambe Information Matrix
MLE	Maximum Likelihood Estimation
PLE	Pseudolikelihood Estimation
CLE	Composite Likelihood Estimation

Chapter 1: Introduction

In many real-world applications, the data consists of both continuous and discrete observations. Specifying joint distributions for such data presents some challenges whereas the conditionals of the continuous part given the discrete and vice versa are easily specified. Specifying the joint distribution in terms of conditionals is also convenient in many other situations. In many cases, the conditionals are commonly used distributions whereas the joint distribution may belong to an intractable family.

In this dissertation, we focused on a joint distribution for data consisting of a binary random variable and bivariate continuous measurements. More specifically, the model postulates a joint distribution determined by Binomial (in fact by a logistic regression model) conditional and a bivariate Normal conditional. This is a special case of a well-known and well-developed literature on conditional specified joint distributions. We provide a complete study of the important special case which is bivariate normal and logistic conditionals. We also address problem of compatibility and we present a real-world application and explore different methods of estimation. In the next three sections, We present two real-world examples where bivariate normal-logistic conditionals can be applied.

1.1 Example 1: Proxy data in aging studies

In aging studies, researchers are interested in identifying the mutable factors related to the disease/disabilities of older adults. However, some of these factors cannot be directly quantifiable. Therefore, researchers have developed various measurement scales Shardell et al. (2010) and/or questionnaires to obtain the required information. Yet, Some older adults may be unwilling or unable to provide necessary information about the disease/disability due to reasons such as medical issues or declining cognitive ability. As a solution, a proxy is asked to respond. A proxy is someone who knows the older adult such as a relative or caregiver. There are advantages and disadvantages of using a proxy. Given below in Table 1.1 is a schematic representation of a longitudinal study Hosseini (2017). Typically, proxy data are not collected if the subject is a self-report. Therefore, for each subject, either the subject or proxy, respond Shardell et al. (2010). This data structure is shown in the Table 1.2. Note that, either the self or proxy observation can be seen in each time point.

As we further investigate the problem, we can posit that the probability distribution of subject responses is different from the probability distribution of the proxy responses. On the other hand, whether or not the subject will require a proxy or not is also dependent on the condition of the subject and hence the past and present values of the response variable. Suppose for the i^{th} subject, $i = 1, 2, ..., n$, the response at the t^{th} time point $t = 1, 2, ..., T$, Y_{it} denotes the response and an indicator R_{it} represents whether the response is truly coming from a proxy ($R_{it} = 0$)

Table 1.1: Self data and Proxy data in a longitudinal study *[Hosseini. M.,2017]*. S = "Self" and P= "Proxy".

Subject ID	Time			
	1	2	$\cdots$	T
1	S	S	P	P
2	S	P	P	P
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	S	P	S	S

or from the subject themselves ($R_{it} = 1$). That is,

$$R_{it} = \begin{cases} 1 & \text{if answered by subject} \\ 0 & \text{if answered by proxy} \end{cases}$$

As shown in Table 1.2, for each time point we have two data points; one is the response and the other is the indicator of whether the response is from a proxy or self. Moreover, typically the response distribution is continuous while the proxy indicator clearly follows a discrete distribution. Since Y depends on R and R depends on Y, we can also state the problem as follows.

Suppose We have the distribution of observed data vector of a single subject (say Y) given the self-proxy pattern (say R), $f_{Y|R}(y|R = r)$ and the distribution of

Table 1.2: Self and Proxy data Structure

Subject ID	Time			
	1	2	.....	T
1	(Y_{11}, R_{11})	(Y_{12}, R_{12})	.....	(Y_{1T}, R_{1T})
2	(Y_{21}, R_{21})	(Y_{22}, R_{22})	.....	(Y_{2T}, R_{2T})
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	(Y_{n1}, R_{n1})	(Y_{n2}, R_{n2})	.....	(Y_{nT}, R_{nT})

self proxy pattern given observed data vector of the subject, $f_{R|Y}(r|Y = y)$. Further, R is a binary variable and Y is a continuous variable such that $f_{Y|R}(y|R = 0) \propto f_0(y)$ and $f_{Y|R}(y|R = 1) \propto f_1(y)$. Further if we let,

$$f_{R|Y}(r|Y = y) = g^{-1}(y, \alpha)^r \{1 - g^{-1}(y, \alpha)\}^{1-r}$$

and

$$f_{Y|R}(y|R = r) = I(r = 0) f_0(y) + I(r = 1) f_1(y)$$

We will have that $g(y, \alpha)$ is a function of y parameterized by α. Detailed proof of the above form can be found in Hosseini (2017). The current approach of analyzing this type of data is either to treat the self and proxy data as interchangeable or simply analyze self and proxy data separately (Shardell et al. (2010), Snow et al. (2005)). The first approach leads to biased estimates and incorrect standard errors whereas the latter results in two sets of estimates without an obvious way of combining them.

Thus, there is a need to develop a single framework to analyze both subject data and proxy data. A single framework, in a sense, is a joint distribution of (Y, R).

1.2 Example 2: Synthetic data in survey sampling

Synthetic data are generated to fill specific observations that may be missing in the original, real data. Synthetic data are also used in publishing survey data subject to confidentiality requirement. Missing data is one of the places where synthetic data are used. Missing data can occur due to non-response. In longitudinal studies, this situation is very common and there are many ways such as partial imputation and interpolation to replace these missing observations with imputed values. Similar to in Example 1, we can define an indicator function R_i as follows.

$$R_i = \begin{cases} 1 & \text{if original observation} \\ 0 & \text{if imputed observation} \end{cases}$$

The data structure is given in Table 1.3.

Table 1.3: Toy data set with imputed values. (Bold observations represent the imputed data.)

i	1	2	3	4	5
Y_i	0.76	1.45	**3.54**	7.63	**0.23**
R_i	1	1	**0**	1	**0**

One can analyze these data assuming the data as interchangeable or simply analyze

original data and self data separately. The first approach leads to biased estimates and incorrect standard errors whereas the latter results in two sets of estimates without an obvious way of combining them. Thus, as in Example 1, there is a need to develop a single framework to analyze both original data and imputed data, so that the information contained in the label can be incorporated into the data analysis.

1.3 Example 3: Stock market recommendations

Stock market analysts classify a stock as either a "buy", "hold" or a "sell" based on their research. This research includes the price history of the stock and also the current status of the market. Analyst can decide that a particular stock is a "buy", " sell", or "hold" depending on their research (Grant (2020)). A "buy" recommendation means that the analyst is expecting the stock price to go up and a "sell" recommendation is given when the analyst is expecting the stock price to go down. A "hold" recommendation is simply an indication that it is not the right time to sell if you have the stock, and perhaps is also not the right time to acquire more of this stock.

Assume we have a data of stock prices over $t = 1, 2, ..., T$ for $s = 1, 2, ..., S$ stocks. Suppose Y_{st} is the price of s^{th} stock at time t^{th} and the indicator function R_{st} that specifies the recommendation by the analyst. That is,

$$R_{st} = \begin{cases} 1 & \text{if stock is a “buy”,} \\ 2 & \text{if stock is a “hold”,} \\ 3 & \text{if stock is a “sell”.} \end{cases}$$

Here, the distribution of the stock price on the “buy” days, the distribution of the stock price on the “hold” days are clearly different from the price distribution on “sell” days. Thus, our choice of the model parameters would differ if we knew that the stock has been classified as a buy instead of as a sell. On the other hand, the price history of the stock will influence the classification (buy, hold or sell) decision of the analyst. Thus, even though it is cumbersome and inconvenient to think of a joint distribution of the stock prices (continuous) and the analyst recommendations (discrete), it is easier to think of the conditional distribution of the recommendations given the price history and the conditional distribution of the stock price given its buy, hold or sell status. Moreover, if we assume we only have “buy” status and “sell” status, the R_{st} is further simplified to a binary indicator as

$$R_{st} = \begin{cases} 1 & \text{if stock is a “buy”,} \\ 0 & \text{if stock is a “sell”.} \end{cases}$$

Though stock prices are usually published every working day, if we use only Monday and Friday stock prices the data structure boils down to more a simplified form as shown in Table (1.4). Thus, the distribution of the observed data vector of a single stock (say y) given the buy-sell pattern (say R) is denoted by $f_{y|R}(y|X = x)$ and the distribution of the buy-sell pattern given the observed data vector of the stocks

Table 1.4: Data structure

		Closing Prices (Monday, Friday)	Recommendation
Stock 1	Week 1	$(y_{11}^{(1)}, y_{11}^{(2)})$	r_{11}
	Week 2	$(y_{12}^{(1)}, y_{12}^{(2)})$	r_{12}
	$\vdots$	$\vdots$	$\vdots$
	Week 40	$(y_{1,40}^{(1)}, y_{1,40}^{(2)})$	$r_{1,40}$
Stock 2	Week 1	$(y_{21}^{(1)}, y_{21}^{(2)})$	r_{21}
	Week 2	$(y_{22}^{(1)}, y_{22}^{(2)})$	r_{22}
	$\vdots$	$\vdots$	$\vdots$
	Week 40	$(y_{2,40}^{(1)}, y_{2,40}^{(2)})$	$r_{2,40}$
Stock_i	week_j	$(y_{ij}^{(1)}, y_{ij}^{(2)})$	r_{ij}

is denoted by $f_{R|Y}(R|Y = y)$. Further, suppose R is a binary variable and Y is a continuous variable such that $f_{Y|R}(y|R = 0) \propto f_0$ and $f_{Y|R}(y|R = 1) \propto f_1$. Further more, we may assume a linear function $g(y, \beta)$, parameterized by β, to model the success probability of r. Then, the conditional distribution of r given $Y = y$ is given

by

$$f_{R|Y}(r|Y=y) = g^{-1}(y,\beta)^r(1-g^{-1}(y,\beta))^{1-r}.$$

One can assume f_0 follows a bivariate normal distribution with mean vector $\underset{\sim}{\mu}^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)})$ and variance covariance matrix as $\Sigma^{(0)}$ and f_1 follows a bivariate normal distribution with mean vector $\underset{\sim}{\mu}^{(1)} = (\mu_1^{(1)}, \mu_2^{(1)})$ and variance covariance matrix as $\Sigma^{(1)}$. We can then write

$$f_{Y|R}(y|R=r) \sim N_2(\underset{\sim}{\mu}^{(r)}, \Sigma^{(r)})$$

We discuss this example more extensively in chapter 5 accompanied by data analysis.

1.4 Organization of the thesis

The rest of the thesis content arranged as follows. In chapter 2, Conditionally Specified (CS) models will be introduced with an extensive literature survey. Some important results in the literature of CS models will be also discussed with examples. Further, important concepts like compatibility, uniqueness of a CS distribution are also discussed in chapter 2. Step by step derivation of the Normal-binary logistic joint model will be presented in chapter 3. Data generation using the new joint model is also briefly discussed in the third chapter. In chapter 4, we present estimation procedures that can be used to find estimates in our newly derived joint model. Application of maximum likelihood estimation and composite likelihood estimation will be discussed in this section with numerical comparisons among the two methods.

Chapter 5 contains a data analysis that involves stock price data example (Example 3) in chapter 1.

Chapter 2: Conditionally Specified Models

A bivariate density is easy to understand/visualize in-terms of its conditional densities. Therefore, one can earn better insight into the form of conditional distributions of experimental variables rather than the joint distribution. For instance, let us cite a classical example by Arnold et al. (2001). Suppose we want to visualize the distribution of height given the weight in some populations. The distribution will be unimodal with the mode varying monotonically with weight. Similarly, the distribution of weight given the height in some populations. The distribution will be unimodal with the mode varying monotonically with height. Yet, it is difficult to visualize features of the appropriate joint distribution of weight and height. One can blindly assume a distribution for this bivariate joint distribution. According to Hauser (1993), Galton's assertion that a unimodal bivariate distribution is one of the picks for the joint distribution of height and weight that turned out to be appropriate. That been said, there's no guarantee that blindly choosing a distribution for joint distribution will be appropriate all the time. Looking at the conditionals instead might be much reasonable in that case. Joint distributions that can be specified using conditional distributions to specified parametric families are called conditionally specified joint models.

In the next section, we present the literature review on the general concept of conditionally specified models in the exponential family of distributions with some useful theorems. Further, we address the issues with these models such as compatibility and uniqueness.

2.1 Conditionally Specified Models: Literature Review

A bivariate density is understandable in terms of its conditional densities Arnold et al. (2001). For instance, instead of providing a model for (X, Y), one can propose families of conditional distributions of X given values of y, and of Y given values x of X. Considerable research has been done on conditionally specified models.

Besag (1974), applied the concept of the conditional specification to spatial stochastic interactions. He examined the spatial interaction of random variables for a finite system by the use of conditional probability models and concluded that the conditional probability approach to the specification and analysis of spatial interaction is more useful than the joint probability approach. Castillo and Galambos (1989), identified the class of all analytic bivariate densities $f(x, y)$ defined on $\Re^2$ for which all conditional models are normally distributed. According to Castillo and Galambos, this class includes the classical bivariate normal density and interesting distributions with non-normal marginals and non-linear regression functions. Based on Castillo and Galambos's work, Arnold (1987), derived the class of all bivariate densities on $\mathbb{R}^{+2}$ for which the conditionals are Pareto (α) densities. Further, Arnold

and Strauss (1991a), studied bivariate distributions with conditionals in exponential families. In their paper, they developed a theorem that can be applied to obtain joint densities using conditional distributions in any specified exponential families. We will discuss this theorem in future sections. Moreover, Arnold and Strauss proposed pseudo-likelihood method as an alternative to the maximum likelihood method to estimate parameters. Arnold et al. (2001), showed that certain conditionally specified densities can provide convenient flexible conjugate prior families in certain multi-parameter Bayesian settings. Further, Arnold et al. (2001) introduced multivariate extensions for conditionally specified concepts. In recent work, Kuo and Wang (2018) and Kuo and Wang (2019), discussed how Gibbs sampling can be used to generate data from conditionally specified models. Moreover, they address the importance of compatibility of conditionals when using gibbs sampling and proved a necessary and sufficient condition for gibbs sampling to simulate the stationary joint probability density from conditionally specified models.

The compatibility of the conditional distributions and the Uniqueness of the joint distribution is also widely discussed in the literature. Arnold et al. (1989), Arnold and Gokhale (1994, 1998), Arnold et al. (2001, 2004, 1989) and Chen (2010) have done some significant work in this area and proposed several approaches to handle the compatibility issue. In more recent papers, Ghosh and Nadarajah (2016, 2017), studied the problem of determining whether a given set of constraints involving marginal and conditional probabilities and expectations of functions are compatible or most nearly compatible when both conditionals are discrete.

2.2 Conditionally Specified Models in Exponential Families (CEF)

As mentioned in the previous section, Arnold and Strauss (1991b), studied bivariate distributions with conditionals in exponential families by extending the work done by Castillo and Galambos (1989) and Arnold (1987). In this section, we use Arnold and Strauss (1991a) and Arnold et al. (2001) as our major references. The section divides into two parts; the first part briefly explains the exponential family of distributions with some classic examples and the second part is dedicated to introducing Arnold et.al. theorem which we use throughout this dissertation.

2.2.1 Exponential Family

The Exponential family is a widely used family of distributions on finite dimensional Euclidean spaces parametrized by a finite dimensional parameter vector Bickel and Doksum (2006) which are practically convenient. The exponential family includes most of the standard discrete and continuous distributions such as normal, poisson, binomial, gamma, multivariate normal and so on. This family is given a special interest in the field of statistics due to its' special properties such as algebraic convenience, special structure and sufficiency properties. Therefore, in next section we will briefly introduce the canonical form of exponential family of distributions and its' variation with some examples. For further information and clarifications reader can use the following classic explanations such as Barndorff-Nielsen (1978), Brown (1986), Casella and Lehmann (2006) and Bickel and Doksum (2006).

Definition 2.1. *Let $X = (X_1, X_2, ..., X_d)$ be a d- dimensional random vector with a distribution $P_\theta, \theta \in \Theta \subseteq \Re$.*

1. *Suppose $X_1, X_2, ..., X_d$ are jointly continuous. The family of distributions $\{P_\theta, \theta \in \Theta\}$ is said to belong to the one parameter Exponential family if the density of $X = (X_1, X_2, ..., X_d)$ may be represented in the form*

$$f(x \mid \theta) = e^{\eta(\theta)T(x) - \psi(\theta)} h(x)$$

for some real valued functions $T(x), \psi(\theta)$ and $h(x) \geq 0$

2. *Suppose $X_1, X_2, ..., X_d$ are jointly discrete. Then $\{P_\theta, \theta \in \Theta\}$ is said to belong to the one parameter Exponential family if the joint pmf of $p(x \mid \theta) = P_\theta(X_1 = x_1, X_2 = x_2, ..., X_d = x_d)$ may be written in the form*

$$p(x/\theta) = e^{\eta(\theta)T(x) - \psi(\theta)} h(x)$$

for some real valued functions $T(x), \psi(\theta)$ and $h(x) \geq 0$

Note that the functions η, T and h are not unique. For example, in the product ηT, we can multiply T by some constant c and divide η by it. Similarly, we can play with constants in the function h. Further, $T(x)$ is called the natural sufficient statistic for the family $\{P_\theta\}$. Apart from the canonical representation, there are other alternative representations which is equivalent to canonical form. We are presenting the alternative representation used in Arnold and Strauss (1991a),

Definition 2.2. *(Exponential Family) : An l_1-parameter family of densities $\{f_1(x;\underline{\theta}) : \underline{\theta} \in \Theta\}$, with respect to dominating measure μ_1 (frequently, Lebesgue measure or counting measure) on D_1 (some subset of Euclidean space of finite dimension), of the form*

$$f_1(x;\underline{\theta}) = r_1(x)\beta_1(\underline{\theta}) \exp\left\{ \sum_{i=1}^{l_1} \theta_i q_{1i}(x) \right\} \tag{2.1}$$

Similarly, we can define,

$$f_2(y;\underline{\tau}) = r_2(y)\beta_2(\underline{\tau}) \exp\left\{ \sum_{j=1}^{l_2} \tau_j q_{2j}(y) \right\} \tag{2.2}$$

Here Θ and T are the natural parameter spaces and $q_{1i}(x)$'s and $q_{2j}(y)$'s are sufficient statistics that assumed to be linearly independent. Further, $\psi(\theta) = -\log \beta_1(\theta)$ and $\psi(\tau) = -\log \beta_2(\tau)$.

Example 2.1. *(Normal Distribution)*

Normal distribution belongs to two parameter exponential family. Therefore, $l_1 = 2$.

$$f(y;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y-\mu)^2}{2\sigma^2} \right\} \text{ , where } \mu \in \mathrm{R} \text{ and } \sigma \in \mathrm{R}^+$$

Rewrite the distribution as exponential form using definition 2.2,

$$f(y;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{-\mu^2}{2\sigma^2} \right\} \exp\left\{ \begin{pmatrix} \frac{\mu}{\sigma^2} & -\frac{1}{2\sigma^2} \end{pmatrix} \begin{pmatrix} y \\ y^2 \end{pmatrix} \right\}$$

where, $\theta_i = \begin{pmatrix} \frac{\mu}{\sigma^2} & -\frac{1}{2\sigma^2} \end{pmatrix}$, $q_{1i}(y) = \begin{pmatrix} y \\ y^2 \end{pmatrix}$, $r_1(y) = 1$ *and* $\beta_1(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-\mu^2}{2\sigma^2}\right)$

Example 2.2. *(Bernoulli Distribution)*

Bernoulli distrbution belongs to one parameter exponential family ($l_1 = 1$). Proba-

bility mass function of Bernoulli distribution is

$$P(R = r) = \pi^r (1 - \pi)^{1-r}, \qquad r = 0, 1$$

Rewrite the distribution as exponential form using definition 2.2,

$$\begin{aligned} P(R = r) &= (1 - \pi) \exp\left\{ \log\left(\frac{\pi}{1 - \pi}\right) r \right\} \\ &= (1 - \pi) \exp\left\{ \theta_1 q_{11}(r) \right\} \end{aligned}$$

Where, $\theta_1 = \log\left(\frac{\pi}{1-\pi}\right)$, $q_{11}(r) = r$, $r_1(r) = 1$ *and* $\beta_1(\theta) = 1$.

Our goal is to identify the class of bivariate densities $f(x, y)$ w.r.t. $\mu_1 \times \mu_2$ on $D_1 \times D_2$ for which conditional densities $f(x/y)$ and $f(y/x)$ are well defined and belongs to the family of exponential for some $\underline{\theta}$ which may depend on y and $\underline{\tau}$ which may depend on x respectively. In next section, we present the theorem which describe the general class of bivariate distributions where their conditionals belong to exponential family.

2.2.2 Arnold and Strauss Theorem

Following is the theorem by Arnold and Strauss (1991a).

Theorem 2.1. *Let* $f(x, y)$ *be a bivariate density whose conditional densities satisfy*

$$f(x \mid y) = f_1(x; \underline{\theta}(y))$$

$$f(y \mid x) = f_2(y; \underline{\tau}(x))$$

for some function $\underline{\theta}(y)$ *and* $\underline{\tau}(x)$ *where* f_1 *and* f_2 *are defined in 2.1 and 2.2. It follows that* $f(x,y)$ *is of the form*

$$f(x,y) = r_1(x)r_2(y)\exp\left\{\underline{q}^{(1)}(x)'M\underline{q}^{(2)}(y)\right\} \tag{2.3}$$

where

$$\underline{q}^{(1)}(x) = (q_{10}(x), q_{11}(x), q_{12}(x), ..., q_{1l_1}(x)),$$

$$\underline{q}^{(2)}(y) = (q_{20}(y), q_{21}(y), q_{22}(y), ..., q_{2l_2}(y))$$

where $q_{10}(x) = q_{20}(y) = 1$, $q_{1i}(x), q_{2j}(x)$; $i = 1, .., l_1$ *and* $j = 1, ..., l_2$ *are sufficient statistics of* $f(x \mid y)$ *and* $f(y \mid x)$ *respectively and* M *is a matrix of constants parameters of appropriate dimensions (i.e.,* $(l_1+1)\times(l_2+1)$*) subject to the requirement that*

$$\int_{D_1}\int_{D_2} f(x,y)d\mu_1(x)d\mu_2(y) = 1$$

The matrix M *is given as,*

$$M = \begin{pmatrix} m_{00} & m_{01} & ... & m_{0l_2} \\ m_{10} & & & \\ ... & & \underline{M} & \\ m_{l_10} & & & \end{pmatrix}$$

Note that when $\underline{M} \equiv 0$*, the two conditionals are independent. That is* $f_{X,Y}(x,y) = f_X(x)f_Y(y)$*.* m_{00} *represent the normalizing constant, which can be derived using the fact that the joint distribution is a valid distribution so integrates to 1.*

Proof: Consider a joint density with conditionals in the given exponential families. Denote the marginal distributions by $g(x), x \in S(X) = \{x : r_1(x) > 0\}$ and $h(y), y \in S(Y) = \{y : r_2(y) > 0\}$ respectively.

The joint can be written as the product of marginal and a conditional in two ways.

$$f(y|x)g(x) = f(x|y)h(y)$$

$$r_2(y)\beta_2(\underline{\tau}(x)) \exp\left\{\sum_{j=1}^{l_2} \tau_j q_{2j}(y)\right\} g(x) = r_1(x)\beta_1(\underline{\theta}(y)) \exp\left\{\sum_{i=1}^{l_1} \theta_i q_{1i}(x)\right\} h(y) \tag{2.4}$$

Define,

$$\tau_0(x) = log[g(x)\beta_2(\tau(x))/r_1(x)]$$

$$\theta_0(y) = log[h(y)\beta_1(\theta(y))/r_2(y)]$$

So equation 2.4 can be written as,

$$\exp(\tau_0) \exp\left\{\sum_{j=1}^{l_2} \tau_j q_{2j}(y)\right\} = \exp(\theta_0) \exp\left\{\sum_{i=1}^{l_1} \theta_i q_{1i}(x)\right\}$$

and

$$\exp\left\{\sum_{j=0}^{l_2} \tau_j q_{2j}(y)\right\} = \exp\left\{\sum_{i=0}^{l_1} \theta_i q_{1i}(x)\right\}$$

This follows that,

$$\exp\left\{\sum_{j=0}^{l_2} \tau_j q_{2j}(y)\right\} = \exp\left\{\sum_{i=0}^{l_1} \theta_i q_{1i}(x)\right\} = \underline{q}^{(1)'}(x) M \underline{q}^{(2)}(y)$$

$$r_1(x)r_2(y) \exp\left\{\sum_{j=0}^{l_2} \tau_j q_{2j}(y)\right\} = r_1(x)r_2(y) \exp\left\{\sum_{i=0}^{l_1} \theta_i q_{1i}(x)\right\}$$

So we can obtain equation (2.3). □

The normalizing constant most of time needs to be evaluated numerically. Because of this awkward normalizing constant, the true likelihood becomes much more complicated and often times intractable. As a result, standard estimation methods such maximum likelihood techniques are difficult to implement. However, there are other approaches have been suggested which we will discussed in chapter 4. However, explicit knowledge about the normalizing constant is not required when generating data from a joint density. Two simple example were given below to understant the application of theorem 2.1

Example 2.3. *(Normal Conditionals) : Here we are dealing with two parameter* $(l_1 = l_2 = 2)$ *exponential family. Assume unknown mean and variance. And we have* $r_1(t) = r_2(t) = 1$. *Then we have,*

$$\underline{q}^{(1)}(t) = \underline{q}^{(2)}(t) = \begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix}$$

From Theorem 2.1, We obtain,

$$f(x,y) = \exp\left\{\underline{q}^{(1)}(t)' M \underline{q}^{(2)}(t)\right\} = \exp\left\{\begin{pmatrix} 1 & x & x^2 \end{pmatrix} M \begin{pmatrix} 1 \\ y \\ y^2 \end{pmatrix}\right\}$$

M *is a* 3×3 *matrix,*

$$M = \begin{pmatrix} m_{00} & m_{01} & m_{02} \\ m_{10} & m_{11} & m_{12} \\ m_{20} & m_{21} & m_{22} \end{pmatrix}$$

The choice $m_{22} = m_{12} = m_{21} = 0$ *gives the classic bivariate normal provide that,*

$$m_{20} < 0, m_{02} < 0, m_{11}^2 < 4m_{02}m_{20}$$

correlations of both signs are possible. And non classical normal conditional models are controlled by the following parametric constraints,

$$m_{22} < 0, 4m_{22}m_{02} > m_{12}^2, 4m_{22}m_{20} > m_{21}^2$$

Example 2.4. *(Poisson - Gamma) : Random variable* (X, Y) *such that,*

$$X|Y = y \sim poisson(y)$$

and assume

$$Y \sim \Gamma(\alpha, \lambda)$$

The resulting distribution X is compound poisson distribution.

$$f(x) = \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)x!}\left(\frac{\lambda}{\lambda+1}\right)^{\alpha}\left(\frac{1}{\lambda+1}\right)^{x} \qquad x = 0, 1, 2, ...$$

which will give us,

$$Y|X = x \sim \Gamma(x+\alpha, \lambda+1)$$

Therefore, joint density belongs to CEF, where l_1*=1 and* l_2*=2. So, M matrix is* 2×3*. Further,* $r_1(x) = \frac{1}{x!}$ *and* $r_2(y) = \frac{1}{y}$*. Joint can be rewritten as,*

$$f(x,y) = \frac{1}{x!y}\exp\left\{\begin{pmatrix}1 & x\end{pmatrix} M \begin{pmatrix}1\\ -y\\ ln(y)\end{pmatrix}\right\} \qquad x = 0, 1, ...; y > 0$$

where,

$$m_{01} > 0, m_{02} > 0, m_{11} \geq 0, m_{12} \geq 0$$

Note that, when $m_{11} = 0$ *and* $m_{12} = 1$ *the joint distribution becomes compound Poisson distribution and* X *and* Y *becomes independent when* $m_{11} = m_{12} = 0$.

2.3 Compatibility

Although using the above theorem we can derive the CS joint model, there is no guarantee that there exists a joint distribution with the given families as its conditionals. That is, when there are candidate families of conditional distributions for X given Y and Y given X, there exist at least one joint distribution for (X, Y) with given families if those candidates families are compatible. There are several ways to check and identify the compatibility of conditionals. First we will consider when X, Y discrete and each have finite set of possible values. All the definitions and theorems are borrowed from Arnold and Gokhale (1994) and Arnold et al. (2001).

2.3.1 Finite discrete case

Let us Consider X and Y to be discrete random variables with possible values $x_1, x_2, ..., x_I$ and $y_1, y_2, ..., y_J$ respectively. Generally, conditional model for the joint distribution of (X, Y) can be associated with two $I \times J$ matrices A and B with either a_{ij} and b_{ij} separately. These a_{ij}'s and b_{ij}'s are assumed to be non negative.

Let

$$a_{ij} = P(X = x_i | Y = y_j) \qquad \forall i, j, \tag{2.5}$$

$$b_{ij} = P(Y = y_j | X = x_i) \qquad \forall i, j, \tag{2.6}$$

$$\sum_{i=1}^{I} a_{ij} = 1 \qquad \forall j, \tag{2.7}$$

and

$$\sum_{j=1}^{J} b_{ij} = 1 \qquad \forall i. \tag{2.8}$$

One important requirement for compatibility is that A and B should have identical incident sets.

Definition 2.3. *(Incident set of a matrix (Arnold et.al. 2001))- Given a matrix A the set $\{(i,j) : a_{ij} > 0\}$ is called the incident set of A and is denoted by N^A.*

According to Arnold and Gokhale (1994), conditional probability matrices A and B are compatible if $N^A = N^B$. If a bivariate random variable (X, Y) exists with conditionals (2.5) and (2.6) then its corresponding marginals are heavily constrained. Suppose the marginals of X and Y are,

$$\tau_i = P(X = x_i) \qquad i = 1, 2, ..., I \tag{2.9}$$

$$\eta_j = P(Y = y_j) \qquad j = 1, 2, ..., J \tag{2.10}$$

Since we can write the joint distribution using its conditionals and marginals in two ways,

$$P(X = x_i, Y = y_j) = P(X = x_i).P(Y = y_j|X = x_i) = P(Y = y_j).P(X = x_i|Y = y_j)$$

Therefore, $\underline{\tau}$ and $\underline{\eta}$ must satisfy,

$$\tau_i b_{ij} = \eta_j a_{ij} \tag{2.11}$$

Theorem 2.2. *(Arnold and Press (1989)) A and B, satisfying (1.3) and (1.4), are compatible iff:*

1. $N^A = N^B = N$ *say; and*

2. $\exists$ *vector* $\underline{u}$ *and* $\underline{v}$ *of appropriate dimensions for which*

$$c_{ij} = a_{ij}/b_{ij} = u_i v_j \qquad \forall (i,j) \in N \tag{2.12}$$

Proof. Suppose first that A and B are compatible, then from (2.11) we have $c_{ij} = \tau_i/\eta_j$ so (8) holds, an appropriate choice for $\underline{\tau}$ is provided by $\tau_i = u_i / \sum_{i=1}^{I} u_i$ and an appropriate choice for $\underline{\eta}$ is provided by $\eta_j = v_{j'}^{-1} / \sum_{j'=1}^{J'} v_{j'}$. A and B have non negative elements.

We will start by introducing the definition for uniform marginal representation followed by a theorem by Arnold and Gokhale (1994),

Definition 2.4. *(Uniform marginal representation of matrix) (Mosteller (1968); Arnold, Arnold et al. (1999))*

Given an $I \times J$ *matrix with non negative elements (with at least one positive element in each row and column), we iteratively normalize rows and columns to have sums* $1/I$ *and* $1/J$*, respectively, until the procedure converges. The limiting matrix is called the uniform marginal representation (UMR) of the original matrix.*

Thus, Arnold and Gokhale (1994) theorem is as follows.

Theorem 2.3. *A and B, satisfying (1.3) and (1.4), are compatible iff:*

1. *A and B are compatible if and only if they have identical uniform marginal representations (UMRs).*

2. *A and B are compatible if and only if all cross product ratios (Arnold et al. (1999), Definitions 2.2 and 2.3) of A are identical to those of B.*

2.3.2 Continuous case

If we relax the condition of (X, Y) being discrete random variables with finite number of possible values, the only change in the theorem 2.3 is notation changes. The only complicated problem is to propose appropriate summable or integrable conditions instead of (2.7) and (2.8).

The joint, marginal and conditional densities defined using usual notation.

$$a(x, y) = f_{X|Y}(x|y), \;\; x \in S(X), \;\; y \in S(Y) \tag{2.13}$$

$$b(x, y) = f_{Y|X}(y|x), \;\; x \in S(X), \;\; y \in S(Y) \tag{2.14}$$

$$N_A = \{(x, y) : a(x, y) > 0\} \tag{2.15}$$

$$N_B = \{(x, y) : b(x, y) > 0\} \tag{2.16}$$

Theorem 2.4. *(Arnold et al. (1999, 2001) A joint density $f(x, y)$, with $a(x, y)$ and $b(x, y)$ as its conditional densities, will exist iff*

1. *N_A=N_B=N;*

2. *$\exists$ functions u and v such that $\forall \; x, y, \in N$;*

$$a(x, y)/b(x, y) = u(x)v(y) \tag{2.17}$$

where,

$$\int_{S(X)} u(x) d\mu_1(x) < \infty.$$

Proof. In order for $a(x, y)$ and $b(x, y)$ to be compatible, suitable marginal densities $f(x)$ and $g(y)$ must exist. Clearly 2.17 must hold with $f(x) \propto u(x)$ and $g(y) \propto$

$1/v(y)$. The condition $\int u(x)d\mu_1(x) < \infty$ is equivalent to $\int [1/v(y)]d\mu_2(y) < \infty$ and only one needs to be check in practice. These integrability conditions reflect the fact that the marginal densities must be integrable and indeed must integrate to 1. We will now look into some examples of compatibility.

Example 2.5. *(Example for continuous case:) Consider the following candidate family of conditional densities with respective to Lebesgue measure.*

$$f_{X|Y}(x|y) = a(x,y) = (y+2)e^{-(y+2)x}I(x>0)$$

$$f_{X|Y}(x|y) = a(x,y) = (x+3)e^{-(x+3)y}I(y>0)$$

We can observe that $S(X) = S(Y) = (0,\infty)$. *By using the theorem,*

$$\frac{a(x,y)}{b(x,y)} = \frac{(y+2)e^{-(y+2)x}}{(x+3)e^{-(x+3)y}} = \left(\frac{e^{-2x}}{x+3}\right)\left(\frac{y+2}{e^{-3y}}\right)$$

and (1.13) holds with $u(x) = \frac{e^{-2x}}{(x+3)}$ *and* $v(y) = (y+2)e^{3y}$

Notice that

$$\int_0^\infty u(x)dx < \infty$$

And thus confirmed the compatibility of two models.

By assuming (X,Y) *random vector is absolutely continuous with respect to some product measure* $\mu_1 \times \mu_2$ *on* $S(X) \times S(Y)$ *where* $S(X)$ *and* $S(Y)$ *are support of* X *and* Y *which can be finite, countable , or uncountable. This can allow one variable to be discrete and the other continuous.*

Example 2.6. *(Example for continuous-discrete case): Consider the following candidate family of conditional densities.*

$$f_{Y|X}(Y/X = x) = a(x,y) = \Gamma(x,\lambda)$$

$$f_{X|Y}(X/Y = y) = b(x, y) = poisson(y)$$

Observe that S(X)= $[0, \infty)$ (discrete) and S(Y) = $(0, \infty)$. So from the theorem (1.4),

$$\frac{a(x, y)}{b(x, y)} = \frac{\lambda^x(\Gamma(x))^{-1}y^{x-1}e^{-\lambda y}}{y^x e^{-y}(x!)^{-1}} = \lambda^x x \times \frac{e^{-(\lambda-1)y}}{y} = u(x)v(y)$$

Where,

$$\sum_{x=0}^{\infty} \lambda^x x = \frac{\lambda}{(1-\lambda)^2} < \infty \ \lambda = 2, ..,$$

Thus confirmed compatibility.

Example 2.7. *(Logistic Regression) : Suppose X takes values in the set $\{x_1, x_2, ..., x_k\}$.*

and suppose all $y \in \mathbf{R}$. For each x we have $Y|X = x \sim N(\theta_x, \sigma_x^2)$.

and for each y we have,

$$P(X = x|Y = y) = \frac{\exp[-(a_x + b_x y)]}{\sum_{x=1}^{k} \exp[-(a_x + b_x y)]}.$$

These two conditionals are compatible if for $x = 1, 2, ..., k$,

1. $\sigma_x^2 = \sigma^2$ *and*

2. $b_x = \theta_x/\sigma^2$.

Note that a_x's are unconstrained.As a remark, When $k = 2$, we have binary X taking values $x_1 = 1$ or $x_2 = 2$, say, and hence a binary logistic conditional. That is,

$$a_x = \begin{cases} 0 \text{ for } x = 1 \\ \alpha_0 \text{ for } x = 2 \end{cases}$$

$$b_x = \begin{cases} 0 \text{ for } x = 1 \\ \\ \alpha_1 \text{ for } x = 2 \end{cases}$$

Proof: *Let,*

$$a(x,y) = f(y|X = x) = N(\theta_x, \sigma_x^2)$$

$$b(x,y) = f(x|Y = y) = \frac{\exp[-(a_x + b_x y)]}{\sum_{x=1}^{k} \exp[-(a_x + b_x y)]}$$

By applying Theorem (2.4) we have:

$$\begin{aligned}
\frac{a(x,y)}{b(x,y)} &= \frac{\frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{ - \frac{(y-\theta_x)^2}{2\sigma_x^2} \right\} \sum_{x=1}^{k} \exp\left\{ - (a_x + b_x y) \right\}}{\exp\left\{ - (a_x + b_x y) \right\}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{ (a_x + b_x y) - \frac{(y-\theta_x)^2}{2\sigma_x^2} \right\} \sum_{i=1}^{k} \exp\left\{ - (a_i + b_i y) \right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{ - \frac{1}{2\sigma_x^2} \left[y^2 + \theta_x^2 - 2\theta_x y - 2\sigma_x^2 a_x - 2\sigma_x^2 b_x y \right] \right\} \sum_{i=1}^{k} \exp\{-(a_i + b_i y)\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{ - \frac{1}{2\sigma_x^2} \left[y^2 - 2(\theta_x + \sigma_x^2 b_x) y \right] + a_x - \frac{\theta_x^2}{2\sigma_x^2} \right\} \sum_{i=1}^{k} \exp\{-(a_i + b_i y)\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{ a_x - \frac{\theta_x^2}{2\sigma_x^2} \right\} \exp\left\{ - \frac{1}{2\sigma_x^2} \left[y^2 - 2(\theta_x + \sigma_x^2 b_x) y \right] \right\} \sum_{i=1}^{k} \exp\{-(a_i + b_i y)\}
\end{aligned}$$

If the two conditionals are compatible then the above expression should factored. In order to be factored we need to have $b_x = -\frac{\theta_x}{\sigma_x^2}$ *and* $\sigma_x^2 = \sigma^2$ *where* a_x*'s are unconstrained. Thus we substitute* b_x *and* σ_x^2 *to the expression.*

$$\begin{aligned}
\frac{a(x,y)}{b(x,y)} &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ a_x + \frac{\theta_x b_x}{2} \right\} \exp\left\{ - \frac{y^2}{2\sigma^2} \right\} \sum_{i=1}^{k} \exp\{-(a_i + b_i y)\} \\
&= \underbrace{\exp\left\{ a_x + \frac{\theta_x b_x}{2} \right\}}_{U(x)} \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ - \frac{y^2}{2\sigma^2} \right\} \sum_{i=1}^{k} \exp\{-(a_i + b_i y)\}}_{V(y)} \\
&= U(x)V(y)
\end{aligned}$$

Where, $\int_{y \in R} V(y)dy < \infty$ *and* $\sum_{x=1}^{k} U(x) < \infty$

Therefore, the two distributions are compatible provided that $b_x = -\frac{\theta_x}{\sigma_x^2}$ *and* $\sigma_x^2 = \sigma^2$ *where* a_x*'s are unconstrained.*

2.4 Uniqueness

After addressing the issue of existence of the CS model so that compatibility of the conditional distributions is assured, question of uniqueness of the CS model also should be addressed. Uniqueness in simple words is that there maybe many different joint distributions consistent with given specification of conditionals. In particular, whether or not the given solution form yield the same conditionals.

Amemiya (1981), Gourieroux et al. (1979) are two of the early references which discussed the uniqueness issue. Arnold et al. (1989) presented the methods to check for the uniqueness for finite discrete case, countable discrete case and absolutely continuous case. Moreover, they extended their work to higher dimensions. Arnold et al. (1999) and Arnold et al. (2001) summarized the existing theoretical methods available to check uniqueness. Kuo and Wang (2017) and Kuo and Wang (2019), presented new computational methods to check the uniqueness of a CS joint model. In this section, we will present some of the major theorems related to the uniqueness issue.

According to Arnold et al. (1999), the necessary and sufficient conditions for uniqueness can be viewed as a Markov chain problem. Suppose (X, Y) is absolutely

continuous with respect to $\mu_1 \times \mu_2$ with support $S(X)$ and $S(Y)$. If $a(x, y)$ and $b(x, y)$ defined as (2.13) and (2.14) are compatible; and $\tau(x)$ is the marginal density of X. Consider the stochastic kernel ba, following the notation by Arnold et al. (1999).

$$ba(x/z) = \int_{S(Y)} a(x, y)b(z, y)d\mu_2(y) \tag{2.18}$$

Note that τ is a stationary distribution of a Markov chain with state space $S(X)$ and transition kernel ba. The distribution $\tau(x)$ is unique if and only if the chain is indecomposible. Reader can refer to Arnold et al. (2001). Given below is a classic example of determining uniqueness from Arnold et al. (2001) paper.

Example 2.8. *Suppose $a(x, y)$ and $b(x, y)$ compatible but with a nonunique compatible density. We will define the sets*

$$A_1 = \{(x, y) : -1 < x < 0, -1 < y < 0\}$$

and

$$A_2 = \{(x, y) : 0 < x < 1, 0 < y < 1\}$$

and set

$$a(x, y) = b(x, y) = I((x, y) \in A_1 \cup A_2) \tag{2.19}$$

Joint density can be obtained in the form of

$$f(x, y) = \frac{\lambda}{2} I((x, y) \in A_1) + \frac{(1 - \lambda)}{2} I((x, y) \in A_2)$$

The joint is compatible with (2.19) when $\lambda \in (0, 1)$. Arnold et al. (2001) states that it is fairly easy to verify that the Markov chain with transition kernel ba defined using

(2.19) is decomposable. The state space $S(X) = [-1, 0] \cup [0, 1]$ *is a disjoint union of two closed subsets of states, namely* $[-1, 0]$ *and* $[0, 1]$*. Thus, non-uniqueness of the above joint density has been proven.*

Further, Arnold et al. (2001) states that the simplest sufficient condition for indecomposibility of the Markov chain with kernel ba is a "positivity" condtion. The assumption that $N_a = N_b = S(X) \times S(Y)$ is sufficient because the kernel ba in (2.18) will be positive for every x and every z.

Chapter 3: Logistic and Bivariate Conditionals

In this section, we present the derivation of a conditionally specified model starting with the Logistic distribution and the Bivariate Normal distribution as conditionals. We consider the stock price data as the motivating example. We propose a logistic regression model for the conditional distribution of the binary valued analyst recommendation given the stock prices from the first and last day of the trading week and the conditional distribution of the stock prices given the analyst recommendation as a Bivariate Normal. We first set up the problem in the notations of Theorem 2.1 and then obtain the form of the joint. It turns out that the normalizing constant can be obtained in a closed form using some results on multivariate normal integrals. We discuss some properties of the resulting joint distribution. We have created a shiny (by Chang et al. (2021)) application which can be used to explore the structure of the joint distribution for various parameter values.

3.1 Setting up the problem

Suppose we have the distribution of observed beginning and end price vector (2×1) of a single trading week (say y) given the analyst recommendation (say r),

$f_{Y|R}(y \mid R = r)$ and the distribution of analyst recommendation given observed data vector of a week, $f_{R|Y}(r|Y = y) \sim Ber[\pi(y, \alpha)]$, where, $\pi(y, \alpha)$ is a function of y parameterized by α. We assume a logistic link:

$$\text{logit}[\pi(y, \alpha)] = log\left(\frac{\pi(y, \alpha)}{1 - \pi(y, \alpha)}\right) = \alpha_0 + \alpha_1 y_1 + \alpha_2 y_2 \tag{3.1}$$

Further we assume that, R is a binary variable and Y is a continuous variable such that $f_{Y|R}(y|R = r) \sim N_2(\mu^{(r)}, \Sigma^{(r)})$, where,

$$\mu^{(r)} = \begin{pmatrix} \mu_1^{(r)} \\ \mu_2^{(r)} \end{pmatrix} \text{ and } \Sigma^{(r)} = \begin{pmatrix} \sigma_1^{(r)2} & \rho^{(r)}\sigma_1^{(r)}\sigma_2^{(r)} \\ \rho^{(r)}\sigma_1^{(r)}\sigma_2^{(r)} & \sigma_2^{(r)2} \end{pmatrix}, \text{ for } r = 0, 1.$$

Under this model, the conditional distributions of the stock price given its buy/sell status is assumed to be normal, with different set of parameter values depending upon the classification. As we shall see in the next section, we will need to assume that the variance-covariance matrices of the two conditionals must be the same (that is, $\Sigma^{(0)} = \Sigma^{(1)}$) in order to ensure compatibility.

3.2 Compatibility of the conditionally specified model

Confirming the existence of the joint model given the two conditionals is essential before the model can be fit to real data. In other words, the two conditionals should be compatible. That is, in this problem, we will start by checking whether the functional forms $f_{R|Y}(r \mid Y = y)$ and $f_{Y|R}(y \mid R = r)$ satisfy the conditions of Theorem 2.4 of Arnold et al. (1989). Following Arnold et al. (1989) notations, let

$$a(r, y) = (2\pi)^{-1}|\Sigma^{(r)}|^{-1/2} exp[-\frac{1}{2}(y - \mu^{(r)})^T \Sigma^{(r)-1}(y - \mu^{(r)})]$$

and

$$b(r, y) = \pi(y)^r [1 - \pi(y)]^{(1-r)}$$

where the logistic link function is

$$\pi(y) = \frac{\exp(\alpha_0 + \alpha_1 y_1 + \alpha_2 y_2)}{1 + \exp(\alpha_0 + \alpha_1 y_1 + \alpha_2 y_2)}.$$

That is,

$$b(1, y) = \frac{\exp(\alpha_0 + \alpha_1 y_1 + \alpha_2 y_2)}{1 + \exp(\alpha_0 + \alpha_1 y_1 + \alpha_2 y_2)}$$

and

$$\begin{aligned} b(0, y) &= 1 - \pi(y) \\ &= 1 - \frac{\exp(\alpha_0 + \alpha_1 y_1 + \alpha_2 y_2)}{1 + \exp(\alpha_0 + \alpha_1 y_1 + \alpha_2 y_2)} \\ &\frac{1}{1 + \exp(\alpha_0 + \alpha_1 y_1 + \alpha_2 y_2)}. \end{aligned}$$

Thus, if we let

$$a_r = \begin{cases} 0 \text{ for } r = 0 \\ \\ \alpha_0 \text{ for } r = 1 \end{cases}$$

$$b_r = \begin{cases} 0 \text{ for } r = 0 \\ \\ \alpha_1 \text{ for } r = 1 \end{cases}$$

and

$$c_r = \begin{cases} 0 \text{ for } r = 0 \\ \\ \alpha_2 \text{ for } r = 1 \end{cases}$$

we can write

$$\pi(y) = \frac{\exp(a_r + b_r y_1 + c_r y_2)}{\sum_{i=0}^{1} \exp(a_i + b_i y_1 + c_i y_2)},$$

and

$$\frac{a(r,y)}{b(r,y)} = \frac{(2\pi)^{-1}|\Sigma^{(r)}|^{-1/2}\exp\Big\{ -\frac{1}{2}(y-\mu^{(r)})^T\Sigma^{(r)-1}(y-\mu^{(r)})\Big\}}{\exp(a_r+b_ry_1+c_ry_2)\Big\{\sum_{i=0}^{1}\exp(a_i+b_iy_1+c_iy_2)\Big\}^{-1}}$$

$$= \frac{(2\pi)^{-1}|\Sigma^{(r)}|^{-1/2}\exp\Big\{ -\frac{1}{2}(y-\mu^{(r)})^T\Sigma^{(r)-1}(y-\mu^{(r)})\Big\}\Big\{\sum_{i=0}^{1}\exp(a_i+b_iy_1+c_iy_2)\Big\}}{\exp(a_r+b_ry_1+c_ry_2)}$$

$$= \frac{(2\pi)^{-1}|\Sigma^{(r)}|^{-1/2}\exp\Big\{ -\frac{1}{2}(y-\mu^{(r)})^T\Sigma^{(r)-1}(y-\mu^{(r)})\Big\}\Big\{\sum_{i=0}^{1}\exp(a_i+(b_i,c_i)y)\Big\}}{\exp\Big\{a_r+(b_r,c_r)y\Big\}}$$

$$= (2\pi)^{-1}|\Sigma^{(r)}|^{-1/2}\exp\Big\{ -a_r - B_r^Ty - \frac{1}{2}(y-\mu^{(r)})^T\Sigma^{(r)-1}(y-\mu^{(r)})\Big\}\sum_{i=0}^{1}\exp(a_i+B_i^Ty)$$

$$= (2\pi)^{-1}|\Sigma^{(r)}|^{-1/2}\exp\Big\{ -a_r - B_r^Ty - \frac{1}{2}y^T\Sigma^{(r)-1}y + 2\mu^{(r)T}\Sigma^{(r)-1}y - \mu^{(r)T}\Sigma^{(r)-1}\mu^{(r)}\Big\}\times \sum_{i=0}^{1}\exp(a_i+B_i^Ty)$$

where we have used B_r to denote the column vector $(b_r, c_r)^T$, expanded the quadratic form in $(\underset{\sim}{y}-\mu^{(r)})$ in terms of y, and separated terms containing only y or only r and those containing both y and r.

Now we note that, if we apply the condition $B_r = 2\Sigma^{(r)-1}\mu^{(r)}$, and that $\Sigma^{(r)} = \Sigma$, we can write

$$\frac{a(r,y)}{b(r,y)} = (2\pi)^{-1}|\Sigma^{(r)}|^{-1/2}\exp\Big\{ -a_r - \frac{1}{2}y^T\Sigma^{(r)-1}y - \mu^{(r)T}\Sigma^{(r)-1}\mu^{(r)}\Big\}\sum_{r=0}^{1}\exp(a_r+B_r^T\underset{\sim}{y})$$

$$= U(r).V(y),$$

where,

$$U(r) = (2\pi)^{-1}|\Sigma^{(r)}|^{-1/2}\exp\Big\{ -a_r - \mu^{(r)T}\Sigma^{(r)-1}\mu^{(r)}\Big\}$$

$$V(y) = \exp\left\{\frac{1}{2} y^T \Sigma^{(r)-1} y\right\} \sum_{r=0}^{1} \exp(a_r + B_r^T \underset{\sim}{y}).$$

This verifies the conditions of theorem (2.4) of Arnold et al. (1989) and hence establishes the compatibility of bivariate Normal and binary logistic distributions as conditionals.

Since $\int_y V(y) < \infty$, compatibility of the given family of conditional densities is assured provided the integrability restriction is also satisfied. That is, the two conditionals are compatible under common variance-covariance matrix Σ and under the condition that $\exp[y^T \Sigma^{-1} y - \alpha_1 y_1 - \alpha_2 y_2]$ integrates to 1.

3.3 Deriving the conditionally specified model

Now that the compatibility condition is satisfied, we can apply Theorem 2.1 to obtain the form of the joint distribution. Note that, in order to apply the theorem 2.1 checking the compatibility beforehand is not necessary. However, checking the compatibility beforehand ensures the existence of the joint distribution obtained from the theorem 2.1. Now, theorem 2.1 was used to obtain the joint distribution of $f(y, r)$. Recall that,

$$\ln(f(y,r)) = (1, r) \begin{pmatrix} m_{00} & m_{01} & m_{02} & m_{03} & m_{04} & m_{05} \\ m_{10} & m_{11} & m_{12} & m_{13} & m_{14} & m_{15} \end{pmatrix} \begin{pmatrix} 1 \\ y_1 \\ y_2 \\ y_1^2 \\ y_1 y_2 \\ y_2^2 \end{pmatrix}.$$

For simplicity we write,

$$f(y,r) = exp(m_{00} + rm_{10}).exp(Q(y,r)); \quad r = 0,1 \text{ and } y \in \mathbf{R},$$

where, $Q(y,r) = (m_{01} + rm_{11})y_1 + (m_{02} + rm_{12})y_2 + (m_{03} + rm_{13})y_1^2 + (m_{04} + rm_{14})y_1y_2 + (m_{05} + rm_{15})y_2^2$.

Obviously, the tiresome part of this derivation is expressing the m_{ij} in terms of α's, μ's and Σ. We will now start by finding the solutions for all the m_{ij} values except for m_{00} which is the normalizing constant. We will present a solution for m_{00} at the end of this section. The usual way to find m values (except m_{00}) is comparing $f(y|R=r)$ and $f(r|Y=y)$ derived from the joint distribution with the original $f(y|R=r)$ and $f(r|Y=y)$. To obtain $f_Y(y)$:

$$\begin{aligned} f_Y(y) &= \sum_{r=0}^{1} f(y,r) \\ &= exp\{m_{00} + m_{01}y_1 + m_{02}y_2 + m_{03}y_1^2 + m_{04}y_1y_2 + m_{05}y_2^2\} \\ &\quad + exp\{m_{00} + m_{10} + (m_{01} + m_{11})y_1 + (m_{02} + m_{12})y_2 + (m_{03} + m_{13})y_1^2 \\ &\quad + (m_{04} + m_{14})y_1y_2 + (m_{05} + m_{15})y_2^2\}; \quad y \in \mathbf{R}. \end{aligned}$$

Let

$$\begin{aligned} A &= exp\{m_{00} + m_{01}y_1 + m_{02}y_2 + m_{03}y_1^2 + m_{04}y_1y_2 + m_{05}y_2^2\} \text{ and} \\ B &= m_{10} + m_{11}y_1 + m_{12}y_2 + m_{13}y_1^2 + m_{14}y_1y_2 + m_{15}y_2^2. \end{aligned} \tag{3.2}$$

So $f(Y=y)$ can be rewritten as,

$$f(Y=y) = A + A.exp(B).$$

In a similar manner, we can rewrite $f(y,r)$ density using A and B as follows,

$$f(Y=y, R=r) = A.exp(r.B).$$

Now the conditional pdf of $f(R \mid Y = y)$ is given by,

$$f(R = r \mid Y = y) = \frac{f(Y = y, R = r)}{f(Y = y)} = \frac{exp(rB)}{1 + exp(B)}.$$

Thus, by substituting equation (3.2) we obtain,

$$f(R = r \mid Y = y) = \frac{exp(r[m_{10} + m_{11}y_1 + m_{12}y_2 + m_{13}y_1^2 + m_{14}y_1y_2 + m_{15}y_2^2])}{1 + exp(m_{10} + m_{11}y_1 + m_{12}y_2 + m_{13}y_1^2 + m_{14}y_1y_2 + m_{15}y_2^2)}.$$

By comparing the true density and the density of the desired logistic regression model, we can obtain the $\pi(y, \alpha)$ as

$$\pi(y, \alpha) = \frac{exp(m_{10} + m_{11}y_1 + m_{12}y_2 + m_{13}y_1^2 + m_{14}y_1y_2 + m_{15}y_2^2)}{1 + exp(m_{10} + m_{11}y_1 + m_{12}y_2 + m_{13}y_1^2 + m_{14}y_1y_2 + m_{15}y_2^2)}. \tag{3.3}$$

The $\pi(y, \alpha)$ in the problem that we are interested, is defined using only linear terms as shown in equation (3.1). Thus, desired $\pi(y, \alpha)$ needs to be obtained by setting the quadratic terms to zero. However, having quadratic terms in the general setting reveals that the derived joint density represents a larger class of joint densities. The particular problem that we are interested in is a special case where the logit link is constructed with a linear function only. Thus, by comparing equation (3.3) and (3.1) we get,

$$m_{10} = \alpha_0 \tag{3.4}$$

$$m_{11} = \alpha_1 \tag{3.5}$$

$$m_{12} = \alpha_2 \tag{3.6}$$

$$m_{13} = m_{14} = m_{15} = 0 \tag{3.7}$$

Similarly, to obtain $f(Y \mid R = r)$, we first derive $f(R = r)$ using $f(Y = y, R = r)$. That is,

$$\begin{aligned} f(R = r) &= \int_{y_2=-\infty}^{\infty} \int_{y_1=-\infty}^{\infty} f(y, r) dy_1 dy_2 \\ &= \int_{y_2=-\infty}^{\infty} \int_{y_1=-\infty}^{\infty} \frac{f(y, r)}{f(y|r)} \times f(y|r) dy_1 dy_2 \\ &= exp(m_{00} + r m_{10})(2\pi)|\Sigma|^{1/2} \int_{y_1, y_2=-\infty}^{\infty} exp\left\{ \frac{1}{2}(y - \mu^{(r)})^T \Sigma^{-1} (y - \mu^{(r)}) + Q(y, r) \right\} f(y|r) d\underline{y}. \end{aligned}$$

In order to obtain a closed form expression for the normalizing constant, we will need to finish the integration. It turns out that an old result involving multivariate normal density function can be used to accomplish this. The complete result is given below for the reader's convenience.

Theorem 3.1. ***(Moment Generating Function (MGF) Theorem):*** *Let* $\underline{X} \sim N_n(\mu, \Sigma)$ *and take* $q_i = 2b_i\underline{x} + \underline{x}^{'} A_i \underline{x}$ *where* b_i *is an* $n \times 1$ *non-random vector and* A_i *is an* $n \times n$ *non-random symmetric matrix (i=1,2,...,k). Take* Γ *to be any* $r \times n$ *matrix such that* $\Sigma = \Gamma^{'}\Gamma$*, where* $r = rank(\Sigma)$*. The joint MGF of* $q_1, ..., q_k$ *is given by,*

$$\begin{aligned} m_{q_1,...,q_k}(t_1, ..., t_k) = |I - 2\sum_{i=1}^{k} t_i \Gamma A_i \Gamma^{'}|^{-1/2} . exp\Bigg\{ 2\left[\sum_{i=1}^{k} t_i (b_i + A_i \mu) \right]^{'} \Gamma^{'} \left[I - 2\sum_{i=1}^{k} t_i \Gamma A_i \Gamma^{'} \right]^{-1} \Gamma \\ \left[\sum_{i=1}^{k} t_i (b_i + A_i \mu) \right] + \mu^{'} \sum_{i=1}^{k} t_i (2b_i + A_i \mu) \Bigg\} \end{aligned}$$

$$\begin{aligned} = |I - 2\sum_{i=1}^{k} t_i \Gamma A_i \Gamma^{'}|^{-1/2} . exp\Bigg\{ 2\left[\sum_{i=1}^{k} t_i b_i \right]^{'} \Sigma \left[I - 2\sum_{i=1}^{k} t_i A_i \Sigma \right]^{-1} \left[\sum_{i=1}^{k} t_i b_i \right] \\ + \mu^{'} \left[I - 2\sum_{i=1}^{k} t_i A_i \Sigma \right]^{-1} \sum_{i=1}^{k} t_i (2b_i + A_i \mu) \Bigg\} \end{aligned}$$

where $(|t_i| < h_i; i = 1, ..., k)$. *for sufficiently small positive constants* $h_1, ..., h_k$

Note that, the above theorem can be used to derive the MGF of linear or quadratic form in a normal random vector $\underline{x}$ or, more generally, of a second degree polynomial in $\underline{x}$. Once the joint distribution becomes available, the marginals of Y and R can be obtained from the joint distribution. Thus, $f(R = r)$ can be written as follows:

$$f(R=r) = exp(m_{00} + rm_{10})(2\pi)|\Sigma|^{1/2} \int_{y_1,y_2=-\infty}^{\infty} exp\Big\{2b\underline{y} + y^T Ay\Big\} f(y|r)d\underline{y}$$
$$= \frac{exp(m_{00} + rm_{10})(2\pi)|\Sigma|^{1/2}}{|I - 2A\Sigma|^{1/2}} exp\Big\{2b^T\Sigma(I - 2A\Sigma)^{-1}b + \mu^{(r)^T}(I - 2A\Sigma)^{-1}(2b + A\mu^{(r)})\Big\}$$

where,

$$b = \left(\frac{2\rho\sigma_1\sigma_2\mu_2^{(r)} - 2\sigma_2^2\mu_1^{(r)}}{4(1-\rho^2)} + \frac{m_{01} + rm_{11}}{2}, \quad \frac{2\rho\sigma_1\sigma_2\mu_1^{(r)} - 2\mu_2^{(r)}\sigma_1^2}{4(1-\rho^2)\sigma_1^2\sigma_2^2} + \frac{m_{02} + rm_{12}}{2}\right)^T$$

and

$$A = \begin{pmatrix} \frac{1}{2(1-\rho^2)\sigma_1^2} + m_{03} + rm_{13} & m_{04} + rm_{14} \\ \frac{-\rho}{(1-\rho^2)\sigma_1\sigma_2} & \frac{1}{2(1-\rho^2)\sigma_2^2} + m_{05} + rm_{15} \end{pmatrix}.$$

Let us now obtain the conditional distribution $f(Y|R = r)$:

$$f(Y|R=r) = \frac{f(Y=y, R=r)}{f(R=r)}$$
$$= \frac{|I - 2A\Sigma_r|^{1/2}}{(2\pi)|\Sigma_r|^{1/2}} exp\Big\{Q(y,r) - 2b^T\Sigma_r(I - 2A\Sigma_r)^{-1}b - \mu_r^T(I - 2A\Sigma_r)^{-1}(2b + A\mu_r)\Big\}$$

We will compare the conditional distribution $f(Y|R = r)$ expressed in terms of the m_{ij} values to the originally specified form of the same distribution $(Y|R = r)$ expressed in terms of the parameters $\underset{\sim}{\mu}^{(r)}$'s and Σ's etc. to obtain the relationships

between the two sets of parameters. These relationships are captured in the following equations:

$$\begin{aligned}
m_{01} + rm_{11} &= \frac{\mu_1^{(r)} - \rho_r(\sigma_{11}/\sigma_{22})\mu_2^{(r)}}{(1-\rho^2)\sigma_{11}^2} \\
m_{02} + rm_{12} &= \frac{\mu_2^{(r)}(\sigma_{11}/\sigma_{22})^2 - \rho(\sigma_{11}/\sigma_{22})\mu_1^{(r)}}{(1-\rho^2)\sigma_{11}^2} \\
m_{03} + rm_{13} &= \frac{-1}{2(1-\rho^2)\sigma_{11}^2} \\
m_{04} + rm_{14} &= \frac{\rho}{(1-\rho^2)\sigma_{11}\sigma_{22}} \\
m_{05} + rm_{15} &= \frac{-1}{2(1-\rho^2)\sigma_{22}^2}
\end{aligned} \tag{3.8}$$

The above equations can be solved for m_{ij} values in terms of the $\mu_1^{(r)}$ and $\mu_2^{(r)}$ values etc. Expressions for $m_{01}, m_{02}, m_{03}, m_{04}$ and m_{05} are as follows:

$$\begin{aligned}
m_{01} &= \frac{\mu_1^{(r)} - \rho(\sigma_{11}/\sigma_{22})\mu_2^{(r)}}{(1-\rho^2)\sigma_{11}^2} - r\alpha_1 \\
m_{02} &= \frac{\mu_2^{(r)}(\sigma_{11}/\sigma_{22})^2 - \rho(\sigma_{11}/\sigma_{22})\mu_1^{(r)}}{(1-\rho^2)\sigma_{11}^2} - r\alpha_2 \\
m_{03} &= \frac{-1}{2(1-\rho^2)\sigma_{11}^2} \\
m_{04} &= \frac{\rho}{(1-\rho^2)\sigma_{11}\sigma_{22}} \\
m_{05} &= \frac{-1}{2(1-\rho^2)\sigma_{22}^2}
\end{aligned} \tag{3.9}$$

From equation (3.9), one can also obtain the original parameters $\mu_1^{(r)}, \mu_2^{(r)}, \rho_r, \sigma_{11}^2$ and σ_{22}^2 in terms of m_{ij}'s. Thus, we have,

$$\rho = \frac{m_{04}}{2\sqrt{m_{03}.m_{05}}} \tag{3.10}$$

$$\sigma_{22}^2 = \frac{2m_{03}}{m_{04}^2 - 4m_{03}.m_{05}} \tag{3.11}$$

$$\sigma_{11}^2 = \frac{2m_{05}}{m_{04}^2 - 4m_{03}.m_{05}} \tag{3.12}$$

According to equations (3.10),(3.11) and (3.12), it is clear that ρ,σ_{11}^2 and σ_{22}^2 do not depend on the value of R. This implies a common variance-covariance matrix for the conditional distributions $f(y|R=0)$ and $f(y|R=1)$. Further,

$$\mu_2^{(r)} = \frac{\rho \times \frac{m_{01}+r\alpha_1}{m_{03}} + \frac{m_{02}+r\alpha_2}{m_{05}}}{\rho^2\sqrt{\frac{m_{05}}{m_{03}}} - \frac{m_{05}}{m_{03}}} \tag{3.13}$$

$$\mu_1^{(r)} = \rho\sqrt{\frac{m_{05}}{m_{03}}} \times \mu_2^{(r)} - \frac{m_{01}+r\alpha_1}{m_{03}} \tag{3.14}$$

The above solutions are verified using simulation studies in chapter 4.

3.4 Deriving normalizing constant (m_{00})

In general, obtaining closed form solution to m_{00} is known to be difficult and sometimes such a closed form may not even exist. For our problem however, we were able to derive a closed form expression for m_{00}. Since m_{00} should be such that $\int_{y=-\infty}^{\infty} f(y,r) = 1$, it follows that

$$m_{00} = -ln\left[\sum_{r=0}^{1} exp(rm_{10}) \int_{y1=-\infty}^{\infty}\int_{y2=-\infty}^{\infty} exp(Q(y,r))dy_2dy_1\right].$$

By using the theorem (3.1), the double integral above can be written as,

$$\int_{y_1=-\infty}^{\infty}\int_{y_2=-\infty}^{\infty} exp(Q(y,r)) = \frac{2\pi\sigma_{11}\sigma_{22}\sqrt{1-\rho^2}}{exp(C(r))}|I-2A\Sigma|^{-1/2}$$
$$exp\left[2b^T\Sigma(I-2A\Sigma)^{-1}b + \mu^{(r)^T}(I-2A\Sigma)^{-1}(2b+A\mu^{(r)})\right],$$

where,

$$C(r) = \frac{-(\sigma_{22}^2\mu_1^{(r)^2} - 2\rho\sigma_{11}\sigma_{22}\mu_1^{(r)}\mu_2^{(r)} + \sigma_{11}^2\mu_2^{(r)^2})}{2(1-\rho^2)\sigma_{11}^2\sigma_{22}^2},$$

$$b = \left(\frac{2\rho\sigma_{11}\sigma_{22}\mu_2^{(r)} - 2\sigma_{22}^2\mu_1^{(r)}}{4(1-\rho^2)} + \frac{m_{01} + rm_{11}}{2}, \quad \frac{2\rho\sigma_{11}\sigma_{22}\mu_1^{(r)} - 2\mu_2^{(r)}\sigma_{11}^2}{4(1-\rho^2)\sigma_{11}^2\sigma_{22}^2} + \frac{m_{02} + rm_{12}}{2} \right)^T,$$

and

$$A = \begin{pmatrix} \frac{1}{2(1-\rho^2)\sigma_{11}^2} + m_{03} + rm_{13} & m_{04} + rm_{14} \\ \frac{-\rho}{(1-\rho^2)\sigma_{11}\sigma_{22}} & \frac{1}{2(1-\rho^2)\sigma_{22}^2} + m_{05} + rm_{15} \end{pmatrix}.$$

Therefore, m_{00} is given by,

$$m_{00} = -ln\Bigg[2\pi\sigma_{11}\sigma_{22}\sqrt{1-\rho^2} \sum_{r=0}^{1} |I - 2A\Sigma|^{-1/2}.exp\Big(rm_{10} - C(r) + 2b^T\Sigma(I-2A\Sigma)^{-1}b + \mu^{(r)T}(I-2A\Sigma)^{-1}(2b + A\mu^{(r)})\Big)\Bigg].$$

3.5 Data Generation

Although the proposed joint model has a closed form expression, it is very complex and has a messy normalizing constant. Therefore, generating data directly from the joint model is immensely difficult and may even not be feasible. However, since the model is conditionally specified we can apply other numerical algorithms such as Gibb's Sampling. This underlines another important advantage of a conditionally specified distribution.

3.5.1 Gibbs Sampling method

Gibbs sampling algorithm, named by Geman and Geman (1984), is a special case of Metropolis-Hastings algorithm. This algorithm can be used to generate data

from multivariate distributions when the univariate conditional densities are fully specified. For more details the reader is referred to Rizzo (2019). For instance, let us assume that we want to generate data from a bivariate density $f_{X,Y}(x, y)$ and the conditional distributions of the model are fully specified. Let $f_{X|Y}(x \mid y)$ and $f_{Y|X}(y \mid x)$ be conditional densities of x given y and y given x respectively. Suppose simulating data from the bivariate density is complicated, Gibbs sampling algorithm is an effective method to generate data from the conditional densities despite having information about the marginal distributions $f_X(x)$ and $f_Y(y)$. Presented below is the Gibbs sampling algorithm.

Algorithm 3.1

1. Let (X_0, Y_0) be initial values.

2. Suppose we already generated $(X_0, Y_0), (X_1, Y_1), .., (X_t, Y_t)$, Then to draw X_{t+1}, Y_{t+1} we follow the below Gibbs cycle,

$$\text{Gibbs Cycle} : \begin{cases} \text{Generate} \quad X_{t+1} \sim f_{X|Y}(x \mid Y_t) \\ \text{Generate} \quad Y_{t+1} \sim f_{Y|X}(y \mid X_{t+1}) \end{cases}$$

The cycle will ultimately generate a Markov chain which will represent the target distribution. The transition kernel given by,

$$K((x_0, y_0), (x_1, y_1)) = f_{X|Y}(x_1 \mid y_0) f_{Y|X}(y_1 \mid x_1)$$

It is recommended that instead of taking complete chain as the random sample, skip several iteration in between or remove the first 1000 iterations to reduce possible auto correlation among the sampled values.

Presented below is the algorithm which can be used to generate data from the proposed density.

Algorithm 3.2

1. Initial value : Set $R^{(0)} = 1$. So, $Y^{(0)} \sim f_1(y)$.

2. Set $\mu^{(0)}$,$\mu^{(1)}$,Σ and the vector α.

3. Suppose we generated $(Y^{(0)}, R^{(0)}), (Y^{(1)}, R^{(1)}), ..., (Y^{(t)}, R^{(t)})$.

   (a) if $R^{(t)} = 0$ Generate $Y^{(t+1)} \sim f(y \mid R = 0)$ and $R^{(t+1)} \sim f(r \mid Y = y^{(t+1)})$

   (b) if $R^{(t)} = 1$ Generate $Y^{(t+1)} \sim f(y \mid R = 1)$ and $R^{(t+1)} \sim f(r \mid Y = y^{(t+1)})$

Contour plots and surface plots for empirical density of the joint distribution (3.1) for $f(y, R = 0)$ and $f(y, R = 1)$ suggest that the joint model is not a unimodal distribution. We have built an interactive tool for exploring the shape of the joint distribution in Shiny. The reader may download this tool from https://github.com/nadeesriw/ShinyAp Further in Figure 3.2 and 3.3 shows how the shape of the density changes. Other parameters are set to $_{(y|R=0)} \sim N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.2 & 0.23 \\ 0.23 & 1.4 \end{pmatrix} \right]$ and $_{(y|R=1)} \sim N_2 \left[\begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 10 & 3.3 \\ 3.3 & 2 \end{pmatrix} \right]$

1. $\rho = \mathbf{0}$

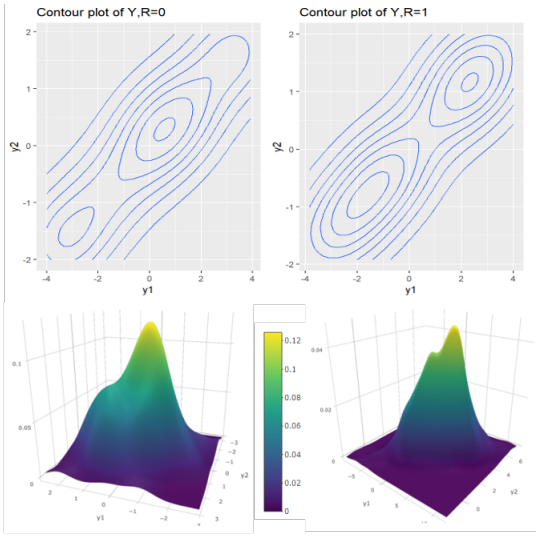


Figure 3.1: Contour plots and Surface plots of the joint distribution. Top and bottom left: $f(Y, R = 0)$ and top and bottom right: $f(Y, R = 1)$.

2. $\underline{\rho = \mathbf{0.5}}$

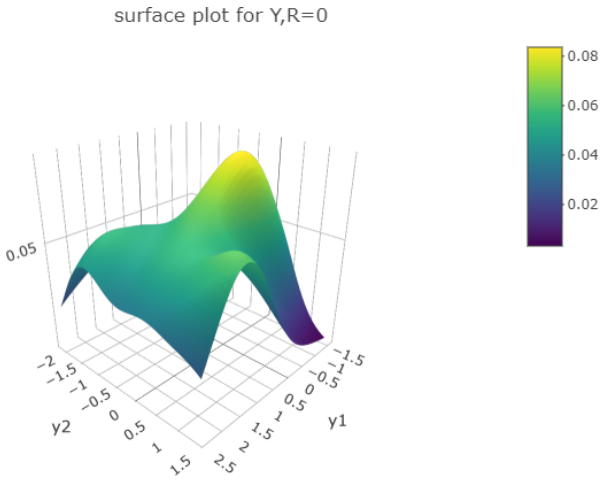


Figure 3.2: Surface plots of the joint distribution $f(Y, R = 0)$

$\underline{\rho = \mathbf{1}}$

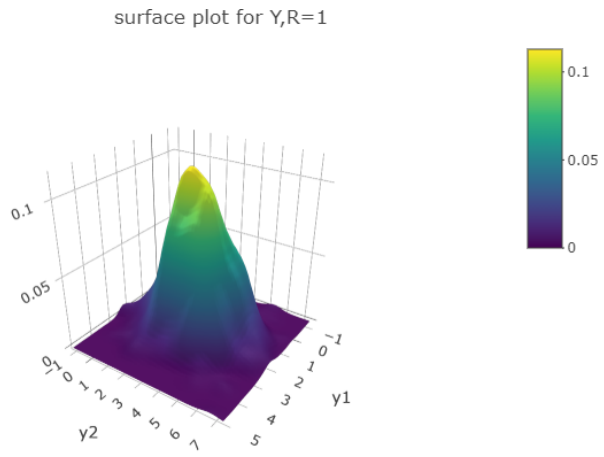


Figure 3.3: Surface plots of the joint distribution $f(Y, R = 1)$

3.6 Bivariate Normal and Multinomial Distribution

Up to this point, we assumed that the discrete part of our data follows a Bernoulli distribution. In practice, we may have more than two values for the discrete part of the data, and maybe assumed to be distributed multinomially. Suppose $s_1, s_2, ..., s_k$ denote the k different states. (In the case of Bernoulli, $K = 2$ and $s_1 = 0, s_2 = 1$) Then the new joint distribution derived using theorem (2.1) based on conditional distributions $f_{Y|R}(y \mid R = s_i) \sim N_2(\mu^{(s_i)}, \Sigma)$ and $f_{R|Y}(r \mid y) \sim \text{Multinomial}(\pi_1(y, \underset{\sim}{\alpha}_1), \pi_2(y, \underset{\sim}{\alpha}_2), ..., \pi_k(y, \underset{\sim}{\alpha}_k))$ where $i = 1, ..., k$; will be different from the joint distribution we derived previously using Bernoulli distribution ($f_{R|Y}(r \mid y) \sim \text{Ber}(\pi(y, \alpha))$). Thus, in this section, we briefly discuss the important steps of deriving the joint distribution $f_{Y,R}(y, r)$. We will start by introducing the conditional distributions.

Suppose $(Y \mid R = s_i)$ follows a bivariate normal distribution with mean vec-

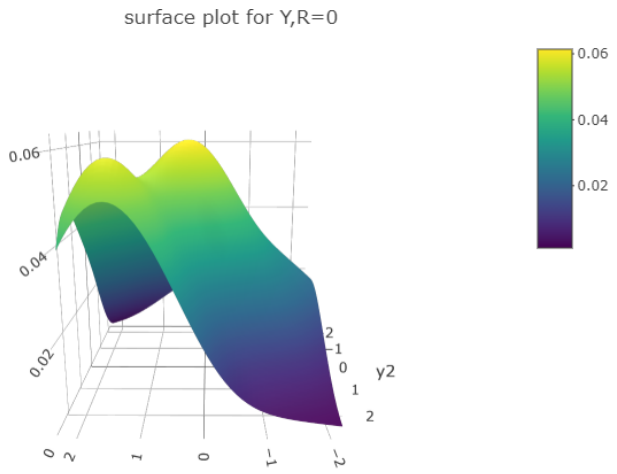


Figure 3.4: Surface plots of the joint distribution $f(Y, R = 0)$

tor $\mu^{(s_i)}$ and common variance covariance matrix Σ. That is, $f_{Y|R}(y \mid R = s_i) \sim N_2(\mu^{(s_i)}, \Sigma)$, $i = 1, ..., k$. Now consider, probabilities $\pi_0(y, \underset{\sim}{\alpha}_0), \pi_1(y, \underset{\sim}{\alpha}_1), ..., \pi_{k-1}(y, \underset{\sim}{\alpha}_{k-1})$, are $\sum_{i=0}^{k-1} \pi_i(y, \underset{\sim}{\alpha}_i) = 1$. where, $\pi_i(y, \underset{\sim}{\alpha}_i)$ is a function of y parameterized by α_i. We assume a logistic link, for probabilities:

$$\text{logit}[\pi_i(y, \underset{\sim}{\alpha}_i)] = log\left(\frac{\pi_i(y, \underset{\sim}{\alpha}_i)}{1 - \pi_i(y, \underset{\sim}{\alpha}_i)}\right) = \alpha_{i0} + \alpha_{i1}y_1 + \alpha_{i2}y_2.$$

If we let s_i denote the result in outcome number i, then

$$f_{R|Y}(R = r \mid Y = y) \sim \prod_{i=0}^{k-1} \pi_i(y, \underset{\sim}{\alpha}_i)^{I_{s_i}(r)} \tag{3.15}$$

where $s_1, s_2, ..., s_k$, denote the different categories, and $I_{s_i}(r)$, $i = 0, .., k-1$ are indicator functions given by:

$$I_{s_i}(r) = \begin{cases} 1 & \text{if } r = s_i \\ 0 & \text{o.w.} \end{cases}$$

Clearly, for $k = 2$, we end up with a Bernoulli distribution with logistic link function. While we denote by $\underset{\sim}{\alpha}$'s, potentially distinct parameters for each category i, it is

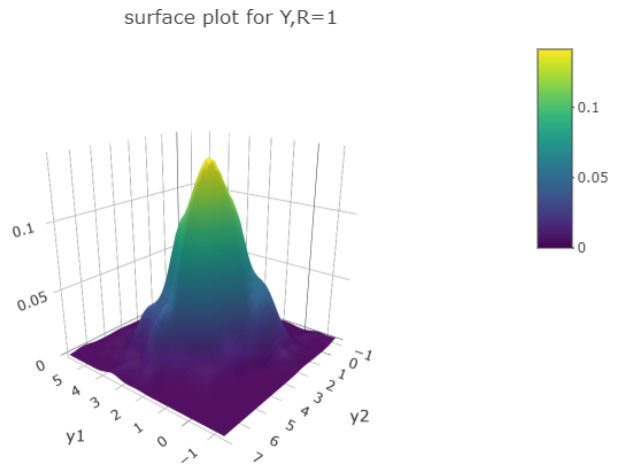


Figure 3.5: Surface plots of the joint distribution $f(Y, R = 1)$

important to note that normally they would be chosen to satisfy one of the standard links such as cumulative logit or generalized logit or proportional odds model. The above model (3.15) can be rewritten as follows:

$$f_{R|Y}(R = r \mid Y = y) \propto \pi_{k-1}(y, \underset{\sim}{\alpha}_{k-1}) \prod_{i=0}^{k-1} \left[\frac{\pi_i(y, \underset{\sim}{\alpha}_i)}{\pi_{k-1}(y, \underset{\sim}{\alpha}_{k-1})} \right]^{I_{s_i}(r)}$$

Since the multinomial distribution belongs to the exponential family of distributions, once you write the canonical form of $f(R = r \mid Y = y)$, you immediately see the following expression. That is,

$$\begin{aligned} f_{R|Y}(R = r \mid Y = y) &= \exp\left[\log(\pi_{k-1}(y, \underset{\sim}{\alpha}_{k-1})) + \sum_{i=0}^{k-1} I_{s_i}(r).\log\left(\frac{\pi_i(y, \underset{\sim}{\alpha}_i)}{\pi_{k-1}(y, \underset{\sim}{\alpha}_{k-1})} \right) \right] \\ &= \exp\left[\left(1, I_{s_1}(r), ..., I_{s_{k-1}}(r) \right) \left(\log\pi_0(y, \underset{\sim}{\alpha}_0), ..., \log\frac{\pi_{k-1}(y, \underset{\sim}{\alpha}_{k-1})}{\pi_k(y, \underset{\sim}{\alpha}_k)} \right)^T \right]. \end{aligned}$$

Now we can easily write down the basic form of the joint distribution $f(y, r)$ using the sufficient statistics and M matrix by applying theorem 2.1. Thus, we can write the joint distribution as,

$$\ln(f(y, r)) = (1, I_{s_1}(r), I_{s_2}(r), ..., I_{s_{k-1}}(r)) M (1, y_1, y_2, y_1^2, y_1 y_2, y_2^2)^T$$

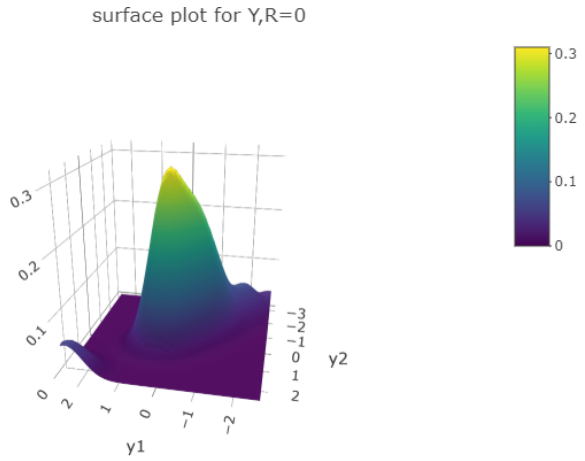


Figure 3.6: Surface plots of the joint distribution $f(Y, R = 0)$

where,

$$M = \begin{pmatrix} m_{00} & m_{01} & m_{02} & m_{03} & m_{04} & m_{05} \\ m_{10} & m_{11} & m_{12} & m_{13} & m_{14} & m_{15} \\ m_{20} & m_{21} & m_{22} & m_{23} & m_{24} & m_{25} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ m_{(k-1)0} & m_{(k-1)1} & m_{(k-1)2} & m_{(k-1)3} & m_{(k-1)4} & m_{(k-1)5} \end{pmatrix}_{k\times 6}.$$

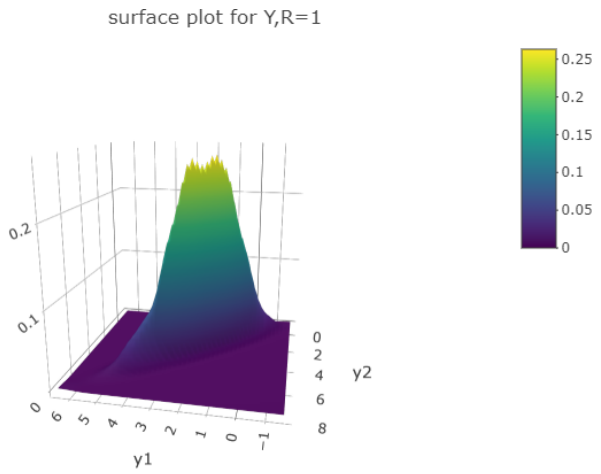


Figure 3.7: Surface plots of the joint distribution $f(Y, R = 1)$

Multiplying out the above expression we get,

$$\ln(f(y,r)) = \left(m_{00} + \sum_{i=1}^{k-1} I_{s_i}(r)m_{i0}, m_{01} + \sum_{i=1}^{k-1} I_{s_i}(r)m_{i1}, ..., m_{05} + \sum_{i=1}^{k-1} I_{s_i}(r)m_{i5}\right) \begin{pmatrix} 1 \\ y_1 \\ y_2 \\ y_1^2 \\ y_1y_2 \\ y_2^2 \end{pmatrix}$$

$$= (m_{00} + \sum_{i=1}^{k-1} I_{s_i}(r)m_{i0}) + (m_{01} + \sum_{i=1}^{k-1} I_{s_i}(r)m_{i1})y_1 + (m_{02} + \sum_{i=1}^{k-1} I_{s_i}(r)m_{i2})y_2 +$$
$$(m_{03} + \sum_{i=1}^{k-1} I_{s_i}(r)m_{i3})y_1^2 + (m_{04} + \sum_{i=1}^{k-1} I_{s_i}(r)m_{i4})y_1y_2 + (m_{05} + \sum_{i=1}^{k-1} I_{s_i}(r)m_{i5})y_2^2.$$

Note that, for each $r = s_i$, $i = 1, 2, ..., k$ we have

$$\log(f(y, s_i)) = (m_{00} + m_{i0}) + (m_{01} + m_{i1})y_1 + ... + (m_{05} + m_{i5})y_2^2$$

$$= (m_{00} + m_{i0}) + Q(y, r)$$

where $Q(y, r) = (m_{01} + m_{i1})y_1 + (m_{02} + m_{i2})y_2 + (m_{03} + m_{i3})y_1^2 + (m_{04} + m_{i4})y_1y_2 +$

$(m_{05} + m_{i5})y_2^2$

To derive elements of M one can follow the same steps we used to derive the Bivariate Normal Bernoulli distribution. We can start by finding the solutions for all the m_{ij} values except for m_{00} which is the normalizing constant. To find m values (except for m_{00}) we have to compare $f(Y|R = r)$ and $f(R|Y = y)$ derived from the joint distribution with the original $f(Y|R = r)$ and $f(R|Y = y)$. To find the normalizing constant m_{00}, we can use the theorem 3.1. However, finding a single expression for m_{00} is not feasible as the the multinomial case has k categories. That is, for each of the $k - 1$ categories $i = 1, \ldots, k$ we have to derive a unique m_{0i}.

Chapter 4: Estimation

In practice, Maximum Likelihood (ML) Estimation is the preferred and the most popular estimation method for parametric models. The goal of ML estimation is to find the values of the model parameters that maximize the likelihood function/log likelihood function over the parameter space. However, finding a closed form analytical solutions cannot be guaranteed all the time. As a remedy, one can use numerical optimization which can most of the time be computationally demanding. These issues rise in conditionally specified models also. As stated in Arnold et al. (2001); Arnold and Strauss (1991a) papers, standard estimation methods are often difficult to implement when dealing with conditionally specified models. This is mainly because of the awkward normalizing constant which is often intractable. And, if an explicit expression is available for the constant, it is usually complicated. Thus, differentiating the likelihood and obtaining estimates by solving the equation is not a viable option. According to Arnold and Strauss (1991a), for exponential family of conditionals ML estimation is a reasonably viable method but it comes with a heavy computational burden. As a remedy, one can use an approach known as Pseudolikelihood (PL) estimation (Besag (1974, 1975)). Pseudolikelihood estimation is perhaps the precedent of composite likelihood (CL) (Lindsay (1988)).

In this chapter, we will present the form of the score functions, fisher information of both ML and PL estimates. We will compare ML and PL estimates in terms of relative efficiency and computational cost.

4.1 Preliminaries

We will start by stating some definitions and theorems which are in the derivations in forthcoming sections. We will first present the definition of ML estimation following the notation of Arnold and Strauss (1991a).

Definition 4.1. *Suppose that we have n observations* $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ *from some bivariate conditionally specified density* $f(x, y; \underset{\sim}{\theta}), \underset{\sim}{\theta} \in \Theta$. *The maximum likelihood estimate of* $\underset{\sim}{\theta}$, *say* $\hat{\underset{\sim}{\theta}}$ *is a value of* $\underset{\sim}{\theta}$ *for which*

$$\prod_{i=1}^{n} f(X_i, Y_i; \hat{\underset{\sim}{\theta}}) = \underset{\underset{\sim}{\theta} \in \Theta}{\operatorname{argmax}} \prod_{i=1}^{n} f(X_i, Y_i; \underset{\sim}{\theta}).$$

In general, several numerical approaches are available for the computation of ML estimates. While a direct search method might be the only way to ensure that the obtained solution in a unique maxima, many other numerical methods are used to obtain a solution which may be a local maxima. One such approach is to solve the log likelihood equation, which are obtained by setting the first derivative of the log likelihood function (also knows as the score function) equal to zero.

Next we state a well known result on the derivation of the score function corresponding to a reparameterization. The result is useful in our case because the joint distribution function corresponding to the given conditionals are expressed simply in terms of a complicated functions of the natural parameters of the conditionals.

This result is commonly available in many basic Mathematical Statistics textbooks and we use the statement given in page 126 of Casella and Lehmann (2006).

Theorem 4.1. *Let X be distributed with density $p_\theta, \theta \in \Omega$, with respect to measure μ where θ is vector-valued, say $\theta = (\theta_1, \theta_2, ..., \theta_s)^T$. Suppose that $\theta_i = h_i(\xi_1, ..., \xi_s)$, $i = 1, 2, ..., s$, is a reparameterization. Let $\mathcal{D}$ be the matrix of derivatives,*

$$\mathcal{D} = \left(\left(\frac{\partial \theta_j}{\partial \xi_i} \right) \right) \qquad i = 1, 2, ..., s \quad \text{and} \quad j =, 1, 2, ..., s.$$

Let the information matrix for $\xi = (\xi_1, \xi_2, ..., \xi_s)^T$ be denoted by $I^*(\xi) = \left(\left(I^*_{ij}(\xi) \right) \right)$, where,

$$I^*_{ij}(\xi) = E\left[\frac{\partial}{\partial \xi_i} \ log \ p_{\theta(\xi)}(X) . \frac{\partial}{\partial \xi_j} \ log \ p_{\theta(\xi)}(X) \right],$$

where $\theta(\xi) = (h_1(\xi_1), ..., h_s(\xi_s))'$.

Proof : It is seen from the chain rule for differentiating a function of several variables that

$$I^*_{ij}(\xi) = \sum_{k=1}^{s} \sum_{l=1}^{s} I_{kl}(\theta) \frac{\partial \theta_k}{\partial \xi_i} \frac{\partial \theta_l}{\partial \xi_j}$$

and hence

$$I^*(\xi) = \mathcal{D} \mathbf{I} \mathcal{D}^T \tag{4.1}$$

where $I \equiv I(\theta)$ is the information matrix for the parameter vector $\underset{\sim}{\theta}$. By further investigating the theorem (4.1), we can write $\mathbf{I}^*(\xi) = -S^*(\xi) S^*(\xi)^T$ and $\mathbf{I}(\theta) = -S(\theta) S(\theta)^T$. Then, by equation (4.1) we have,

$$-S^*(\xi) S^*(\xi)^T = \mathcal{D}[-S(\theta) S(\theta)^T] \mathcal{D}^T$$

$$-S^*(\xi) S^*(\xi)^T = -\mathcal{D} S(\theta) [\mathcal{D} S(\theta)]^T.$$

By comparing we get,

$$S^*(\xi) = \mathcal{D}S(\theta). \tag{4.2}$$

We will be using the theorem 4.1 in section 4.2.

Statistical estimation methods are usually characterized by either minimization or maximization of an objective function, Least Squares (LS) estimation and ML estimation are being two famous examples. These are in contrast with the method of moments (MOM) where estimators are defined as solution to certain equations without an associated objective function to be optimized. In principle, one may consider any reasonable objective function which is a function of the data and the model assumptions to derive an estimator. Of course, the quality of the resulting estimator would depend on how well does the objective function capture the stochastics of the true model underlying the observed data. The ML method is widely used and is reputed to lead to estimators who enjoy many optimality properties. However, in many models, the likelihood function is either difficult or sometimes even impossible to derive. In such cases, one may consider objective functions which are close to the likelihood function and easier to handle computationally. Such functions are called pseudo likelihood (PL) functions or composite likelihood (CL) functions. The idea is for the PL or CL function to capture the key features of the likelihood function while dropping its' computationally complex aspects. There is extensive literature on PL and CL methods (see Lindsay (1988), Varin et al. (2011), Besag (1975), Arnold and Strauss (1991a)). Clearly, using PL and CL will lead to loss of efficiency. Varin et al. (2011) provides a very reader friendly account of the CL

method including a frame work for evaluating the loss of efficiency in using CL (or PL) relative to ML method. Here, we provide a brief review of their work which will be used later to propose a CL method for the conditionally specified models and evaluate its efficiency. We begin with the definition of CL estimator.

Definition 4.2. *Consider an m-dimensional vector random variable Y, with probability density function $f(y;\theta)$ for some unknown p-dimensional parameter vector $\theta \in \Theta$. Denote by $\{A_1, ..., A_K\}$ a set of marginal or conditional events with associated likelihoods $L_k(\theta; y) \propto f(y \in A_k; \theta)$.* ***Composite likelihood*** *can be written as,*

$$L_{CL}(\theta; y) = \prod_{k=1}^{K} L_k(\theta; y)^{\omega_k} \tag{4.3}$$

where, ω_k are non-negative weights to be chosen

Example 1: The product of the marginal likelihood function is simplest example of a composite marginal likelihood:

$$L_{ind}(\theta; y) = \prod_{r=1}^{m} f(y_r, \theta).$$

Because of the underlying independence assumption composite marginal likelihood sometimes referred to as the independence likelihood. This composite likelihood only permits inference on marginal parameters. The reader is referred to Varin et al. (2011), Chandler and Bate (2007), Cox and Reid (2004) and Varin (2008) for more information.

Example 2: In the context of longitudinal studies, Molenberghs and Verbeke (2006) and in the context of bioinformatics, Mardia et al. (2008) construct composite like-

lihoods by pooling pairwise conditional densities

$$L(\theta; y) = \prod_{r=1}^{m} \prod_{s=1}^{m} f(y_r \mid y_s; \theta).$$

The maximum composite likelihood estimator $\hat{\theta}_{CL}$ locates the maximum of the composite likelihood function, or equivalently of the composite log-likelihood function $\ell_{cl}(\theta; y) = \sum_{k=1}^{K} \ell_k(\theta; y)\omega_k$, where $\ell_k(\theta; y) = \log L_k(\theta; y)$. In standard problems $\hat{\theta}_{CL}$ may be found by solving the composite likelihood equations, obtained by setting the composite score function $U_{cl}(\theta; y) = \nabla_\theta \ell_{cl}(\theta; y)$, equal to zero. Note that, $U_{cl}(\theta; y)$ is a linear combination of the scores associated with each log-likelihood term $l_k(\theta; y)$. Further, sensitivity matrix defined as,

$$H(\theta) = E_\theta\{-\nabla_\theta U(\theta; Y)\} = \int \{-\nabla_\theta U(\theta; y)\} f(y; \theta) dy,$$

where $U(\theta; Y) = \nabla_\theta \ell_{cl}(\theta; y)$.

Note that the sensitivity matrix is the expected value of the Hessian of the composite log likelihood equation with respect to the true probability distribution of the data. Similarly, the variability matrix is defined as:

$$J(\theta) = Var\{\nabla_\theta U(\theta; Y)\},$$

where variance is computed with respect to the true probability distribution of the observed data. It is important to note that the if we were to compute the sensitivity matrix and variability matrix with true likelihood as the composite likelihood, they will be equal to each other. And since we no longer are dealing with the full likelihood the Fisher information needs to be substituted by Godambe information matrix

(Godambe (1960)) which is also called the sandwich information matrix:

$$G(\theta) = H(\theta)J(\theta)^{-1}H(\theta). \tag{4.4}$$

Note that if we have the true log likelihood function then $G = H = I$, where $I \equiv I(\theta)$ is the Fisher information matrix. Further, under regularity conditions on the component log-densities we have a central limit theorem for the composite likelihood score statistic, leading to the result that the composite maximum likelihood estimator, θ_{CL} is asymptotically normally distributed.

$$\sqrt{n}(\hat{\theta}_{CL} - \theta) \xrightarrow{d} N_p[0, G^{-1}(\theta)] \tag{4.5}$$

where $N_p(\mu, \Sigma)$ denoted the p-dimensional normal distribution with mean μ and variance covariance matrix Σ.

As noted earlier, composite likelihood functions are constructed in cases where full maximum likelihood functions are not computationally convenient. Although, every effort is to be made to ensure that the composite likelihood capture all features of the full likelihood, one would not expect it to be fully efficient. The asymptotic properties of maximum likelihood estimator is well chronicled. Standard reference include in Rao (1973) and DasGupta (2008). In particular, under certain regularity conditions, the ML estimate $\hat{\theta}_{ML}$ is consistent and asymptotically normal:

$$\sqrt{n}(\hat{\theta}_{ML} - \theta) \to N(0, I^{-1}(\theta)).$$

Comparing the asymptotic distributional results for $\hat{\theta}_{ML}$ and $\hat{\theta}_{CL}$, we note that both are consistent and asymptotically normal with $I^{-1}(\theta)$ and $G^{-1}(\theta)$ as the respective variance covariance matrices. Thus, the efficiency of $\hat{\theta}_{CL}$ with respect to $\hat{\theta}_{ML}$ can

be measured using the concept of Joint Asymptotic relative efficiency (JARE), as defined in Cramer (1999):

$$JARE(\theta) = \frac{\left|G^{-1}(\theta)\right|}{\left|I^{-1}(\theta)\right|},$$

where the symbol $\left|A\right|$ denotes the determinant of a square matrix A. In literature, one may find other definitions of efficiency. It may be noted that the determinant of the variance covariance matrix corresponds to the volume of the confidence ellipsoid obtained based on the asymptotic distribution. The determinant of variance covariance matrix is also known as generalized variance (Wilks (1932), Sengupta (2004)). JARE can then be interpreted as the ratio of generalized variances, and its value indicates which of $\hat{\theta}_{CL}$ or $\hat{\theta}_{ML}$ provides tighter confidence regions.

As noted in chapter 2, most of these joint distributions obtained by specification of conditionals either have a normalizing constant which makes the ML estimation inconvenient. Without a closed form expression for the normalizing constant, computation of ML estimates is not possible. Thus we will consider a composite likelihood. In the context of conditionally specified distributions, product of the conditionals offers itself as a ready option. Note that, under independence this will reduce to the true joint distribution. Furthermore, it meaningfully incorporates the joint parameters using the two marginals. Arnold and Strauss (1991a) refer to this function as Pseudolikelihood (PL). They show that, under regularity conditions, the resulting PL estimators are consistent and asymptotically normal. However, they also note that, PL estimators have slightly reduced efficiency compared to ML esti-

mators. Note that the PL is a special case of CL. Therefore, we will use the general theory about CL presented earlier to obtain the Godambe information matrix, and JARE of the PL with respect to ML estimators. To formally connect the definition of CL presented here to the PL given in Arnold et al. (2001), we consider observations $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ from some bivariate conditionally specified density $f(x, y; \theta), \theta \in \Theta$; and $f(x|y; \theta)$ and $f(y|x; \theta)$ are the compatible conditional distributions. By following the definition (4.2), let $f(x|y; \theta)$ and $f(y|x; \theta)$ be the associated likelihoods and let $k = 1$ and $\omega_k = 1$. Arnold et al. (2001) define the pseudolikelihood estimates as follows.

Definition 4.3. *Suppose that we have n observations* $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ *from some bivariate conditionally specified density* $f(x, y; \theta), \theta \in \Theta$. *The maximum pseudo likelihood estimate of* θ, *say* $\hat{\theta}$, *is a usually unique value of* θ *for which,*

$$pl(\hat{\theta}) = \prod_{i=1}^{n} f_{X|Y}(x_i|y_i; \hat{\theta}) f_{Y|X}(y_i|x_i; \hat{\theta}) = \max_{\theta \in \Theta} \prod_{i=1}^{n} f_{X|Y}(x_i|y_i; \theta) f_{Y|X}(y_i|x_i; \theta).$$

Thus, we note that PL defined above is a special case of CL of definition (4.2).

4.2 Maximum Likelihood Estimation for Bivariate Normal and Bernoulli Conditionals

We will start by stating the likelihood function of the our joint distribution derived in chapter 3. The likelihood function of the joint distribution is,

$$L(y, r; \underset{\sim}{\theta}) = \prod_{i=1}^{n} exp(m_{00} + r m_{10}).exp(Q(y, r)), \qquad r = 0, 1 \text{ and } y \in \mathbf{R} \tag{4.6}$$

where,

$$Q(y,r) = (m_{01}+rm_{11})y_1+(m_{02}+rm_{12})y_2+(m_{03}+rm_{13})y_1^2+(m_{04}+rm_{14})y_1y_2+(m_{05}+rm_{15})y_2^2$$

We can also write (4.6) as,

$$L(f(y,r;\underset{\sim}{\hat{\theta}})) = \prod_{i=1}^{n} \frac{exp(rm_{10}+Q(y,r))}{\sum_{r=0}^{1} exp(rm_{10}) \int_{y1=-\infty}^{\infty} \int_{y2=-\infty}^{\infty} exp(Q(y,r))dy_2dy_1} \tag{4.7}$$

Notice that,the normalizing constant m_{00} is present in the equation (4.6). The expression for m_{00} can be found in chapter 3 which is

$$m_{00} = -ln\Bigg[2\pi\sigma_{11}\sigma_{22}\sqrt{1-\rho^2} \sum_{r=0}^{1} |I-2A\Sigma|^{-1/2}.exp\Bigg(rm_{10} - C(r) + 2b^T\Sigma(I-2A\Sigma)^{-1}b + \mu^{(r)T}(I-2A\Sigma)^{-1}(2b+A\mu^{(r)})\Bigg)\Bigg],$$

where

$$C(r) = \frac{-(\sigma_{22}^2\mu_1^{(r)^2} - 2\rho\sigma_{11}\sigma_{22}\mu_1^{(r)}\mu_2^{(r)} + \sigma_{11}^2\mu_2^{(r)^2})}{2(1-\rho^2)\sigma_{11}^2\sigma_{22}^2}$$

$$b = \left(\frac{2\rho\sigma_{11}\sigma_{22}\mu_2^{(r)} - 2\sigma_{22}^2\mu_1^{(r)}}{4(1-\rho^2)} + \frac{m_{01}+rm_{11}}{2}, \quad \frac{2\rho\sigma_{11}\sigma_{22}\mu_1^{(r)} - 2\mu_2^{(r)}\sigma_{11}^2}{4(1-\rho^2)\sigma_{11}^2\sigma_{22}^2} + \frac{m_{02}+rm_{12}}{2}\right)^T$$

and

$$A = \begin{pmatrix} \frac{1}{2(1-\rho^2)\sigma_{11}^2} + m_{03} + rm_{13} & m_{04}+rm_{14} \\ \frac{-\rho}{(1-\rho^2)\sigma_{11}\sigma_{22}} & \frac{1}{2(1-\rho^2)\sigma_{22}^2} + m_{05} + rm_{15} \end{pmatrix}$$

It is clear that in spite of the likelihood function having a closed form, the expression for the normalizing constant and obtaining its' derivatives is intractable. We present the form of the score function and briefly discuss the Fisher information matrix in the next section.

4.2.1 Derivation of Score Function and Fisher Information Matrix

Our goal is to obtain the information matrix for the joint distribution given in equation (4.6). Consider the conditionally specified joint distribution $f(y \mid \underset{\sim}{\theta})$ and y now denote all the data and θ the parameters of both conditionals. Thus the $s \times 1$ vector valued score function can be written as,

$$S(\underset{\sim}{\theta}) = \frac{\partial log\ \mathrm{f}(y \mid \underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}}$$

The expression for the joint distribution provided by equation (4.6) is in terms of the natural parameters m_{ij}'s. We can do this by using the Theorem (4.1) stated earlier which essentially based on the chain rule of differentiation. We follow the notations of Theorem (4.1) and define the following 10×1 parameter vectors:

$\underset{\sim}{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10})^T = (m_{01}, m_{02}, m_{03}, m_{04}, m_{05}, m_{10}, m_{11}, m_{10}, m_{11}, m_{12})$ and

$$\underset{\sim}{\xi} = (\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6, \xi_7, \xi_8, \xi_9, \xi_{10})^T = (\mu_1^{(0)}, \mu_2^{(0)}, \mu_1^{(1)}, \mu_2^{(1)}, \sigma_{11}, \sigma_{22}, \rho, \alpha_0, \alpha_1, \alpha_2) \tag{4.8}$$

The objective is to obtain $S(\underset{\sim}{\theta})$ and $S^*(\underset{\sim}{\xi})$, the reparameterizations of the score vectors which can be used to define the maximum likelihood equations. Further, using the equations (3.9) - (3.14), each component of θ, namely one of the m_{ij}'s, can be written in terms of the components of ξ. Then, we have in the notations of Theorem (4.1),

$$\theta_i = h_i(\underset{\sim}{\xi}) \qquad i = 1, 2, ..., 10$$

and the 10×10 matrix $\mathcal{D}$ of derivatives,

$$\mathcal{D} = \left(\left(\frac{\partial h_j(\xi)}{\partial \xi_i} \right) \right) \qquad i \neq j = 1, 2, ..., 10$$

Elements of $\mathcal{D}$ can be obtained by differentiating $h_j(\xi)$ with respect to ξ_i's. Hence, $\mathcal{D}$ is as follows.

$$\mathcal{D} = \begin{pmatrix} \frac{\partial m_{01}}{\partial \mu_1^{(0)}} & \frac{\partial m_{01}}{\partial \mu_2^{(0)}} & \frac{\partial m_{01}}{\partial \mu_1^{(1)}} & \frac{\partial m_{01}}{\partial \mu_2^{(1)}} & \frac{\partial m_{01}}{\partial \sigma_{11}} & \frac{\partial m_{01}}{\partial \sigma_{22}} & \frac{\partial m_{01}}{\partial \rho} & 0 & -1 & 0 \\ \frac{\partial m_{02}}{\partial \mu_1^{(0)}} & \frac{\partial m_{02}}{\partial \mu_2^{(0)}} & \frac{\partial m_{02}}{\partial \mu_1^{(1)}} & \frac{\partial m_{02}}{\partial \mu_2^{(1)}} & \frac{\partial m_{02}}{\partial \sigma_{11}} & \frac{\partial m_{02}}{\partial \sigma_{22}} & \frac{\partial m_{02}}{\partial \rho} & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & \frac{\partial m_{03}}{\partial \sigma_{11}} & 0 & \frac{\partial m_{03}}{\partial \rho} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{\partial m_{04}}{\partial \sigma_{11}} & \frac{\partial m_{04}}{\partial \sigma_{22}} & \frac{\partial m_{04}}{\partial \rho} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{\partial m_{05}}{\partial \sigma_{22}} & \frac{\partial m_{05}}{\partial \rho} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Similarly, the score function $S(\theta)$ is,

$$S(\theta)_{10\times 1} = \begin{pmatrix} \frac{\partial l}{\partial m_{01}} & \frac{\partial l}{\partial m_{02}} & \frac{\partial l}{\partial m_{03}} & \frac{\partial l}{\partial m_{04}} & \frac{\partial l}{\partial m_{05}} & \frac{\partial l}{\partial m_{10}} & \frac{\partial l}{\partial m_{11}} & \frac{\partial l}{\partial m_{10}} & \frac{\partial l}{\partial m_{11}} & \frac{\partial l}{\partial m_{12}} \end{pmatrix}^T.$$

Now $S(\xi)$, the score function in terms of the parameter ξ can be obtained by using equation (4.2). For example, the first component of $S(\xi)$ is given as,

$$(S(\xi))_1 = \frac{\partial l}{\partial m_{01}} \times \frac{\partial m_{01}}{\partial \mu_1^{(0)}} + \frac{\partial l}{\partial m_{02}} \times \frac{\partial m_{01}}{\partial \mu_2^{(0)}} + \frac{\partial l}{\partial m_{03}} \times \frac{\partial m_{01}}{\partial \mu_1^{(1)}} + \frac{\partial l}{\partial m_{04}} \times \frac{\partial m_{01}}{\partial \mu_2^{(1)}} + \frac{\partial l}{\partial m_{05}} \times \frac{\partial m_{01}}{\partial \sigma_{11}} + \frac{\partial l}{\partial m_{10}} \times \frac{\partial m_{01}}{\partial \sigma_{22}} + \frac{\partial l}{\partial m_{11}} \times \frac{\partial m_{01}}{\partial \rho} - r_i \frac{\partial l}{\partial m_{11}}$$

where,

$$\frac{\partial l}{\partial m_{10}} = \sum_{i=1}^{n} \frac{\partial m_{00}}{\partial m_{10}} + \sum_{i=1}^{n} r_i, \qquad \frac{\partial l}{\partial m_{12}} = \sum_{i=1}^{n} \frac{\partial m_{00}}{\partial m_{12}} + \sum_{i=1}^{n} r_i y_{2i}$$

$$\frac{\partial l}{\partial m_{01}} = \sum_{i=1}^{n} \frac{\partial m_{00}}{\partial m_{01}} + \sum_{i=1}^{n} y_{1i}, \qquad \frac{\partial l}{\partial m_{03}} = \sum_{i=1}^{n} \frac{\partial m_{00}}{\partial m_{03}} + \sum_{i=1}^{n} y_{1i}^2$$

$$\frac{\partial l}{\partial m_{11}} = \sum_{i=1}^{n} \frac{\partial m_{00}}{\partial m_{11}} + \sum_{i=1}^{n} r_i y_{1i}, \qquad \frac{\partial l}{\partial m_{04}} = \sum_{i=1}^{n} \frac{\partial m_{00}}{\partial m_{04}} + \sum_{i=1}^{n} y_{1i} y_{2i}$$

$$\frac{\partial l}{\partial m_{02}} = \sum_{i=1}^{n} \frac{\partial m_{00}}{\partial m_{02}} + \sum_{i=1}^{n} y_{2i} \text{ and} \qquad \frac{\partial l}{\partial m_{05}} = \sum_{i=1}^{n} \frac{\partial m_{00}}{\partial m_{05}} + \sum_{i=1}^{n} y_{2i}^2$$

Similarly, the remaining components of the score function $S(\xi)$ can be obtained as follows:

$$(S(\xi))_2 = \frac{\partial l}{\partial m_{01}} \times \frac{\partial m_{02}}{\partial \mu_1^{(0)}} + \frac{\partial l}{\partial m_{02}} \times \frac{\partial m_{02}}{\partial \mu_2^{(0)}} + \frac{\partial l}{\partial m_{03}} \times \frac{\partial m_{02}}{\partial \mu_1^{(1)}} + \frac{\partial l}{\partial m_{04}} \times \frac{\partial m_{02}}{\partial \mu_2^{(1)}} + \frac{\partial l}{\partial m_{05}} \times \frac{\partial m_{02}}{\partial \sigma_{11}} + \frac{\partial l}{\partial m_{10}} \times \frac{\partial m_{02}}{\partial \sigma_{22}} + \frac{\partial l}{\partial m_{11}} \times \frac{\partial m_{02}}{\partial \rho} - r_i \frac{\partial l}{\partial m_{12}}$$

$$(S(\xi))_3 = \frac{\partial l}{\partial m_{05}} \times \frac{\partial m_{03}}{\partial \sigma_{11}} + \frac{\partial l}{\partial m_{11}} \times \frac{\partial m_{03}}{\partial \rho}$$

$$(S(\xi))_4 = \frac{\partial l}{\partial m_{05}} \times \frac{\partial m_{04}}{\partial \sigma_{11}} + \frac{\partial l}{\partial m_{10}} \times \frac{\partial m_{04}}{\partial \sigma_{22}} + \frac{\partial l}{\partial m_{11}} \times \frac{\partial m_{04}}{\partial \rho}$$

$$(S(\xi))_5 = \frac{\partial l}{\partial m_{10}} \times \frac{\partial m_{05}}{\partial \sigma_{22}} + \frac{\partial l}{\partial m_{11}} \times \frac{\partial m_{05}}{\partial \rho}$$

$$(S(\xi))_6 = \frac{\partial l}{\partial m_{10}}$$

$$(S(\xi))_7 = \frac{\partial l}{\partial m_{11}}$$

$$(S(\xi))_8 = \frac{\partial l}{\partial m_{10}}$$

$$(S(\xi))_9 = \frac{\partial l}{\partial m_{11}}$$

$$(S(\xi))_{10} = \frac{\partial l}{\partial m_{12}}$$

All the required partial derivatives are given in appendix (A.1). Therefore, the Fisher information matrix (10×10) based on original parameter space is:

$$I^*(\xi) = E[-S(\xi)S(\xi)^T]_{(10\times 10)}$$

As noted before, because of the intractability of m_{00}, finding analytical solutions for estimates is difficult. However, finding numerical solutions for the log likelihood equations in feasible.

4.3 Composite Likelihood Estimation

As mentioned in Section 4.1, we follow the Arnold and Strauss (1991a) and replace the full likelihood by the product of conditional densities. This is motivated by the fact that our full likelihood is constructed from the conditionals. Furthermore, the conditionals (one is Bernoulli; the other bivariate Gaussian) are relatively easier to handle.

Hence, by applying equation (4.3) to our problem we get the likelihood as,

$$\mathrm{L}(\Theta, y, r) = \prod_{i=1}^{n} f(\underset{\sim}{y_i} \mid R = r_i) f(r_i \mid Y = \underset{\sim}{y_i})$$

where, $\Theta = (\mu^{(r)}, \Sigma, \alpha_0, \alpha_1, \alpha_2)$ and $\omega_k = 1$. Arnold and Strauss (1991a) refers to the above as the pseudolikelihood (PL) function and parameter estimates called as PL estimators. Now, we can write $\mathrm{CL}(\Theta, y, r) = \mathrm{PL}(\Theta, y, r)$. Note that, $f(\underset{\sim}{y} \mid R = r)$ is bivariate normal with mean $\mu^{(r)}$ and variance covariance matrix Σ; and $\mu^{(r)}$ can be rewritten as $\mu^{(r)} = \mu^{(0)}(1 - r) + \mu^{(1)} r$. Thus the likelihood can be written as,

$$\begin{aligned}
\mathrm{L}_{CL}(\Theta, y, r) = & \prod_{i=1}^{n} \frac{1}{2\pi |\Sigma|^{1/2}} \\
& \times exp\left\{ -\frac{1}{2} (\underset{\sim}{y_i} - \mu^{(1)} r_i - \mu^{(0)}(1 - r_i))^T \Sigma^{-1} (\underset{\sim}{y_i} - \mu^{(1)} r_i - \mu^{(0)}(1 - r_i)) \right\} \\
& \times \pi(\underset{\sim}{y_i})^{r_i} (1 - \pi(\underset{\sim}{y_i}))^{1 - r_i}.
\end{aligned}$$

Now, using the expression for $\pi(y)$ and letting $y^* = (1, y_1, y_2)^T$,

$$\begin{aligned} \mathrm{L}_{CL}(\Theta, y, r) = &\prod_{i=1}^{n} \frac{1}{2\pi|\Sigma|^{1/2}} \\ &\times exp\left\{ -\frac{1}{2}(\underset{\sim}{y}_i - \mu^{(1)} r_i - \mu_0(1-r_i))^T \Sigma^{-1} (\underset{\sim}{y}_i - \mu^{(1)} r_i - \mu^{(0)}(1-r_i)) \right\} \\ &\times \left\{ \frac{exp(\underset{\sim}{\alpha}^T y^*)}{1 + exp(\underset{\sim}{\alpha}^T y^*)} \right\}^{r_i} \left\{ \frac{1}{1 + exp(\underset{\sim}{\alpha}^T y^*)} \right\}^{1-r_i}, \end{aligned}$$

which can be rewritten as,

$$\begin{aligned} \mathrm{L}_{CL}(\Theta, y, r) = &\prod_{i=1}^{n} \frac{1}{2\pi|\Sigma|^{1/2}} \\ &\times exp\left\{ -\frac{1}{2}(\underset{\sim}{y}_i - \mu^{(1)} r_i - \mu_0(1-r_i))^T \Sigma^{-1} (\underset{\sim}{y}_i - \mu^{(1)} r_i - \mu^{(0)}(1-r_i)) \right\} \\ &\times \frac{\{exp(\underset{\sim}{\alpha}^T y^*)\}^{r_i}}{1 + exp(\underset{\sim}{\alpha}^T y^*)}, \end{aligned}$$

where, $\alpha = (\alpha_0, \alpha_1, \alpha_2)^T$ and $\underset{\sim}{y}^* = (1, \underset{\sim}{y}_i)^T$. Thus, the log likelihood function becomes,

$$\begin{aligned} \ell_{CL}(\Theta, y, r) = &-n\ln(2\pi) - \frac{n}{2}\ln(|\Sigma|) \\ &- \frac{1}{2}\sum_{i=1}^{n} (\underset{\sim}{y}_i - \mu^{(1)} r_i - \mu^{(0)}(1-r_i))^T \Sigma^{-1} (\underset{\sim}{y}_i - \mu^{(1)} r_i - \mu^{(0)}(1-r_i)) \\ &+ \sum_{i=1}^{n} r_i \alpha^T y^* - \sum_{i=1}^{n} \ln(1 + exp(\alpha^T y^*)). \end{aligned} \tag{4.9}$$

By solving the log likelihood equation one can find the numerical solutions for the parameters without much hassle due to the fact that now we have a much simpler likelihood function which does not involves a normalizing constant that is difficult to evaluate.

4.4 Godambe Information Matrix (GIM)

In this section we consider the theoretical properties of the composite likelihood function in defined using the conditionals specifically belongs to the exponential family . The core of this section is the derivation of GIM for conditionals belongs to exponential family of distributions.

4.4.1 Conditionally Exponential family (CEF)

We start by presenting our theorem followed by the complete proof. Note that, we are following the notations of definition 2.2.2; exponential family of distributions defined by Arnold and Strauss (1991a).

Theorem 4.2. *Suppose $f(Y, X; \theta, \tau)$ be the joint distribution consists with $f_1(x; \underset{\sim}{\theta}(y))$ and $f_2(y; \underset{\sim}{\tau}(x))$ as its conditional distributions belongs to l_1-parameter and l_2- parameter exponential family of distributions respectively and has the form,*

$$f(x \mid Y = y) = f_1(x; \underset{\sim}{\theta}(y)) = r_1(x)\beta_1(\underset{\sim}{\theta}(y))exp\Big\{ \sum_{i=1}^{l_1} \theta_i(y) q_{1i}(x) \Big\}$$

$$f(y \mid X = x) = f_2(y; \underset{\sim}{\tau}(x)) = r_2(y)\beta_2(\underset{\sim}{\tau}(x))exp\Big\{ \sum_{j=1}^{l_2} \tau_j(x) q_{2j}(y) \Big\}$$

And the composite likelihood of the form,

$$L_{CL}(\underset{\sim}{\gamma}; x, y) = \prod_{i=1}^{n} f_{x|Y=y_i}(x_i, \theta(y)) f_{y|X=x_i}(y_i, \tau(x))$$

where, $\underset{\sim}{\gamma} = \Big(\underset{\sim}{\theta}(y), \underset{\sim}{\tau}(x)\Big) = \Big(\theta_1(y), \theta_2(y), ..., \theta_{l_1}(y), \tau_1(x), \tau_2(x), ..., \tau_{l_2}(x)\Big)_{1\times(l_1+l_2)}$

Further, $H(\cdot)$,$J(\cdot)$ are sensitivity matrix and variability matrix respectively; $G(.)$ Godambe information matrix. Then, $G(\gamma)$ can be written as,

$$G(\gamma) = \begin{pmatrix} -E_{(Y,X)}(V_{X/Y}(\gamma)) & \underset{\sim}{0} \\ \underset{\sim}{0} & -E_{(Y,X)}(V_{Y/X}(\gamma)) \end{pmatrix}$$

where,

$$V_{Y|X}(\gamma) = \begin{pmatrix} \frac{\partial^2 \ell_{y/x}}{\partial^2 \tau_1} & \frac{\partial^2 \ell_{y/x}}{\partial \tau_1 \partial \tau_2} & \cdots & \frac{\partial^2 \ell_{y/x}}{\partial \tau_1 \partial \tau_{l_2}} \\ \frac{\partial^2 \ell_{y/x}}{\partial \tau_1 \partial \tau_2} & \cdots & \cdots & \frac{\partial^2 \ell_{y/x}}{\partial \tau_2 \partial \tau_{l_2}} \\ \vdots & & & \vdots \\ \frac{\partial^2 \ell_{y/x}}{\partial \tau_1 \partial \tau_{l_2}} & \frac{\partial^2 \ell_{y/x}}{\partial \tau_2 \partial \tau_{l_2}} & \cdots & \frac{\partial^2 \ell_{y/x}}{\partial^2 \tau_{l_2}} \end{pmatrix}_{l_2 \times l_2}$$

and

$$V_{X|Y}(\gamma) = \begin{pmatrix} \frac{\partial^2 \ell_{x/y}}{\partial^2 \theta_1} & \frac{\partial^2 \ell_{x/y}}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \ell_{x/y}}{\partial \theta_1 \partial \theta_{l_1}} \\ \frac{\partial^2 \ell_{x/y}}{\partial \theta_1 \partial \theta_2} & \cdots & \cdots & \frac{\partial^2 \ell_{x/y}}{\partial \theta_2 \partial \theta_{l_1}} \\ \vdots & & & \vdots \\ \frac{\partial^2 \ell_{x/y}}{\partial \theta_1 \partial \theta_{l_1}} & \frac{\partial^2 \ell_{x/y}}{\partial \theta_2 \partial \theta_{l_1}} & \cdots & \frac{\partial^2 \ell_{x/y}}{\partial^2 \theta_{l_1}} \end{pmatrix}_{l_1 \times l_1}$$

Also, $H(\gamma) = J(\gamma)$; called as information unbiased. Note that, $G(\gamma) = H(\gamma)$.

Proof. We will start by defining the composite likelihood. Suppose, $k = n$ and $\omega_i = 1$. Then the equation (4.7) becomes,

$$L_{CL}(\underset{\sim}{\gamma}; x, y) = \prod_{i=1}^{n} L_i(\underset{\sim}{\gamma}; y, x)$$

Where, $\underset{\sim}{\gamma} = \left(\theta(\underset{\sim}{y}), \tau(\underset{\sim}{x})\right) = \left(\theta_1(y), \theta_2(y), ..., \theta_{l_1}(y), \tau_1(x), \tau_2(x), ..., \tau_{l_2}(x)\right)_{1 \times (l_1 + l_2)}$

Let the product of the conditionals $f_1(x; \theta(\underset{\sim}{y}))$ and $f_2(y; \tau(\underset{\sim}{x}))$ as the composite

likelihood. Thus, the above expression can be rewritten as,

$$\begin{aligned} L_{CL}(\underset{\sim}{\gamma}; x, y) &= \prod_{i=1}^{n} f_1(x_i, \theta(y)) f_2(y_i, \tau(x)) \\ &= \prod_{i=1}^{n} f_{x|Y=y_i}(x_i, \theta(y)) f_{y|X=x_i}(y_i, \tau(x)) \end{aligned}$$

Log composite likelihood is,

$$\ell(\underset{\sim}{\gamma}; x, y) = \sum_{i=1}^{n} \log[f_1(x_i, \theta(y))] + \sum_{i=1}^{n} \log[f_2(y_i, \tau(x))].$$

Composite likelihood score function $U(\underset{\sim}{\gamma}; x, y)$ can be derived as,

$$\begin{aligned} U(\underset{\sim}{\gamma}; x, y) &= \nabla_{\gamma} \ln(L_{CL}(\underset{\sim}{\gamma}; x, y)) \\ &= \frac{\partial}{\partial \underset{\sim}{\gamma}} [\ell_{CL}(\underset{\sim}{\gamma}; x, y))] \\ &= \frac{\partial}{\partial \underset{\sim}{\gamma}} \left[\sum_{i=1}^{n} \log[f_1(x_i, \theta(y))] + \sum_{i=1}^{n} \log[f_2(y_i, \tau(x))] \right] \\ &= \frac{\partial}{\partial \underset{\sim}{\gamma}} \left[\ell_{x|y} + \ell_{y|x} \right] \\ &= U(x \mid y; \underset{\sim}{\gamma}) + U(y \mid x; \underset{\sim}{\gamma}) \end{aligned}$$

where, $\ell_{x|y}$ and $\ell_{y|x}$ are log likelihood functions of $f(x|y) = f_1(x, \theta(y))$ and $f(y|x) = f_2(y, \tau(x))$ respectively. Further, $U(x \mid y; \underset{\sim}{\gamma})$ and $U(y \mid x; \underset{\sim}{\gamma})$ are score functions of $f(x|y) = f_1(x, \theta(y))$ and $f(y|x) = f_2(y, \tau(x))$ respectively. We will start by differentiating $\ell_{x|y}$ with respect to $\underset{\sim}{\gamma}$. Log likelihood function of $f_1(x; \theta(y))$ is,

$$\ell_{x|y} = \log(f(x|y)) = \sum_{k=1}^{n} \log(r_1(x_k)) + n\beta_1(\theta(y)) + \sum_{k=1}^{n} \sum_{i=1}^{l_1} \theta_i(y) q_{1i}(x_k).$$

Since $\underset{\sim}{\gamma} = \left(\theta(\underset{\sim}{y}), \tau(\underset{\sim}{x}) \right)$, We start by differentiating the above function with respect

to $\theta(\underset{\sim}{y})$,

$$\begin{aligned}
\frac{\partial \ell_{x|y}}{\partial \theta_1} &= n\frac{\partial \beta_1(\theta(y))}{\partial \theta_1} + \sum_{k=1}^{n} q_{11}(x_k) \\
\frac{\partial \ell_{x|y}}{\partial \theta_2} &= n\frac{\partial \beta_1(\theta(y))}{\partial \theta_2} + \sum_{k=1}^{n} q_{12}(x_k) \\
\frac{\partial \ell_{x|y}}{\partial \theta_3} &= n\frac{\partial \beta_1(\theta(y))}{\partial \theta_3} + \sum_{k=1}^{n} q_{13}(x_k) \\
&\vdots \\
\frac{\partial \ell_{x|y}}{\partial \theta_{l_1}} &= n\frac{\partial \beta_1(\theta(y))}{\partial \theta_{l_1}} + \sum_{k=1}^{n} q_{1l_1}(x_k)
\end{aligned}$$

General form of the differentiated expression can be written as,

$$\frac{\partial \ell_{x|y}}{\partial \theta_{l_1}} = n\frac{\partial \beta_1(\theta(y))}{\partial \theta_{l_1}} + \sum_{k=1}^{n} q_{1l_1}(x_k), \qquad l_1 = 1, 2, 3, ..., l_1.$$

Differentiate with respect to $\tau(\underset{\sim}{x})$ gives, $\frac{\partial \ell_{x|y}}{\partial \tau_{l_2}}. = 0 \quad ; l_2 = 1, 2, 3, ..., l_2$. This is due to the fact that, $\ell_{x|y}$ contains θ parameters only. Therefore, the score function of $f(x \mid y)$ can be written as,

$$U(x \mid y; \underset{\sim}{\gamma}) = \left(\frac{\partial \ell_{x|y}}{\partial \theta_1}, \frac{\partial \ell_{x|y}}{\partial \theta_2}, \cdots \frac{\partial \ell_{x|y}}{\partial \theta_{l_1}}, \underbrace{\frac{\partial \ell_{x|y}}{\partial \tau_1}, \frac{\partial \ell_{x|y}}{\partial \tau_2}, \cdots \frac{\partial \ell_{x|y}}{\partial \tau_{l_2}}}_{0} \right). \tag{4.10}$$

Similarly we differentiate $f_2(y; \tau(x))$ with respect to $\underset{\sim}{\gamma}$. Log likelihood function of $f_2(y; \tau(x))$ is,

$$\ell_{y|x} = \log(f(y \mid x)) = \sum_{k=1}^{n} \log(r_2(y_k)) + n\beta_2(\tau(x)) + \sum_{k=1}^{n} \sum_{j=1}^{l_2} \tau_j(x) q_{2j}(y_k).$$

Thus the general expression,

$$\frac{\partial l_{y/x}}{\partial \tau_{l_2}} = n\frac{\partial \beta_2(\tau(x))}{\partial \tau_{l_2}} + \sum_{k=1}^{n} q_{2p_2}(y_k) \qquad ; l_2 = 1, 2, 3, ..., l_2$$

Same as previous we have $\frac{\partial \ell_{y|x}}{\partial \theta_{l_2}} = 0 \quad ; l_2 = 1, 2, 3, ..., l_2$. The score function of $f(y \mid x)$ is,

$$U(y \mid x; \underset{\sim}{\gamma}) = \Bigg(\underbrace{\frac{\partial \ell_{y|x}}{\partial \theta_1}, \frac{\partial \ell_{y|x}}{\partial \theta_2}, \cdots \frac{\partial \ell_{y|x}}{\partial \theta_{l_1}}}_{0}, \frac{\partial \ell_{y|x}}{\partial \tau_1}, \frac{\partial \ell_{y|x}}{\partial \tau_2}, \cdots \frac{\partial \ell_{y|x}}{\partial \tau_{l_2}} \Bigg) \tag{4.11}$$

From equation (4.10) and (4.11) we can easily obtain the score function of composite likelihood function given in equation (**??**). Now let us derive the sensitivity matrix $H(\gamma)$,

$$\begin{aligned} H(\gamma) &= E_\gamma(-\nabla_\gamma U(\underset{\sim}{\gamma}; x, y)) \\ &= E_{(X,Y)}\Bigg\{ \frac{-\partial}{\partial \gamma}\Bigg[\sum_{k=1}^{n} U(x_k \mid y_k) + \sum_{k=1}^{n} U(y_k \mid x_k) \Bigg] \Bigg\} \\ &= E_{(X,Y)}\Bigg\{ \frac{-\partial}{\partial \gamma} \sum_{k=1}^{n} U(x_k \mid y_k) \Bigg\} + E_{(X,Y)}\Bigg\{ \frac{-\partial}{\partial \gamma}\Bigg\{ \sum_{k=1}^{n} U(y_k \mid x_k) \Bigg\}. \end{aligned}$$

Thus,

$$H(\gamma) = E_X\Bigg\{ E_{(X|Y)}\Bigg[\frac{-\partial}{\partial \gamma} \sum_{k=1}^{n} U(x_k \mid y_k) \Bigg] \Bigg\} + E_Y\Bigg\{ E_{(Y|X)}\Bigg[\frac{-\partial}{\partial \gamma}\Bigg\{ \sum_{k=1}^{n} U(y_k \mid x_k) \Bigg] \Bigg\}.$$

Further $H(\gamma)$ can be written as,

$$H(\gamma) = E_X\Bigg\{ J(Y \mid X = x_i) \Bigg\} + E_Y\Bigg\{ J(X \mid Y = y_i) \Bigg\} \tag{4.12}$$

where J is the variability matrix. Note that, $J(\gamma) = Var_\gamma\Bigg\{ \nabla_\gamma U(\gamma; X, Y) \Bigg\}$ and this can be rewritten as,

$$J(\gamma) = E_{(X,Y)}[U(\gamma; X, Y)U(\gamma; X, Y)^T] - \Bigg\{ E_{(X,Y)}[U(\gamma; X, Y)] \Bigg\} \Bigg\{ E_{(X,Y)}[U(\gamma; X, Y)] \Bigg\}^T \tag{4.13}$$

Let us first start by finding the expression for $E_{(X,Y)}[U(\gamma; X, Y)]$,

$$\begin{aligned}
E_{(X,Y)}[U(\gamma; X, Y)] &= E_{(X,Y)}\left[\sum_{k=1}^{n} U(X \mid Y = y_k) + \sum_{k=1}^{n} U(Y \mid X = x_k)\right] \\
&= E_{(X,Y)}\left[\sum_{k=1}^{n} U(X \mid Y = y_k)\right] + E_{(X,Y)}\left[\sum_{k=1}^{n} U(Y \mid X = x_k)\right] \\
&= E_Y\Bigg[\underbrace{E_{(X|Y)}\left\{\sum_{k=1}^{n} U(X \mid Y = y_k)\right\}}_{\text{A}}\Bigg] + E_X\Bigg[\underbrace{E_{(Y|X)}\left\{\sum_{k=1}^{n} U(Y \mid X = x_k)\right\}}_{\text{B}}\Bigg]
\end{aligned}$$

The expectation of score function under its true likelihood is zero. Therefore, A and B becomes zero. Thus,

$$\begin{aligned}
E_{(X,Y)}[U(\gamma; X, Y)] &= E_Y(0) + E_X(0) \\
&= 0.
\end{aligned}$$

Then, from equation (4.13) $J(\gamma)$ becomes,

$$\begin{aligned}
J(\gamma) &= E_{(X,Y)}[U(\gamma; X, Y).U(\gamma; X, Y)^T] \\
&= E_{(X,Y)}\Big\{[U(Y \mid X, \gamma) + U(X \mid Y, \gamma)].[U(Y \mid X, \gamma) + U(X \mid Y, \gamma)]^T\Big\} \\
&= E_{(X,Y)}\Big(U(Y \mid X, \gamma).U(Y \mid X, \gamma)^T\Big) + E_{(X,Y)}\Big(U(Y \mid X, \gamma).U(X \mid Y, \gamma)^T\Big) + \\
&\qquad E_{(X,Y)}\Big(U(X \mid Y, \gamma).U(Y \mid X, \gamma)^T\Big) + E_{(X,Y)}\Big(U(X \mid Y, \gamma).U(X \mid Y, \gamma)^T\Big)
\end{aligned}$$

Note that, $U(X \mid Y, \gamma)$ depends only on $\underset{\sim}{\theta}$ and $U(Y \mid X, \gamma)$ depends only on $\underset{\sim}{\tau}$. According to equation (4.10) and (4.11), $U(X \mid Y, \gamma)$ and $U(Y \mid X, \gamma)$ will have zeros in complementary positions. Hence, $U(Y \mid X, \gamma).U(X \mid Y, \gamma)^T = 0$ and $U(X \mid Y, \gamma).U(Y \mid X, \gamma)^T = 0$

$$J(\gamma) = E_{(X,Y)}\Big(U(Y \mid X, \gamma).U(Y \mid X, \gamma)^T\Big) + E_{(X,Y)}\Big(U(X \mid Y, \gamma).U(X \mid Y, \gamma)^T\Big)$$

$$= E_X\left(E_{Y|X}\left[U(Y \mid X,\gamma).U(Y \mid X,\gamma)^T\right]\right) + E_Y\left(E_{X|Y}\left[U(X \mid Y,\gamma).U(X \mid Y,\gamma)^T\right]\right) \tag{4.14}$$

Note that, equation (4.14) can be written as,

$$J(\gamma) = \int_{x\in\mathbf{R}} J_{Y|X=x}(\gamma) f_X(x)dx + \int_{y\in\mathbf{R}} J_{X|Y=y}(\gamma) f_Y(y)dy$$

and equation (4.12) can be express as,

$$H(\gamma) = \int_{x\in\mathbf{R}} J_{Y|X=x}(\gamma) f_X(x)dx + \int_{y\in\mathbf{R}} J_{X|Y=y}(\gamma) f_Y(y)dy$$

Thus, we can conclude that $H(\gamma) = J(\gamma)$. This means that Sensitivity matrix is equal to Variability matrix when the composite likelihood contains the product of conditionals. We called this scenario as information unbiased. Moreover, according to Varin et al. (2011), Godambe information matrix is

$$G(\gamma) = H(\gamma)J(\gamma)^{-1}H(\gamma)$$

Since $H(\gamma) = J(\gamma)$, the information matrix becomes $G(\gamma) = H(\gamma)$. That is, Godambe information matrix is equal to sensitivity matrix.

We will further simplifying the expression $H(\gamma)$.

$$H(\gamma) = E_{(X,Y)}\left[\frac{-\partial}{\partial\gamma}\sum_{k=1}^{n} U(x_k \mid y_k)\right] + E_{(X,Y)}\left[\frac{-\partial}{\partial\gamma}\left\{\sum_{k=1}^{n} U(y_k \mid x_k)\right\}\right]$$

and

$$\frac{\partial}{\partial\gamma}\sum_{k=1}^{n} U(x_k \mid y_k) = \begin{pmatrix} V_{X|Y} & \underset{\sim}{0}_{l_2\times l_1} \\ \underset{\sim}{0}_{l_1\times l_2} & \underset{\sim}{0}_{l_2\times l_2} \end{pmatrix}_{(l_1+l_2)\times(l_1+l_2)} \text{ and } \frac{\partial}{\partial\gamma}\sum_{k=1}^{n} U(y_k \mid x_k) = \begin{pmatrix} \underset{\sim}{0}_{l_1\times l_1} & \underset{\sim}{0}_{l_2\times l_1} \\ \underset{\sim}{0}_{l_1\times l_2} & V_{Y|X} \end{pmatrix}_{(l_1+l_2)\times(l_1+l_2)}$$

where,

$$\mathrm{V}_{Y|X}(\gamma) = \begin{pmatrix} \frac{\partial^2 \ell_{y/x}}{\partial^2 \tau_1} & \frac{\partial^2 \ell_{y/x}}{\partial \tau_1 \partial \tau_2} & \cdots & \frac{\partial^2 \ell_{y/x}}{\partial \tau_1 \partial \tau_{l_2}} \\ \frac{\partial^2 \ell_{y/x}}{\partial \tau_1 \partial \tau_2} & \cdots & \cdots & \frac{\partial^2 \ell_{y/x}}{\partial \tau_2 \partial \tau_{l_2}} \\ \vdots & & & \vdots \\ \frac{\partial^2 \ell_{y/x}}{\partial \tau_1 \partial \tau_{l_2}} & \cdots & \cdots & \frac{\partial^2 \ell_{y/x}}{\partial^2 \tau_{l_2}} \end{pmatrix}_{l_2 \times l_2}$$

and

$$\mathrm{V}_{X|Y}(\gamma) = \begin{pmatrix} \frac{\partial^2 \ell_{x/y}}{\partial^2 \theta_1} & \cdots & \cdots & \frac{\partial^2 \ell_{x/y}}{\partial \theta_1 \partial \theta_{l_1}} \\ \frac{\partial^2 \ell_{x/y}}{\partial \theta_1 \partial \theta_2} & \cdots & \cdots & \frac{\partial^2 \ell_{x/y}}{\partial \theta_2 \partial \theta_{l_1}} \\ \vdots & & & \vdots \\ \frac{\partial^2 \ell_{x/y}}{\partial \theta_1 \partial \theta_{l_1}} & \frac{\partial^2 \ell_{x/y}}{\partial \theta_2 \partial \theta_{l_1}} & \cdots & \frac{\partial^2 \ell_{x/y}}{\partial^2 \theta_{l_1}} \end{pmatrix}_{l_1 \times l_1}$$

Therefore one can write $H(\gamma)$ as,

$$H(\gamma) = \begin{pmatrix} -E_{(Y,X)}(V_{X|Y}(\gamma)) & \underset{\sim}{0} \\ \underset{\sim}{0} & -E_{(Y,X)}(V_{Y|X}(\gamma)) \end{pmatrix} \tag{4.15}$$

Since $G(\gamma) = H(\gamma)$, Godambe matrix can be written as,

$$G(\gamma) = \begin{pmatrix} -E_{(Y,X)}(V_{X|Y}(\gamma)) & \underset{\sim}{0} \\ \underset{\sim}{0} & -E_{(Y,X)}(V_{Y|X}(\gamma)) \end{pmatrix} \tag{4.16}$$

QED

4.4.2 Godambe Information matrix for Bivariate Normal and Bernoulli Conditionals

Results of the previous section are derived for the general CEF case. Now they are applied for the Bivariate normal Bernoulli case. We will start by deriving score functions of $f(y \mid R = r) \sim N_2(\mu^{(r)}, \Sigma)$ and $f(r \mid Y = y) \sim \mathrm{Ber}(\pi(y))$ separately.

Let the full parameter space be,

$$\theta_{1\times 10} = \left(\mu_1^{(0)}, \mu_2^{(0)}, \mu_1^{(1)}, \mu_2^{(1)}, \sigma_1, \sigma_2, \rho, \alpha_0, \alpha_1, \alpha_2\right),$$

$f(y|R = r)$ distribution is,

$$f(Y|R = r, \underset{\sim}{\theta}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\Big\{ - \frac{z}{2(1-\rho^2)} \Big\}$$

where,

$$z = \frac{(y_1 - [r\mu_1^{(1)} + (1-r)\mu_1^{(0)}])^2}{\sigma_1^2} - \frac{2\rho(y_1 - [r\mu_1^{(1)} + (1-r)\mu_1^{(0)}])(y_2 - [r\mu_2^{(1)} + (1-r)\mu_2^{(0)}])}{\sigma_1\sigma_2} + \frac{(y_2 - [r\mu_2^{(1)} + (1-r)\mu_2^{(0)}])^2}{\sigma_2^2}.$$

Log likelihood function of $f(y/r)$ is

$$ln(f(\underset{\sim}{\theta}; y, r)) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma_1) - \frac{n}{2}ln(\sigma_2) - \frac{n}{4}\ln(1-\rho^2) - \frac{1}{2(1-\rho^2)}\sum_{i=1}^{n} z_i.$$

Derivatives of z with respect to full parameter space θ is given in the appendix (A.2). Thus, we can take the derivatives of $\log(y|R = r)$ with respect to the full parameter space θ using the chain rule:

$$\begin{aligned}
\frac{\partial \ell}{\partial \mu_1^{(0)}} &= \frac{-1}{2(1-\rho^2)} \times \frac{\partial z}{\partial \mu_1^{(0)}} \\
\frac{\partial \ell}{\partial \mu_1^{(1)}} &= \frac{-1}{2(1-\rho^2)} \times \frac{\partial z}{\partial \mu_1^{(1)}} \\
\frac{\partial \ell}{\partial \mu_2^{(0)}} &= \frac{-1}{2(1-\rho^2)} \times \frac{\partial z}{\partial \mu_2^{(0)}} \\
\frac{\partial \ell}{\partial \mu_2^{(1)}} &= \frac{-1}{2(1-\rho^2)} \times \frac{\partial z}{\partial \mu_2^{(1)}} \\
\frac{\partial \ell}{\partial \sigma_1} &= \frac{-1}{\sigma_1} - \frac{1}{2(1-\rho^2)} \times \frac{\partial z}{\partial \sigma_1} \\
\frac{\partial \ell}{\partial \sigma_2} &= \frac{-1}{\sigma_2} - \frac{1}{2(1-\rho^2)} \times \frac{\partial z}{\partial \sigma_2} \\
\frac{\partial \ell}{\partial \rho} &= \frac{\rho}{1-\rho^2} - \frac{1}{2(1-\rho^2)} \times \frac{\partial z}{\partial \rho}
\end{aligned}$$

Score function $U(Y|R=r_i,\theta)$ of $f(Y|R=r)$ is:

$$U(Y|R=r,\theta) = \left(\frac{\partial \ell}{\partial \mu_1^{(0)}}, \frac{\partial \ell}{\partial \mu_1^{(1)}}, \frac{\partial \ell}{\partial \mu_2^{(0)}}, \frac{\partial \ell}{\partial \mu_2^{(1)}}, \frac{\partial \ell}{\partial \sigma_1}, \frac{\partial \ell}{\partial \sigma_2}, \frac{\partial \ell}{\partial \rho}, \frac{\partial \ell}{\partial \alpha_0}, \frac{\partial \ell}{\partial \alpha_1}, \frac{\partial \ell}{\partial \alpha_2}\right)^T$$
$$= \left(\frac{\partial \ell}{\partial \mu_1^{(0)}}, \frac{\partial \ell}{\partial \mu_1^{(1)}}, \frac{\partial \ell}{\partial \mu_2^{(0)}}, \frac{\partial \ell}{\partial \mu_2^{(1)}}, \frac{\partial \ell}{\partial \sigma_1}, \frac{\partial \ell}{\partial \sigma_2}, \frac{\partial \ell}{\partial \rho}, 0, 0, 0\right)^T$$

Similarly for $f(R|Y=y)$, The distribution is $f(R|Y=y,\underset{\sim}{\theta}) = \pi(y)^r(1-\pi(y))^{1-r}$

and $\pi(y)$ can be written as,

$\pi(y) = \frac{exp(\alpha_0+\alpha_1 y_1+\alpha_2 y_2)}{1+exp(\alpha_0+\alpha_1 y_1+\alpha_2 y_2)}$ and $1-\pi(y) = \frac{1}{1+exp(\alpha_0+\alpha_1 y_1+\alpha_2 y_2)}$.

Therefore $f(r|Y=y)$ become

$$f(r|Y=y) = \frac{exp[r(\alpha_0+\alpha_1 y_1+\alpha_2 y_2)]}{1+exp(\alpha_0+\alpha_1 y_1+\alpha_2 y_2)}.$$

Log likelihood function of $f(r|Y=y)$,

$$log(f(r|Y=y,\underset{\sim}{\theta})) = \sum_{i=1}^{n} r_i(\alpha_0+\alpha_1 y_{1i}+\alpha_2 y_{2i}) + \sum_{i=1}^{n} log[1+exp(\alpha_0+\alpha_1 y_{1i}+\alpha_2 y_{2i})].$$

Derivatives are given by:

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha_0} &= \sum_{i=1}^{n} r_i - \sum_{i=1}^{n} \frac{exp(\alpha_0 + \alpha_1 y_{1i} + \alpha_2 y_{2i})}{1 + exp(\alpha_0 + \alpha_1 y_{1i} + \alpha_2 y_{2i})} \\ &= \sum_{i=1}^{n} r_i - \sum_{i=1}^{n} \frac{1}{1 + exp[-(\alpha_0 + \alpha_1 y_{1i} + \alpha_2 y_{2i})]}\end{aligned}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha_1} &= \sum_{i=1}^{n} r_i y_{1i} - \sum_{i=1}^{n} \frac{y_{1i}.exp(\alpha_0 + \alpha_1 y_{1i} + \alpha_2 y_{2i})}{1 + exp(\alpha_0 + \alpha_1 y_{1i} + \alpha_2 y_{2i})} \\ &= \sum_{i=1}^{n} r_i y_{1i} - \sum_{i=1}^{n} \frac{y_{1i}}{1 + exp[-(\alpha_0 + \alpha_1 y_{1i} + \alpha_2 y_{2i})]}\end{aligned}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha_2} &= \sum_{i=1}^{n} r_i y_{2i} - \sum_{i=1}^{n} \frac{y_{2i}.exp(\alpha_0 + \alpha_1 y_{1i} + \alpha_2 y_{2i})}{1 + exp(\alpha_0 + \alpha_1 y_{1i} + \alpha_2 y_{2i})} \\ &= \sum_{i=1}^{n} r_i y_{1i} - \sum_{i=1}^{n} \frac{y_{2i}}{1 + exp[-(\alpha_0 + \alpha_1 y_{1i} + \alpha_2 y_{2i})]}\end{aligned}$$

Then, the score function of $f(R|Y = y)$

$$U(R|Y = y, \theta) = \left(0, 0, 0, 0, 0, 0, 0, \frac{\partial \ell}{\partial \alpha_0}, \frac{\partial \ell}{\partial \alpha_1}, \frac{\partial \ell}{\partial \alpha_2}\right)^T,$$

and the score function of $f(y, r)$ is $U(\theta; y, r) = U(Y|R = r, \theta) + U(R|Y = y, \theta)$ is given by

$$U(\theta; y, r) = \left(\frac{\partial \ell}{\partial \mu_1^{(0)}}, \frac{\partial \ell}{\partial \mu_1^{(1)}}, \frac{\partial \ell}{\partial \mu_2^{(0)}}, \frac{\partial \ell}{\partial \mu_2^{(1)}}, \frac{\partial \ell}{\partial \sigma_1}, \frac{\partial \ell}{\partial \sigma_2}, \frac{\partial \ell}{\partial \rho}, \frac{\partial \ell}{\partial \alpha_0}, \frac{\partial \ell}{\partial \alpha_1}, \frac{\partial \ell}{\partial \alpha_2}\right)^T.$$

Thus from equation (4.16), Godambe Information matrix $G(\theta)$ for Bivariate Normal and Bernoulli conditional problem is,

$$G(\theta) = \begin{pmatrix} E_{(Y,R)}(V_{Y|R}(\theta))_{(7\times 7)} & \underset{\sim}{0}_{(7\times 3)} \\ \underset{\sim}{0}_{(3\times 7)} & E_{(Y,R)}(V_{R|Y}(\theta))_{(3\times 3)} \end{pmatrix}_{10\times 10}, \tag{4.17}$$

where,

$$V_{Y|R} = \begin{pmatrix} \frac{\partial^2 \ell}{\partial \mu_1^{(0)2}} & \frac{\partial^2 \ell}{\partial \mu_1^{(0)} \partial \mu_1^{(1)}} & \frac{\partial^2 \ell}{\partial \mu_1^{(0)} \partial \mu_2^{(0)}} & \frac{\partial^2 \ell}{\partial \mu_1^{(0)} \partial \mu_2^{(1)}} & \frac{\partial^2 \ell}{\partial \mu_1^{(0)} \partial \sigma_1} & \frac{\partial^2 \ell}{\partial \mu_1^{(0)} \partial \sigma_2} & \frac{\partial^2 \ell}{\partial \mu_1^{(0)} \partial \rho} \\ & \frac{\partial^2 \ell}{\partial \mu_1^{(1)2}} & & & & & \\ & & \frac{\partial^2 \ell}{\partial \mu_2^{(0)2}} & & & & \\ & & & \frac{\partial^2 \ell}{\partial \mu_2^{(1)2}} & & \underset{\sim}{0} & \\ & \underset{\sim}{0} & & & \frac{\partial^2 \ell}{\partial \sigma_1^2} & & \\ & & & & & \frac{\partial^2 \ell}{\partial \sigma_2^2} & \\ & & & & & & \frac{\partial^2 \ell}{\partial \rho^2} \end{pmatrix}$$

and

$$V_{R|Y} = \begin{pmatrix} \frac{\partial^2 \ell}{\partial \alpha_0^2} & \frac{\partial^2 \ell}{\partial \alpha_0 \partial \alpha_1} & \frac{\partial^2 \ell}{\partial \alpha_0 \partial \alpha_2} \\ & \frac{\partial^2 \ell}{\partial \alpha_1^2} & \underset{\sim}{0} \\ \underset{\sim}{0} & & \frac{\partial^2 \ell}{\partial \alpha_2^2} \end{pmatrix}.$$

Note that, $V_{R|Y}$ and $V_{Y|R}$ are triangular matrices. Thus, numerical computation of GIM is relatively easier. Further, note that, we proved that GIM for exponential family of distributions is equal to sensitivity matrix.

4.5 Comparison between ML and CL : Accuracy and Computational Cost of Estimates

In this section, we provide three simulation studies. We start by comparing computational cost of two estimation methods followed by accuracy of estimates in terms of bias and standard error and efficiency comparison based on information matrices.

4.5.1 Simulation Study 1: $\pi(y)$ with Linear Logistic Link

We carried out a simulation study for different sample sizes varying from 100 to 1000. Estimates of the model parameters were obtained using both MLE and PLE methods. Data were generated using the Gibbs algorithms. 100 such data sets were used for the study. Apart from the estimates, we also calculated the variance, bias and time. The entire analysis was performed using the R 3.5.1 software. The R package called *Rsolnp*, which performs constrained optimization, was used to obtain estimates. Results are shown in Table 4.1 and Table 4.2.

Table 4.1: Wall times of simulation study with respect to sample size and method used: TOL 10^{-3} and 100 data sets

Sample Size (n)	Wall time (in hours)	
	Maximum Likelihood Estimates	Pseudolikelihood Estimates
100	0.24876	0.03046
200	0.48388	0.05411
500	1.01297	0.12695
1000	3.00707	1.63587
10000	27.62709	17.71052

In Table 4.1, the first column shows the wall times for different sample sizes n, when the maximum likelihood method was used. The second set of columns of Table 4.1 presents the wall times for pseudolikelihood method. According to results, we can see that pseudolikelihood method is superior to maximum likelihood method in

terms of computation efficiency. For n=100, wall time of pseudolikelihood estimation for 100 iterations is 0.0304 hours (1.83 mins) while the wall time of maximum likelihood estimation for 100 iterations is 0.2488 hours ($\sim$ 15 mins). It is apparent that for small sample sizes such as $n = 100, 200$ there is no significant computational cost difference between the two methods. However, for a larger sample size, the computational advantage of PLE surpasses that of MLE quickly. Further, we note that even the PLE method shows large wall times when n increases. In that case, parallelizing the code would be more effective. We are planning to explore this in the future.

Table 4.2: Maximum likelihood estimates and pseudolikelihood estimates for different sample sizes (n): TOL 10^{-3}

Parameters	True values	Maximum likelihood		Pseudolikelihood	
		Estimate (sd)	Bias	Estimate(sd)	Bias
		n=100			
σ_0^2	1.2000	1.1986(0.0100)	-0.0014	1.1836(0.0213)	-0.0164
σ_{01}	0.6481	0.6455(0.0100)	-0.0026	0.6510(0.0135)	0.0029
σ_1^2	1.4000	1.3981(0.0100)	-0.0018	1.4273(0.0213)	0.0273
$\mu_1^{(1)}$	2.0000	2.1828(0.0173)	0.1828	1.9879(0.0176)	-0.0120
$\mu_2^{(1)}$	3.0000	3.2157(0.0141)	0.2157	3.0179(0.0253)	0.0179
$\mu_1^{(0)}$	0.0000	-0.0104(0.0223)	-0.0104	0.0053(0.0182)	0.0053
$\mu_2^{(0)}$	0.0000	-0.0213(0.0173)	-0.0213	-0.0008(0.0218)	-0.0008
α_0	0.0010	0.0756(0.0012)	0.0746	-2.3291(0.2972)	-2.3301
α_1	0.0010	0.0009(0.0000)*	-0.0001	0.5229(0.1117)	0.5219
α_2	0.0010	0.0009(0.0000)*	-0.0001	1.2877(0.1710)	1.2867
		n=200			
σ_0^2	1.2000	1.1993(0.0001)	-0.0007	1.2119(0.0085)	0.0119
σ_{01}	0.6481	0.6458(0.0007)	-0.0023	0.6498(0.0053)	0.0017
σ_1^2	1.4000	1.3990(0.0017)	-0.0009	1.4205(0.0073)	0.0205
$\mu_1^{(1)}$	2.0000	2.1824(0.0001)	0.1824	1.9785(0.0076)	-0.0215
$\mu_2^{(1)}$	3.0000	3.2145(0.0001)	0.2145	2.9683(0.0115)	-0.0317
$\mu_1^{(0)}$	0.0000	-0.0124(0.0002)	-0.0124	-0.0049(0.0073)	-0.0049
$\mu_2^{(0)}$	0.0000	-0.0219(0.0001)	-0.0219	0.0013(0.0130)	0.0013
α_0	0.0010	0.0736(0.0000)*	0.0726	-2.2607(0.1565)	-2.2617

α_1	0.0010	0.0009(0.0000)*	-0.0001	0.5220(0.0565)	0.5210
α_2	0.0010	0.0009(0.0000)*	-0.0001	1.1872(0.0994)	1.1862
		n=500			
σ_0^2	1.2000	1.1996(0.0000)*	-0.0003	1.2116(0.0009)	0.0116
σ_{01}	0.6481	0.6458(0.0008)	-0.0023	0.6434(0.0014)	-0.0046
σ_1^2	1.4000	1.3995(0.0000)*	-0.0005	1.4109(0.0010)	0.0109
$\mu_1^{(1)}$	2.0000	2.1842(0.0078)	0.1842	1.9995(0.0036)	-0.0005
$\mu_2^{(1)}$	3.0000	3.2158(0.0080)	0.2158	3.0081(0.0056)	0.0081
$\mu_1^{(0)}$	0.0000	-0.0107(0.0112)	-0.0107	0.0035(0.0045)	0.0035
$\mu_2^{(0)}$	0.0000	-0.0210(0.0069)	-0.0210	0.0069(0.0047)	0.0069
α_0	0.0010	0.0755(0.0037)	0.0745	-2.3054(0.0940)	-2.3064
α_1	0.0010	0.0009(0.0000)*	-0.0001	0.4747(0.0171)	0.4737
α_2	0.0010	0.0009(0.0000)*	-0.0001	1.1533(0.0153)	1.1523
		n=1000			
σ_0^2	1.2000	1.1998(0.0000)*	-0.0002	1.2039(0.0002)	0.0039
σ_{01}	0.6481	0.6457(0.0009)	-0.0024	0.6434(0.0009)	-0.0046
σ_1^2	1.4000	1.3997(0.0000)*	-0.0003	1.4085(0.0003)	0.0085
$\mu_1^{(1)}$	2.0000	2.1814(0.0088)	0.1814	1.9947(0.0019)	-0.0053
$\mu_2^{(1)}$	3.0000	3.2237(0.0107)	0.2237	2.9945(0.0032)	-0.0055
$\mu_1^{(0)}$	0.0000	-0.0250(0.0096)	-0.0250	-0.0042(0.0021)	-0.0042
$\mu_2^{(0)}$	0.0000	-0.0241(0.0064)	-0.0241	0.0009(0.0019)	0.0009
α_0	0.0010	0.0702(0.0035)	0.0692	-2.3886(0.0976)	-2.3896
α_1	0.0010	0.0009(0.0000)*	-0.0001	0.4790(0.0096)	0.4780
α_2	0.0010	0.0009(0.0000)*	-0.0001	1.1613(0.0094)	1.1603

$*$: very small non zero values

In Table 4.2, we present the estimates, the bias and the variances for different sample sizes. By looking at the bias values it is clear that the MLE method has less bias (and nearly zero in some cases) than the PLE method. In both methods, the variances of the estimates decrease as the sample size increases. The MLE of the α vector has less bias compared to the PLE. However, there's a noticeable departure from the true α values of PL estimates compared to ML estimates. Overall, it is evident that the MLE outperforms the PLE in terms of accuracy (less bias) and efficiency (lower variance). Therefore, we can conclude that choosing PLE over MLE is a trade off between efficiency and computational cost.

4.5.2 Simulation Study 2: $\pi(y)$ with with Quadratic Terms

To further investigate the issue behind the significant departure of the alpha estimates from the true parameter value we performed simulation studies using updated $\pi(y)$ with quadratic terms as given in equation 3.3.

$$\pi(y,\alpha) = \frac{exp(m_{10} + m_{11}y_1 + m_{12}y_2 + m_{13}y_1^2 + m_{14}y_1y_2 + m_{15}y_2^2)}{1 + exp(m_{10} + m_{11}y_1 + m_{12}y_2 + m_{13}y_1^2 + m_{14}y_1y_2 + m_{15}y_2^2)}.$$

The simulation was carried out in a similar manner as before. Results were obtain for sample sizes $n = 100, 200, 500$ and 1000. We obtained the estimates of original parameters as well as the estimates of the elements of M matrix (m space).In Tables 4.4 and 4.6, we present original parameter estimates of both ML and PL, Table 4.5 has ML estimates of m parameter space and Table 4.7 presents the estimates of m parameter space from PL method. Further, we present the Bias of the estimates and standard deviation (SD) of the estimates.

In Tables 4.4 and 4.6, we present the parameter estimates, the bias and the variances

Table 4.4: ML estimates of the original parameters

	n=100					n=200			
	Parameter	Estimate	Bias	SD		Parameter	Estimate	Bias	SD
σ_1^2	1.2000	1.2052	0.0052	0.067	σ_1^2	1.2000	1.2149	0.0149	0.073
ρ	0.6481	0.6561	0.0080	0.056	ρ	0.6481	0.6591	0.0110	0.062
σ_2^2	1.4000	1.4187	0.0187	0.146	σ_2^2	1.4000	1.4070	0.0070	0.079
$\mu_1^{(1)}$	2.0000	1.9878	-0.0122	0.083	$\mu_1^{(1)}$	2.0000	1.9935	-0.0065	0.051
$\mu_2^{(1)}$	3.0000	2.9854	-0.0146	0.125	$\mu_2^{(1)}$	3.0000	2.9886	-0.0114	0.057
$\mu_1^{(0)}$	0.0000	0.0007	0.0007	0.064	$\mu_1^{(0)}$	0.0000	-0.0055	-0.0055	0.052
$\mu_2^{(0)}$	0.0000	-0.0016	-0.0016	0.057	$\mu_2^{(0)}$	0.0000	0.0015	0.0015	0.061
α_0	0.0010	-0.0829	-0.0839	0.655	α_0	0.0010	-0.0820	-0.0830	0.534
α_1	0.0010	0.0340	0.0330	0.415	α_1	0.0010	0.0327	0.0317	0.356
α_2	0.0010	0.0679	0.0669	0.532	α_2	0.0010	0.0642	0.0632	0.464
α_3	0.0010	-0.0083	-0.0093	0.174	α_3	0.0010	0.0040	0.0030	0.129
α_4	0.0010	0.0234	0.0224	0.226	α_4	0.0010	0.0221	0.0211	0.169
α_5	0.0010	0.0199	0.0189	0.314	α_5	0.0010	0.0119	0.0109	0.117

	n=500					n=1000			
	Parameter	Estimate	Bias	SD		Parameter	Estimate	Bias	SD
σ_1^2	1.2000	1.2126	0.0126	0.062	σ_1^2	1.2000	1.2034	0.0034	0.064
ρ	0.6481	0.6560	0.0079	0.055	ρ	0.6481	0.6583	0.0103	0.053
σ_2^2	1.4000	1.4135	0.0135	0.077	σ_2^2	1.4000	1.4133	0.0133	0.063
$\mu_1^{(1)}$	2.0000	1.9998	-0.0002	0.056	$\mu_1^{(1)}$	2.0000	2.0000	-0.0000	0.050
$\mu_2^{(1)}$	3.0000	3.0083	0.0083	0.053	$\mu_2^{(1)}$	3.0000	3.0023	0.0023	0.055
$\mu_1^{(0)}$	0.0000	0.0006	0.0006	0.052	$\mu_1^{(0)}$	0.0000	-0.0071	-0.0071	0.047
$\mu_2^{(0)}$	0.0000	-0.0067	-0.0067	0.048	$\mu_2^{(0)}$	0.0000	0.0001	0.0001	0.055
α_0	0.0010	-0.0856	-0.0866	0.502	α_0	0.0010	-0.0809	-0.0819	0.455
α_1	0.0010	0.0328	0.0318	0.325	α_1	0.0010	0.0345	0.0335	0.357
α_2	0.0010	0.0648	0.0638	0.375	α_2	0.0010	0.0589	0.0579	0.380
α_3	0.0010	0.0191	0.0181	0.099	α_3	0.0010	0.0047	0.0037	0.098
α_4	0.0010	-0.0092	-0.0102	0.176	α_4	0.0010	0.0201	0.0191	0.181
α_5	0.0010	0.0185	0.0175	0.110	α_5	0.0010	0.0165	0.0155	0.125

Table 4.5: ML estimates of the m parameters

	n=100					n=200			
	Parameters	Estimate	Bias	SD		Parameters	Estimate	Bias	SD
m00	1.5569	4.7517	3.1948	2.4945	m00	1.5569	6.7630	5.2061	4.6195
m01	0.0000	0.0030	0.0030	0.060	m01	0.0000	-0.0063	-0.0063	0.047
m02	0.0000	-0.0032	-0.0032	0.046	m02	0.0000	0.0040	0.0040	0.048
m03	-0.5556	-0.5573	-0.0018	0.031	m03	-0.5556	-0.5542	0.0014	0.029
m04	0.5143	0.5174	0.0031	0.050	m04	0.5143	0.5186	0.0043	0.041
m05	-0.4762	-0.4754	0.0007	0.036	m05	-0.4762	-0.4783	-0.0021	0.025
m10	0.0010	-0.0829	-0.0839	0.655	m10	0.0010	-0.0820	-0.0830	0.534
m11	0.0010	0.0340	0.0330	0.415	m11	0.0010	0.0327	0.0317	0.356
m12	0.0010	0.0679	0.0669	0.532	m12	0.0010	0.0642	0.0632	0.464
m13	0.0010	-0.0083	-0.0093	0.174	m13	0.0010	0.0040	0.0030	0.129
m14	0.0010	0.0234	0.0224	0.226	m14	0.0010	0.0221	0.0211	0.169
m15	0.0010	0.0199	0.0189	0.314	m15	0.0010	0.0119	0.0109	0.117

	n=500					n=1000			
	Parameters	Estimate	Bias	SD		Parameters	Estimate	Bias	SD
m00	1.5569	1.5307	-0.0262	0.069	m00	1.5569	1.5173	-0.0396	0.073
m01	0.0000	0.0039	0.0039	0.050	m01	0.0000	-0.0079	-0.0079	0.045
m02	0.0000	-0.0068	-0.0068	0.041	m02	0.0000	0.0038	0.0038	0.045
m03	-0.5556	-0.5526	0.0029	0.028	m03	-0.5556	-0.5599	-0.0044	0.030
m04	0.5143	0.5128	-0.0015	0.040	m04	0.5143	0.5215	0.0071	0.043
m05	-0.4762	-0.4741	0.0021	0.023	m05	-0.4762	-0.4763	-0.0002	0.022
m10	0.0010	-0.0856	-0.0866	0.502	m10	0.0010	-0.0809	-0.0819	0.455
m11	0.0010	0.0328	0.0318	0.325	m11	0.0010	0.0345	0.0335	0.357
m12	0.0010	0.0648	0.0638	0.375	m12	0.0010	0.0589	0.0579	0.380
m13	0.0010	0.0191	0.0181	0.099	m13	0.0010	0.0047	0.0037	0.098
m14	0.0010	-0.0092	-0.0102	0.176	m14	0.0010	0.0201	0.0191	0.181
m15	0.0010	0.0185	0.0175	0.110	m15	0.0010	0.0165	0.0155	0.125

Table 4.6: PL estimates of the original parameters

	n=100					n=200			
	Parameters	Estimate	Bias	SD		Parameters	Estimate	Bias	SD
σ_0^2	1.2000	1.1615	-0.0385	0.0381	σ_0^2	1.2000	1.2053	-0.0053	0.0359
σ_{01}	0.6481	0.6727	0.0246	0.0231	σ_{01}	0.6481	0.6835	-0.0354	0.0144
σ_1^2	1.4000	1.4716	0.0716	0.0497	σ_1^2	1.4000	1.4266	-0.0266	0.0272
$\mu_1^{(1)}$	2.0000	2.0458	0.0458	0.0151	$\mu_1^{(1)}$	2.0000	1.9564	0.0436	0.0073
$\mu_2^{(1)}$	3.0000	3.0263	0.0263	0.0437	$\mu_2^{(1)}$	3.0000	2.9752	0.0248	0.0192
$\mu_1^{(0)}$	0.0000	0.0096	0.0096	0.0343	$\mu_1^{(0)}$	0.0000	0.0092	-0.0092	0.0237
$\mu_2^{(0)}$	0.0000	0.0561	0.0561	0.0182	$\mu_2^{(0)}$	0.0000	0.0562	-0.0562	0.0331
α_0	0.0010	0.02132	0.0203	0.5742	α_0	0.0010	0.0210	-0.0200	0.3773
α_1	0.0010	0.6859	0.6849	1.5424	α_1	0.0010	0.6247	-0.6237	1.1606
α_2	0.0010	0.3051	0.3041	0.1467	α_2	0.0010	1.8235	-1.8225	0.0027
α_3	0.0010	0.0906	0.0896	0.9671	α_3	0.0010	0.0369	-0.0359	0.0835
α_4	0.0010	0.0018	0.0008	0.1787	α_4	0.0010	0.0015	-0.0005	0.1759
α_5	0.0010	0.0069	0.0059	0.1752	α_5	0.0010	0.0061	-0.0051	0.1416

	n=500					n=1000			
	Parameters	Estimate	Bias	SD		Parameters	Estimate	Bias	SD
σ_0^2	1.2000	1.2021	-0.0021	0.0335	σ_0^2	1.2000	1.2060	-0.0060	0.0150
σ_{01}	0.6481	0.6489	-0.0008	0.0172	σ_{01}	0.6481	0.6478	0.0003	0.0063
σ_1^2	1.4000	1.4032	-0.0032	0.0212	σ_1^2	1.4000	1.3985	0.0015	0.0008
$\mu_1^{(1)}$	2.0000	2.0027	-0.0027	0.0069	$\mu_1^{(1)}$	2.0000	1.9955	0.0045	0.0022
$\mu_2^{(1)}$	3.0000	3.0009	-0.0009	0.0150	$\mu_2^{(1)}$	3.0000	3.0110	-0.0110	0.0132
$\mu_1^{(0)}$	0.0000	0.0379	-0.0379	0.0291	$\mu_1^{(0)}$	0.0000	-0.0071	0.0071	0.0238
$\mu_2^{(0)}$	0.0000	0.0002	-0.0002	0.0172	$\mu_2^{(0)}$	0.0000	-0.0079	0.0079	0.0073
α_0	0.0010	0.0255	-0.0245	0.4014	α_0	0.0010	0.0158	-0.0148	0.4008
α_1	0.0010	0.6550	-0.6540	1.3633	α_1	0.0010	0.7029	-0.7019	0.1587
α_2	0.0010	1.8449	1.8439	0.0669	α_2	0.0010	1.8700	-1.8690	0.0063
α_3	0.0010	-0.0745	0.0755	0.0669	α_3	0.0010	-0.0184	0.0194	0.0703
α_4	0.0010	0.0504	-0.0494	0.1713	α_4	0.0010	0.0071	-0.0061	0.1639
α_5	0.0010	0.0037	-0.0027	0.0874	α_5	0.0010	0.0064	-0.0054	0.0099

Table 4.7: PL estimates of the m parameters

	n=100					n=200			
	Parameters	Estimate	Bias	SD		Parameters	Estimate	Bias	SD
m00	1.5569	2.6732	1.1163	0.6959	m00	1.5569	2.7615	1.2046	0.9926
m01	0.0000	0.0105	0.0105	0.054	m01	0.0000	-0.0111	-0.0111	0.054
m02	0.0000	-0.0047	-0.0047	0.048	m02	0.0000	0.0026	0.0026	0.049
m03	-0.5556	-0.5600	-0.0045	0.027	m03	-0.5556	-0.5582	-0.0026	0.028
m04	0.5143	0.5141	-0.0003	0.046	m04	0.5143	0.5166	0.0022	0.045
m05	-0.4762	-0.4750	0.0012	0.022	m05	-0.4762	-0.4797	-0.0035	0.024
m00	0.0010	0.02132	0.0203	0.574	m00	0.0010	0.0210	-0.0200	0.377
m01	0.0010	0.6859	0.6849	0.1542	m01	0.0010	0.6247	-0.6237	0.1161
m02	0.0010	0.3051	0.3041	0.147	m02	0.0010	1.8235	-1.8225	0.003
m03	0.0010	0.0906	0.0896	0.967	m03	0.0010	0.0369	-0.0359	0.083
m04	0.0010	0.0018	0.0008	0.179	m04	0.0010	0.0015	-0.0005	0.176
m05	0.0010	0.0069	0.0059	0.175	m05	0.0010	0.0061	-0.0051	0.142
	n=500					n=1000			
	Parameters	Estimate	Bias	SD		Parameters	Estimate	Bias	SD
m00	1.5569	3.9067	2.3498	0.5068	m00	1.5569	1.5231	-0.0338	0.077
m01	0.0000	0.0424	0.0424	0.050	m01	0.0000	-0.0038	-0.0038	0.049
m02	0.0000	-0.0196	-0.0196	0.061	m02	0.0000	-0.0038	-0.0038	0.045
m03	-0.5556	-0.5561	-0.0005	0.024	m03	-0.5556	-0.5537	0.0019	0.025
m04	0.5143	0.5144	0.0001	0.041	m04	0.5143	0.5131	-0.0013	0.039
m05	-0.4762	-0.4770	-0.0008	0.025	m05	-0.4762	-0.4775	-0.0013	0.022
m00	0.0010	0.0255	-0.0245	0.401	m00	0.0010	0.0158	-0.0148	0.401
m01	0.0010	0.6550	-0.6540	0.136	m01	0.0010	0.7029	-0.7019	0.159
m02	0.0010	1.8449	1.8439	0.067	m02	0.0010	1.8700	-1.8690	0.006
m03	0.0010	-0.0745	0.0755	0.067	m03	0.0010	-0.0184	0.0194	0.070
m04	0.0010	0.0504	-0.0494	0.171	m04	0.0010	0.0071	-0.0061	0.164
m05	0.0010	0.0037	-0.0027	0.087	m05	0.0010	0.0064	-0.0054	0.010

for different sample sizes for ML and PL methods. Similarly as before by looking at the bias values it is clear that the ML method has less bias than the PL method. In both methods, the variances of the estimates decrease as the sample size increases.

According to Table 4.4 and 4.6, the ML method estimates the α vector accurately compared to PL method. Further, in-terms of m parameter estimates, ML estimation is still performing superior compared to PL. By looking at Table 4.6 it is evident that α vector estimates coming from the PL method has significantly improved when we incorporate the quadratic terms into the logistic link. Yet, in PL method, bias of α_1 and α_2 are still significant. The same issue exist in PL estimates in m space also. The reason for this issue may be related to uniqueness of the distribution which we will address in our future work after further investigation.

4.5.3 Efficiency Comparison

In order to compare the efficiencies of $\hat{\theta}_{ML}$ and $\hat{\theta}_{CL}$ we used the concept of joint asymptotic relative efficiency (JARE), as defined in (4.1). We increased the sample size from 100 to 1000 and observed the JARE between $\hat{\theta}_{ML}$ and $\hat{\theta}_{CL}$. Results are given in Table 4.8.

Table 4.8: Joint Asymptotic Relative Efficiency between $\hat{\theta}_{ML}$ and $\hat{\theta}_{CL}$

Sample size (n)	JARE
100	4.82×10^{-10}
200	1.06×10^{-6}
500	8.46×10^{-4}
1000	8.21×10^{-3}

According to Table 4.8, when sample size n increases JARE increases. That is, the ratio between estimates coming from ML and estimates coming from CL is approaching 1 when sample size increases. By looking at the results we can state that in terms of efficiency MLE is superior than CLE.

Chapter 5: Bivariate Normal and Bernoulli Distribution Illustrative Example

In this chapter, we will present an illustrative application. We consider stock price data along with analyst recommendations; usually buy, sell or hold. Clearly, the recommendations are discrete/categorical and stock price in a continuous response variable. This provides an opportunity to illustrate the of use of discrete-continuous joint distribution specified via the conditionals.

As background, it is sufficient to know that analysts recommendations are based on the study of the behavior of the stock price, economic indicators, and any other relevant information they can obtain. Then it makes sense to specify a model (such as logistic regression) for the recommendation conditional on the past stock prices. On the other hand, it is reasonable to assume that the price of a stock which is classified as a buy is likely to behave different that its price defining the time it is listed as a sell. Hence, the conditional distribution of the stock price given its recommendation states may also be formulated. We shall then ask if the specifications are compatible and the joint distribution can be estimated.

5.1 Description of the Data Set

Daily closing prices and recommendations were collected for 100 stocks (the list of 100 stock can be found in Appendix) using the website *www.Barchart.com*. The data were collected from July 2019 to August 2020; that is roughly 60 weeks of data. However, we restricted ourselves to the period from July 2019 to February 2020 prior to the market drop due to Covid-19. For the descriptive analysis in section 5.3, we used both the entire data set and partial data set. Only the partial data set (data from July 2019 to February 2020) were used for the advance analysis. For each stock, we have closing price and recommendation for each business day for 60 weeks in full data set (data from July 2019 to August 2020) and the partial data contains only 32 weeks worth of data.

5.2 Data Analysis

The data analysis included an exploratory graphical and descriptive analysis of closing prices and recommendation for each stock. Time series plots, Surface plots and contour plots were used to study the distribution of data.

The derived model assumes a bivariate normal conditional for the continuous part (Y) of data. Hence, we only considered the two closing prices from Monday and Friday of each week and the recommendation from Monday of the following week to match the conditionals of the joint model. That is, conditional distributions of closing prices of Monday (Y_0) and Friday (Y_1) of each week given the end of the week recommendation (R), is assumed to follow a bivariate normal distribution and the end of the week recommendation given the closing prices of Monday and Friday is assumed to follow a standard logistic regression model. By letting $Y = (Y_0, Y_1)^T$, the bivariate vector start and end of the week

prices, and R denote the end of the week analyst recommendation where $R = 0$ is "sell" and $R = 1$ is "Buy"; the specification of conditional distribution is

$$f(\underset{\sim}{y} \mid R = 0) = N_2(\underset{\sim}{\mu}^{(0)}, \Sigma)$$

$$f(\underset{\sim}{y} \mid R = 1) = N_2(\underset{\sim}{\mu}^{(1)}, \Sigma)$$

and $f(R = r \mid y) = Ber[\pi(y, \alpha)]$ where $\pi(y, \alpha)$ is set as in equation 3.1.

Price data for some of the stocks are not consistent with normality assumption. Data for five representative stocks are shown in figure (5.1) and (5.2). Data are divided into two graphs according to the value of R. Then, each graph corresponding to a conditional distribution. We apply Box-Cox transformation to each stock price separately to improve the normality fit. The pseudolikelihood and the maximum likelihood methods are used to estimate the parameters for each stocks.

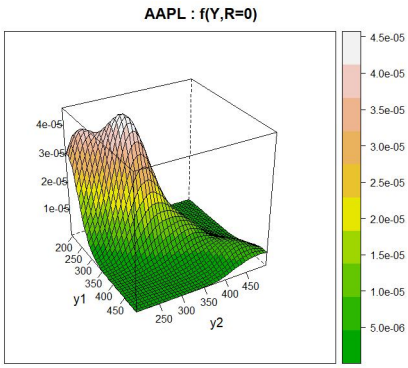
As shown in chapter 3, there exists a joint distribution consistent with the above conditional specification. The corresponding likelihood function is given by equation (4.7)

$$L(f(y, r; \hat{\underset{\sim}{\theta}})) = \prod_{i=1}^{n} \frac{exp(rm_{10} + Q(y, r))}{\sum_{r=0}^{1} exp(rm_{10}) \int_{y1=-\infty}^{\infty} \int_{y2=-\infty}^{\infty} exp(Q(y, r)) dy_2 dy_1}$$

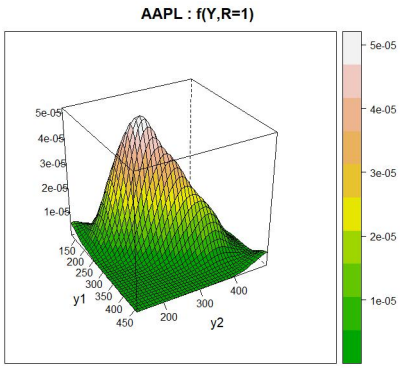
In general, such joint likelihoods are not tractable. But in our case the normality constant does yield an analytical expression given in equation (3.4). Then, the ML estimate can be computed, but it is computationally demanding. In this example, we will also compute the pseudolikelihood estimates for comparison purposes. MLE and PLE are given in Table (5.1). Note that the ML and PL estimates of the conditional distribution of $Y \mid R$ are nearly identical. However, the two methods give very different results of the parameters of $R \mid Y$ in most cases.

Table (5.1) present summary statistics of closing prices of five representative stocks namely Apple Inc. stock (AAPL), Biomerica Inc. (BMRA) stock, PepsiCo, Inc. (PEP)

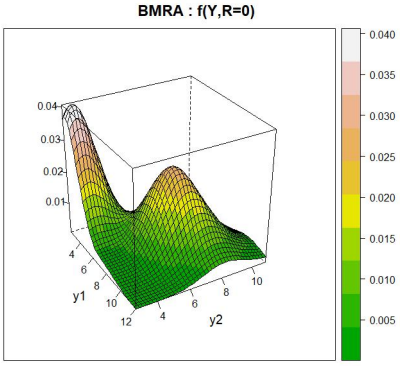
stock, Automatic Data Processing Inc. (ADP) stock and Mid-America Apartment Communities (MAA) stock. Statistics are separated with respective to the recommendation (R). According to the Table (5.1), we can see that Monday closing price and Friday closing price have strong correlations. Further, variance covariance matrix of closing prices when $R = 0$ and $R = 1$ are similar. Thus, we can assume the compatibility requirement is satisfied for these selected stocks.



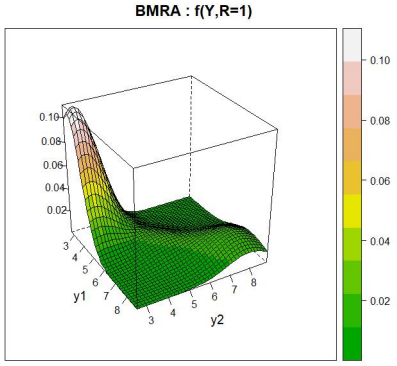
(a)



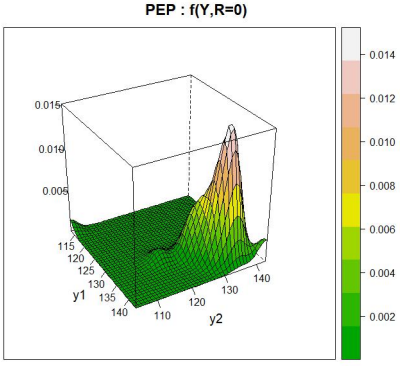
(b)



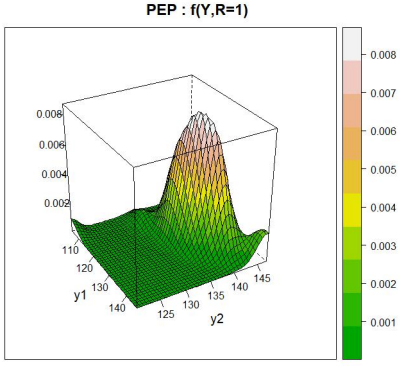
(c)



(d)

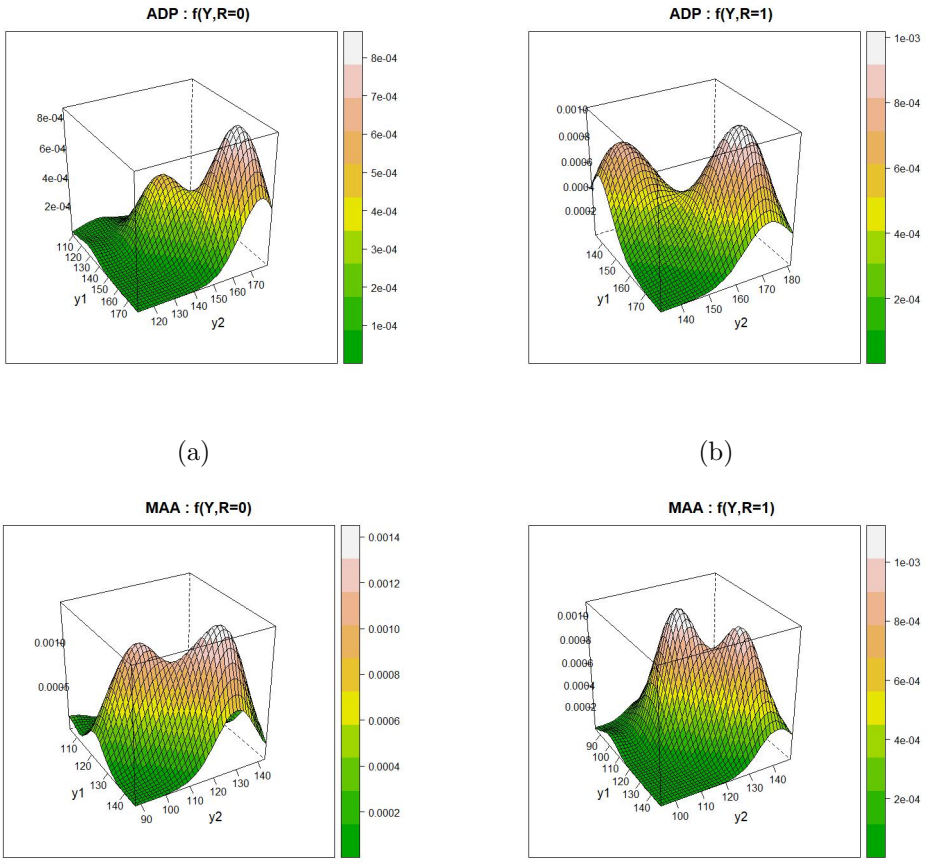


(e)



(f)

Figure 5.1: $f(y, r = 0)$ surface plot for five stocks. (a) $f(y, r = 0)$ of Apple Inc. stock, (b) $f(y, r = 1)$ of Apple Inc. stock, (c) $f(y, r = 0)$ of Biomerica Inc. stock, (d) $f(y, r = 1)$ of Biomerica Inc. stock, (e) $f(y, r = 0)$ of PepsiCo, Inc. stock, (f) $f(y, r = 1)$ of PepsiCo, Inc. stock.



(a) (b)

(c) (d)

Figure 5.2: $f(y, r = 1)$ surface plot for five stocks. (a) $f(y, r = 0)$ of Automatic Data Processing Inc. stock, (b) $f(y, r = 1)$ of Automatic Data Processing Inc. stock, (c) $f(y, r = 0)$ of Mid-America Apartment Communities stock, (d) $f(y, r = 1)$ of Mid-America Apartment Communities stock.

Table 5.1: Summary Statistics for Closing Prices

Summary Statistics	AAPL		ADP		BMRA		MAA		PEP	
(Closing Prices)	MON	FRI	MON	FRI	MON	FRI	MON	FRI	MON	FRI
$R = 0$										
Minimum	193.34	201	158	159	2.87	2.7	121	119	124	129
Mean ($\bar{x}$)	249.91	245.58	167.54	166.83	3.00	3.03	130.12	131.13	135.32	135.36
Median (Q_2)	218.82	219.90	167.09	168.05	3.01	3.01	130.50	132.68	135.63	136.12
Standard Deviation (s)	44.72	45.60	5.37	5.74	0.09	0.18	5.71	5.58	3.75	3.73
ρ	0.99		0.74		0.52		0.93		0.86	
Maximum	309.00	320.03	175.74	179.10	3.18	3.47	139.15	137.21	142.14	142.91
Q_1	210.012	208.74	162.25	163.72	2.98	2.85	127.19	126.82	134.51	133.28
Q_3	281.93	289.80	170.19	171.01	3.09	3.07	135.82	134.59	136.69	136.64
$R = 1$										
Minimum	200	3.03	3.03	159	2.75	2.68	118	119	130	128
Mean ($\bar{x}$)	251.46	249.12	168.78	167.11	3.02	3.06	131.96	129.38	137.17	136.01
Median (Q_2)	251.20	244.78	169.56	167.14	3.00	3.07	132.35	130.55	137.37	136.53
Standard Deviation (s)	41.72	45.69	5.36	5.84	0.19	0.21	5.26	5.19	3.51	4.09
ρ	1		0.96		0.66		0.97		0.93	
Maximum	316.96	318.73	179.36	181.25	3.41	3.46	142.34	145.51	145.66	146.99
Q_1	226.29	229.31	162.48	163.99	2.92	2.89	125.40	128.82	133.74	133.94
Q_3	264.85	269.47	169.07	171.01	3.20	3.17	133.30	136.96	137.10	139.11

In empirical plots based on smoothing the data in figures (5.1) and (5.2), we present the joint distributions, ($f(y; r = 0)$ and $f(y; r = 1)$) of five representative stocks. The reader can refer to appendix to study the rest of 95 stocks. According to figures, it is clear that most of the distributions are bimodal.

5.3 Estimation using Weekly Data

As the first step, we follow the data structure given in (1.4) in example 2 in chapter 1. That is, for a certain week Monday price (y_{MP}) and Friday price (y_{FP}) assumed to follow a bivariate normal distribution. Further, "Buy" or "Sell" recommendation (R) follows a Bernoulli distribution with standard logistic link.

$$(y_{MP}, y_{FP}) \sim f(\underset{\sim}{y} \mid R = r) \sim N_2(\underset{\sim}{\mu}^{(r)}, \Sigma)$$

$$R \sim f_{R|Y}(r|Y = y) = \pi(y, \alpha)^r (1 - \pi(y, \alpha))^{1-r}$$

where,

$$\text{logit}[\pi(y, \alpha)] = log\left(\frac{\pi(y, \alpha)}{1 - \pi(y, \alpha)}\right) = \alpha_0 + \alpha_1 y_{MP} + \alpha_2 y_{FP}$$

We estimate the parameters of each stock using Maximum likelihood estimation and Pseudo likelihood estimation. We only present the results for five representative stocks in Table (5.2). The reader can find the rest of the results are in the Appendix if interested. In Table (5.2), $\mu^{(0)}$ is the mean vector when stock is a "sell" and $\mu^{(1)}$ is the mean vector when stock is a "buy". The mean vectors, variances, covariances and alpha coefficients are presented in Table (5.2). Note that in Table (5.2) the means, variances and correlation estimates obtained by the two methods are very close to each other. However, the estimates of the parameters of the logistic regression model are not so close. The other important point to note is that the α_2 parameter seems to be significantly different from zero. This implies that the product of the Monday and Friday prices is significant in the logistic regression model. This pattern can be observed in the rest of the stocks too.

Table 5.2: Weekly Analysis: Maximum likelihood and pseudolikelihood estimates for AAPL, PEP, MAA, BMRA and ADP stock prices

Stock	Method	$\mu^{(0)}$	$\mu^{(1)}$	σ_1^2	σ_2^2	$\rho\sigma_1\sigma_2$	α_0	α_1	α_2
AAPL	PLE	(289.49,289.71)	(278.25,283.62)	5276.43	5580.58	5387.98	0.53	56.08	64.83
	MLE	(288.98,290.12)	(278.45,283.52)	5276.44	5580.60	5387.95	3.59	462.05	478.19
PEP	PLE	(134.70,133.64)	(134.20,135.33)	39.96	39.52	33.89	0.10	-0.10	0.10
	MLE	(134.64,133.70)	(133.71,134.96)	48.04	33.54	34.61	0.11	0.51	1.42
MAA	PLE	(125.91,123.95)	(122.23,123.25)	146.07	150.43	132.57	0.09	-0.09	0.09
	MLE	(134.52,133.38)	(131.22,124.20)	114.42	163.76	132.80	-6.89	-1012.75	-889.68
BMRA	PLE	(5.41,5.42)	(4.15,4.13)	6.24	5.79	5.65	0.10	-0.10	0.1
	MLE	(5.41,5.42)	(4.15,4.13)	6.24	5.79	5.65	0.10	-0.10	0.1
ADP	PLE	(157.29,156.11)	(156.16,156.47)	227.52	236.93	214.55	0.10	-0.09	0.09
	MLE	(156.09,155.97)	(155.22,156.02)	227.43	236.80	214.81	0.41	37.33	49.06

5.4 Parameter Estimation of Lag Distribution

We extended our analysis to the data structure with lag prices. To further explain the data set, we take consider the Friday price (y_{FP}) before the weekend and Monday price (y_{MP}) after the same weekend follows a bivariate normal distribution. Further, "Buy" or "Sell" recommendation (R) on Monday after the weekend follows a Bernoulli distribution with standard logistic link as before. We estimate the parameters using ML estimation and PL estimation of the same five representative stocks as same in the previous analysis. Results are given in Table (5.4).

Table 5.3: Pseudolikelihood estimates for AAPL, PEP, MAA, BMRA and ADP stock prices.

Stock	$\mu^{(0)}$	$\mu^{(1)}$	Σ	α_0	α_1	α_2	α_3	α_4
AAPL	(246.73,249.03,245.85,245.52)	(244.02,248.02,228.94, 241.52)	$\begin{pmatrix} 1394.035 & 1470.437 & 1285.463 & 1355.924 \\ 1470.437 & 1569.651 & 1370.025 & 1441.065 \\ 1285.463 & 1370.025 & 1223.746 & 1269.654 \\ 1355.924 & 1441.065 & 1269.654 & 1344.374 \end{pmatrix}$	1.18	285.88	294.14	259.56	280.98
ADP	(167.87,167.81,177.03,175.42)	(167.08,168.76,165.86,166.95)	$\begin{pmatrix} 28.66 & 28.48 & 18.96 & 22.75 \\ 28.48 & 34.74 & 22.99 & 24.96 \\ 18.96 & 22.99 & 23.24 & 18.92 \\ 22.75 & 24.96 & 18.92 & 25.05072 \end{pmatrix}$	0.12	3.52	3.69	3.18	3.47
BMRA	(3.03, 3.00, 3.03, 2.96)	(3.06, 3.02,8.15,7.63)	$\begin{pmatrix} 0.03 & 0.02 & 0.01 & 0.03 \\ 0.02 & 0.04 & 0.01 & 0.02 \\ 0.01 & 0.01 & 0.03 & 0.02 \\ 0.03 & 0.02 & 0.02 & 0.04 \end{pmatrix}$	0.10	-0.09	0.10	0.10	-0.09
MAA	(131.13,130.12,140.07,139.75)	(129.38 ,131.96,129.63,129.87)	$\begin{pmatrix} 42.46 & 40.53 & 2.06 & 0.48 \\ 40.53 & 44.93 & 12.69 & 12.22 \\ 2.06 & 12.69 & 581.01 & 603.20 \\ 0.48 & 12.22 & 603.20 & 632.00 \end{pmatrix}$	0.10	-0.10	0.10	0.10	-0.10
PEP	(135.36,135.32,128.97,130.85)	(136.00 ,137.17,135.03,135.36)	$\begin{pmatrix} 17.23 & 16.78 & 13.06 & 16.09 \\ 16.78 & 20.70 & 13.19 & 14.99 \\ 13.06 & 13.19 & 14.63 & 13.83 \\ 16.09 & 14.99 & 13.83 & 16.93 \end{pmatrix}$	0.10	-0.10	0.10	0.10	-0.10

5.5 Estimation of Biweekly Data

To further, investigate behaviour of the estimates and predictions we used consider Monday price and Friday price of two weeks and the recommendation at the end of that two weeks. That is we have $Y = (Y_1, Y_2, Y_3, Y_4)^T$ where y_1 = Monday price of week one, y_2 = Monday price of week one, y_3 = Monday price of week two, and y_4 = Monday price of week two and recommendation r at the end of the second week. Thus, we have $f(y \mid R = r) \sim N_4(\mu^{(r)}, \Sigma)$ and $f(r/Y = y) \sim Ber(\pi(y), \alpha)$; where $\text{logit}[\pi(y)] = \alpha_0 +$

$\alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_3 + \alpha_4 y_4$. Note that we are still considering the linear logistic link.

Deriving the joint distribution for the above two conditionals is quite cumbersome. Therefore, we are not presenting the ML estimates for this particular case. However, PL estimation is still possible since we use product of two conditionals as the likelihood instead of the full likelihood. Table 5.4 presents the PL estimates.

Table 5.4: Pseudolikelihood estimates for AAPL, PEP, MAA, BMRA and ADP stock prices.

Stock	$\mu^{(0)}$	$\mu^{(1)}$	Σ	α_0	α_1	α_2	α_3	α_4
AAPL	(244.71,249.64,251.11,256.48)	(243.22,243.77,244.62,244.57)	$\begin{pmatrix} 1410.17 & 1472.81 & 1437.76 & 1436.04 \\ 1472.81 & 1557.10 & 1526.09 & 1528.32 \\ 1437.76 & 1526.09 & 1508.28 & 1508.88 \\ 1436.04 & 1528.32 & 1508.89 & 1528.35 \end{pmatrix}$	0.10	-0.10	0.10	0.10	-0.10
ADP	(166.98, 167.98, 167.77, 170.14)	(168.10,168.74,167.38,165.51)	$\begin{pmatrix} 27.58 & 26.32 & 24.94 & 18.84 \\ 26.32 & 32.72 & 30.58 & 24.10 \\ 24.94 & 30.58 & 30.84 & 27.51 \\ 18.84 & 24.10 & 27.51 & 30.52 \end{pmatrix}$	-65.64	-1.06	218.32	-526.55	306.80
BMRA	(3.00,2.98,3.00,2.98)	(3.04,3.12,3.12,3.12)	$\begin{pmatrix} 0.03 & 0.02 & 0.01 & 0.01 \\ 0.02 & 0.03 & 0.02 & 0.02 \\ 0.01 & 0.02 & 0.02 & 0.02 \\ 0.01 & 0.02 & 0.02 & 0.04 \end{pmatrix}$	0.10	-0.10	0.10	0.10	-0.10
MAA	(128.06,129.43,130.05,131.07)	(131.24,132.15,132.13,131.07)	$\begin{pmatrix} 38.16 & 37.19 & 36.38 & 35.15 \\ 37.19 & 42.34 & 41.66 & 40.19 \\ 36.38 & 41.66 & 41.75 & 40.21 \\ 35.15 & 40.19 & 40.21 & 41.70 \end{pmatrix}$	8.01	-0.03	-0.83	-0.13	0.93
PEP	(133.84,134.70,135.50,136.75)	(137.07,137.41,136.82,135.90)	$\begin{pmatrix} 18.79 & 18.16 & 16.48 & 18.14 \\ 18.16 & 20.33 & 17.49 & 18.48 \\ 16.48 & 17.49 & 15.97 & 17.62 \\ 18.14 & 18.48 & 17.62 & 22.80 \end{pmatrix}$	112.84	-2.69	0.20	0.15	1.51

5.6 Discussion

In this illustrative example, closing prices with recommendation for 100 stocks collected from July 2019 to August 2020 which is roughly 60 weeks, were analyzed. The main goal of the study is to compare the parameter estimates obtained from ML estimation method and PL estimation method for each stock. Since the joint distribution we derived in chapter 3 has bivariate normal distribution and the standard logistic regression distribution as the two compatible conditionals; for each stock, we assumed the two closing prices from Monday and Friday of each week given the recommendation from Monday of the following week follows a bivariate normal distribution and the end of the week recommendation given the closing prices of Monday and Friday is assumed to follow a standard logistic regression model. Box-Cox transformation was applied to continuous data when necessary to maintain the normality assumption. First the joint distributions of each stock were observed. Further, estimated the parameters using ML and PL methods.

According to the results, for the five representative stocks, the means, variances and correlation estimates obtained by ML method and PL method are very close to each other. However, the estimates of the parameters of the logistic regression model are not so close. The other important point to note is that the α_2 parameter seems to be significantly different from zero. This implies that the product of the Monday and Friday prices is significant in the logistic regression model. This pattern can be observed in the rest of the stocks we chosen also. This computational problem might be a negative characteristic of the composite likelihood (or pseudo likelihood) method. This means that using maximum likelihood estimation is more reliable method to analyze the data. However, due to the complexity of the full model and restriction upon the conditionals it

is not always convenient to analyze the data if the data does not have a specific structure. Composite likelihood is a good choice since the data does not need a specific structure.

Chapter 6: Conclusion

A joint model is often needed in many real-world applications. However, selecting an appropriate model that explains the real-life data observed could be difficult. One can choose a model from well-known parametric families of distributions which approximately explain the observed data. Often the joint model which explains the real observed data is either intractable or not in a closed-form solution. This will lead to many other issues such as interpretation of the model, estimating parameters, inference, and data generation. This is particularly the case where as the data consist of both continuous and discrete parts. Further, even after choosing a model that explains data adequately, visualizing the distribution is often difficult. Often in practice, conditionals are easier to model and interpret while the joint distribution itself is either intractable or not available in closed form. Especially, when the observed response consists of both continuous and discrete components, specifying conditionals is more convenient. If one wants to derive the joint model based on conditionals, one can use the concept called conditional specification. Such models are referred to as conditionally specified models. These conditionally specified models are intuitively appealing, and knowing the conditional distributions makes the problem easier to understand and visualize. In this thesis, our main focus was to derive a joint distribution to describe both the continuous response variables and the discrete variables and to explore the theoretical aspects of conditionally specified models including

parameter estimation and data generation using the joint distributions.

We started the study by presenting several practical motivating examples such as self-proxy data in gerontology studies, synthetic data, and stock market closing prices and recommendations data where the data have a discrete component and a continuous component and can be specified as conditionals. To analyze these data in a single framework one can use the approach we are proposing. To derive the joint we used the stock market closing price and the "Buy"/"Sell" recommendation example where the beginning of the week and end of the week price data given the recommendation follows a bivariate normal distribution and the recommendation given the price data follows a logistic regression model. The conditionally specified joint model was derived by Arnold et al. (2001). The compatibility of the conditional distributions has been verified using standard theorems. The conditionals resulting from this joint distribution are more general than those we started with and hence the approach provides a general class of models for analyzing the stock price and recommendation data. We were able to obtain all the elements of M matrix in terms of true parameters of conditionals as well as a closed-form solution to the normalizing constant. Further, although the proposed joint model has a closed-form expression, it is very complex and has a messy normalizing constant. Therefore, generating data directly from the joint model is immensely difficult and may even not be feasible. However, since the model is conditionally specified we were able to apply other numerical algorithms such as Gibb's Sampling.

Because of awkward normalizing constant differentiating the likelihood and deriving the maximum likelihood equations became challenging. While in practice ML estimation is the preferred estimation method for parametric models, in our case it comes with a heavy computational burden. Therefore, it behooves us to explore other methods such as

composite likelihood estimation. Due to the difficulty of obtaining closed-form expressions for estimates, using our proposed density is not feasible and we use numerical methods to obtain the estimates of parameters. We derived the score function of our proposed joint distribution and investigate the Fisher information matrix. The score function obtained using the joint model is complex as expected due to the complex normalizing constant. Moreover, we presented a theorem for the CL method; the Godambe information matrix for conditional specified models when the respective conditional distributions belong to the exponential family of distribution followed by complete proof.

We carried out a simulation study for different sample sizes to investigate the properties of maximum likelihood estimates and composite likelihood. Estimates of the model parameters were obtained using both MLE and CLE methods. Data were generated using Gibb's algorithms. Apart from the estimates, we calculated the variance, bias, and wall time. Further, joint asymptotic relative efficiency (JARE) between MLE and CLE was calculated for different sample sizes. According to the results, the ML method has less bias (and nearly zero in some cases) than the PL method. In both methods, the variances of the estimates decrease as the sample size increases. The MLE of the vector has less bias compared to the CL method. In terms of wall time, for larger sample size, the computational advantage of the CL method surpasses that of the ML method quickly. Further, we note that even the CL method shows large wall times when the sample size increases. In that case, parallelizing the code would be more effective. For relative efficiency, the ML method is better than the CL method. However, the efficiency of the CL method increases with sample size and has the potential to surpass the ML method for much larger sample sizes. Thus, we can conclude that choosing the CL method over the ML method is a trade-off between efficiency and computational cost.

Appendix A: Useful derivatives for score function and Information matrix

A.1 Useful derivatives for score function: MLE

$$\frac{\partial m_{01}}{\partial \mu_1^{(0)}} = \frac{1-r}{(1-\rho^2)\sigma_{11}^2}, \qquad \frac{\partial m_{02}}{\partial \mu_1^{(0)}} = \frac{-\rho(1-r)}{(1-\rho^2)\sigma_{11}\sigma_{22}}$$

$$\frac{\partial m_{01}}{\partial \mu_2^{(0)}} = \frac{-\rho(1-r)}{(1-\rho^2)\sigma_{11}\sigma_{22}}, \qquad \frac{\partial m_{02}}{\partial \mu_2^{(0)}} = \frac{1-r}{(1-\rho^2)\sigma_{22}^2}$$

$$\frac{\partial m_{01}}{\partial \mu_1^{(1)}} = \frac{r}{(1-\rho^2)\sigma_{11}^2}, \qquad \frac{\partial m_{02}}{\partial \mu_1^{(1)}} = \frac{-\rho r}{(1-\rho^2)\sigma_{11}\sigma_{22}}$$

$$\frac{\partial m_{01}}{\partial \mu_2^{(1)}} = \frac{\rho r}{(1-\rho^2)\sigma_{11}\sigma_{22}}, \qquad \frac{\partial m_{02}}{\partial \mu_2^{(1)}} = \frac{r}{(1-\rho^2)\sigma_{22}^2}$$

$$\frac{\partial m_{01}}{\partial \sigma_{11}} = \frac{-2\mu_1^{(r)}}{(1-\rho^2)\sigma_{11}^3} + \frac{\rho\mu_2(r)}{(1-\rho^2)\sigma_{22}\sigma_{11}^2}, \qquad \frac{\partial m_{02}}{\partial \sigma_{11}} = \frac{\rho\mu_1^{(r)}}{(1-\rho^2)\sigma_{22}\sigma_{11}^2}$$

$$\frac{\partial m_{01}}{\partial \sigma_{22}} = \frac{\rho\mu_2(r)}{(1-\rho^2)\sigma_{11}\sigma_{22}^2}, \qquad \frac{\partial m_{02}}{\partial \sigma_{22}} = \frac{-2\mu_1^{(r)}}{(1-\rho^2)\sigma_{11}^3} + \frac{\rho\mu_2(r)}{(1-\rho^2)\sigma_{22}\sigma_{11}^2}$$

$$\frac{\partial m_{01}}{\partial \rho} = \frac{2\rho\mu_1^{(r)}}{(1-\rho^2)^2\sigma_{11}^2} - \frac{\mu_2^{(r)}}{\sigma_{11}\sigma_{22}}\left[\frac{2\rho^2}{(1-\rho^2)^2} + \frac{1}{(1-\rho^2)}\right], \qquad \frac{\partial m_{02}}{\partial \rho} = \frac{2\rho\mu_1^{(r)}}{(1-\rho^2)^2\sigma_{11}^2} - \frac{\mu_2^{(r)}}{\sigma_{11}\sigma_{22}}\left[\frac{2\rho^2}{(1-\rho^2)^2} + \frac{1}{(1-\rho^2)}\right]$$

$$\frac{\partial m_{01}}{\partial \alpha_0} = 0, \qquad \frac{\partial m_{02}}{\partial \alpha_0} = 0$$

$$\frac{\partial m_{01}}{\partial \alpha_1} = -r, \qquad \frac{\partial m_{02}}{\partial \alpha_1} = 0$$

$$\frac{\partial m_{01}}{\partial \alpha_2} = 0, \qquad \frac{\partial m_{02}}{\partial \alpha_2} = -r$$

$$\frac{\partial m_{03}}{\partial \mu_1^{(0)}} = 0 \qquad \frac{\partial m_{04}}{\partial \mu_1^{(0)}} = 0$$

$$\frac{\partial m_{03}}{\partial \mu_2^{(0)}} = 0 \qquad \frac{\partial m_{04}}{\partial \mu_2^{(0)}} = 0$$

$$\frac{\partial m_{03}}{\partial \mu_1^{(1)}} = 0 \qquad \frac{\partial m_{04}}{\partial \mu_1^{(1)}} = 0$$

$$\frac{\partial m_{03}}{\partial \mu_2^{(1)}} = 0 \qquad \frac{\partial m_{04}}{\partial \mu_2^{(1)}} = 0$$

$$\frac{\partial m_{03}}{\partial \sigma_{11}} = \frac{1}{(1-\rho^2)\sigma_{11}^3} \qquad \frac{\partial m_{04}}{\partial \sigma_{11}} = \frac{-\rho}{(1-\rho^2)\sigma_{11}^2\sigma_{22}}$$

$$\frac{\partial m_{03}}{\partial \sigma_{22}} = 0 \qquad \frac{\partial m_{04}}{\partial \sigma_{22}} = \frac{-\rho}{(1-\rho^2)\sigma_{22}^2\sigma_{11}}$$

$$\frac{\partial m_{03}}{\partial \rho} = \frac{-\rho}{(1-\rho^2)^2\sigma_{11}^2} \qquad \frac{\partial m_{04}}{\partial \rho} = \frac{1+\rho^2}{(1-\rho^2)^2\sigma_{11}\sigma_{22}}$$

$$\frac{\partial m_{03}}{\partial \alpha_0} = 0 \qquad \frac{\partial m_{04}}{\partial \alpha_0} = 0$$

$$\frac{\partial m_{03}}{\partial \alpha_1} = 0 \qquad \frac{\partial m_{04}}{\partial \alpha_1} = 0$$

$$\frac{\partial m_{03}}{\partial \alpha_2} = 0 \qquad \frac{\partial m_{04}}{\partial \alpha_2} = 0$$

$$\frac{\partial m_{05}}{\partial \mu_1^{(0)}} = 0 \qquad \frac{\partial m_{10}}{\partial \mu_1^{(0)}} = 0$$

$$\frac{\partial m_{05}}{\partial \mu_2^{(0)}} = 0 \qquad \frac{\partial m_{10}}{\partial \mu_2^{(0)}} = 0$$

$$\frac{\partial m_{05}}{\partial \mu_1^{(1)}} = 0 \qquad \frac{\partial m_{10}}{\partial \mu_1^{(1)}} = 0$$

$$\frac{\partial m_{05}}{\partial \mu_2^{(1)}} = 0 \qquad \frac{\partial m_{10}}{\partial \mu_2^{(1)}} = 0$$

$$\frac{\partial m_{05}}{\partial \sigma_{11}} = 0 \qquad \frac{\partial m_{10}}{\partial \sigma_{11}} = 0$$

$$\frac{\partial m_{05}}{\partial \sigma_{22}} = \frac{1}{(1-\rho^2)\sigma_{22}^3} \qquad \frac{\partial m_{10}}{\partial \sigma_{22}} = 0$$

$$\frac{\partial m_{05}}{\partial \rho} = \frac{-\rho}{(1-\rho^2)^2\sigma_{22}^2} \qquad \frac{\partial m_{10}}{\partial \rho} = 0$$

$$\frac{\partial m_{05}}{\partial \alpha_0} = 0 \qquad \frac{\partial m_{10}}{\partial \alpha_0} = 1$$

$$\frac{\partial m_{05}}{\partial \alpha_1} = 0 \qquad \frac{\partial m_{10}}{\partial \alpha_1} = 0$$

$$\frac{\partial m_{05}}{\partial \alpha_2} = 0 \qquad \frac{\partial m_{10}}{\partial \alpha_2} = 0$$

$$\frac{\partial m_{11}}{\partial \mu_1^{(0)}} = 0 \qquad \frac{\partial m_{12}}{\partial \mu_1^{(0)}} = 0$$

$$\frac{\partial m_{11}}{\partial \mu_2^{(0)}} = 0 \qquad \frac{\partial m_{12}}{\partial \mu_2^{(0)}} = 0$$

$$\frac{\partial m_{11}}{\partial \mu_1^{(1)}} = 0 \qquad \frac{\partial m_{12}}{\partial \mu_1^{(1)}} = 0$$

$$\frac{\partial m_{11}}{\partial \mu_2^{(1)}} = 0 \qquad \frac{\partial m_{12}}{\partial \mu_2^{(1)}} = 0$$

$$\frac{\partial m_{11}}{\partial \sigma_{11}} = 0 \qquad \frac{\partial m_{12}}{\partial \sigma_{11}} = 0$$

$$\frac{\partial m_{11}}{\partial \sigma_{22}} = 0 \qquad \frac{\partial m_{12}}{\partial \sigma_{22}} = 0$$

$$\frac{\partial m_{11}}{\partial \rho} = 0 \qquad \frac{\partial m_{12}}{\partial \rho} = 0$$

$$\frac{\partial m_{11}}{\partial \alpha_0} = 0 \qquad \frac{\partial m_{12}}{\partial \alpha_0} = 0$$

$$\frac{\partial m_{11}}{\partial \alpha_1} = 1 \qquad \frac{\partial m_{12}}{\partial \alpha_1} = 0$$

$$\frac{\partial m_{11}}{\partial \alpha_2} = 0 \qquad \frac{\partial m_{12}}{\partial \alpha_2} = 1$$

A.2 Derivatives of z with respect to the full parameter space θ

$$\frac{\partial z}{\partial \mu_1^{(0)}} = \sum_{i=1}^{n} \left[\frac{-2(1-r_i)}{\sigma_1^2}(y_{1i} - [r_i\mu_1^{(1)} + (1-r_i)\mu_1^{(0)}]) + \frac{2\rho(1-r_i)}{\sigma_1\sigma_2}(-[r_i\mu_2^{(1)} + (1-r_i)\mu_2^{(0)}]) \right]$$

$$\frac{\partial z}{\partial \mu_1^{(1)}} = \sum_{i=1}^{n} \left[\frac{-2r_i}{\sigma_1^2}(y_{1i} - [r_i\mu_1^{(1)} + (1-r_i)\mu_1^{(0)}]) + \frac{2\rho r_i}{\sigma_1\sigma_2}(-[r_i\mu_2^{(1)} + (1-r_i)\mu_2^{(0)}]) \right]$$

$$\frac{\partial z}{\partial \mu_2^{(0)}} = \sum_{i=1}^{n} \left[\frac{-2(1-r_i)}{\sigma_2^2}(-[r_i\mu_2^{(1)} + (1-r_i)\mu_2^{(0)}]) + \frac{2\rho(1-r_i)}{\sigma_1\sigma_2}(y_{1i} - [r_i\mu_1^{(1)} + (1-r_i)\mu_1^{(0)}]) \right]$$

$$\frac{\partial z}{\partial \mu_2^{(1)}} = \sum_{i=1}^{n} \left[\frac{-2r}{\sigma_2^2}(-[r_i\mu_2^{(1)} + (1-r_i)\mu_2^{(0)}]) + \frac{2\rho r_i}{\sigma_1\sigma_2}(y_{1i} - [r_i\mu_1^{(1)} + (1-r_i)\mu_1^{(0)}]) \right]$$

$$\frac{\partial z}{\partial \sigma_1} = \sum_{i=1}^{n} \left[\frac{-2}{\sigma_1^3}(y_{1i} - [r_i\mu_1^{(1)} + (1-r_i)\mu_1^{(0)}])^2 + \frac{2\rho}{\sigma_1^2\sigma_2}(y_{1i} - [r_i\mu_1^{(1)} + (1-r_i)\mu_1^{(0)}])(-[r_i\mu_2^{(1)} + (1-r_i)\mu_2^{(0)}]) \right]$$

$$\frac{\partial z}{\partial \sigma_2} = \sum_{i=1}^{n} \left[\frac{-2}{\sigma_2^3}(-[r_i\mu_2^{(1)} + (1-r_i)\mu_2^{(0)}])^2 + \frac{2\rho}{\sigma_1\sigma_2^2}(y_{1i} - [r_i\mu_1^{(1)} + (1-r_i)\mu_1^{(0)}])(-[r_i\mu_2^{(1)} + (1-r_i)\mu_2^{(0)}]) \right]$$

$$\frac{\partial z}{\partial \rho} = \sum_{i=1}^{n} \left[\frac{-2}{\sigma_1\sigma_2}(y_{1i} - [r_i\mu_1^{(1)} + (1-r_i)\mu_1^{(0)}])(-[r_i\mu_2^{(1)} + (1-r_i)\mu_2^{(0)}]) \right]$$

$$\frac{\partial z}{\partial \alpha_0} = 0$$

$$\frac{\partial z}{\partial \alpha_1} = 0$$

$$\frac{\partial z}{\partial \alpha_2} = 0$$

Appendix B: Data Analysis

B.1 Names of 100 stocks

Stock Number	Stock	Name of the Company
1	AAL	American Airlines Group Inc
2	AAPL	Apple Inc
3	ADP	Automatic Data Processing, Inc
4	AEE	Ameren Corp
5	AES	AES Corp
6	BDN	Brandywine Realty Trust
7	BF.B	Brown–Forman
8	BMRA	Biomerica
9	CBM	Center Bancorp, Inc.
10	CLI	Mack-Cali Realty Corporation
11	CMI	Cummins
12	CMS	CMS Energy
13	CORE	Core-Mark
14	COUP	Coupa Software

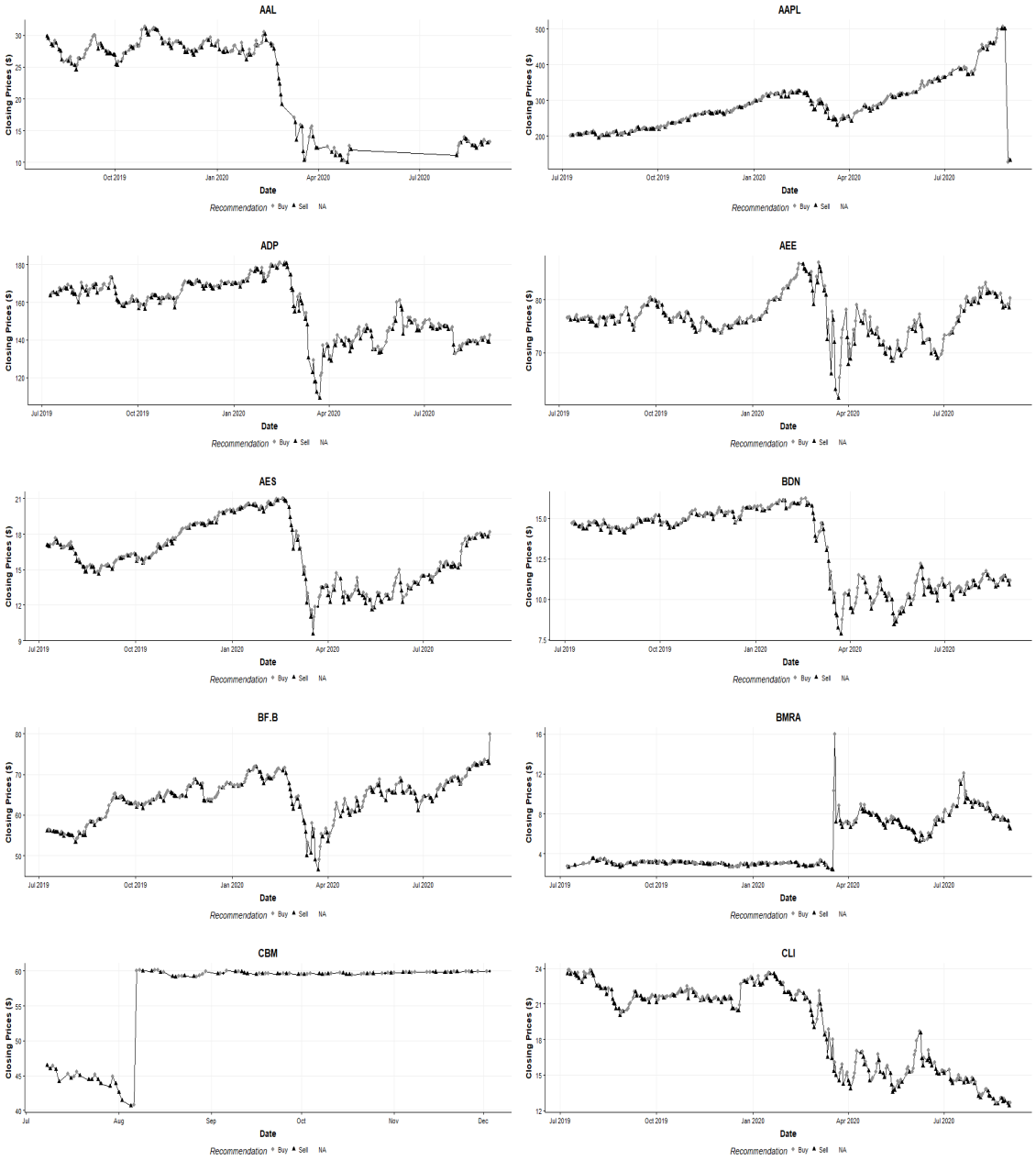
15	CPB	Campbell Soup Company
16	CRM	Salesforce.com, Inc.
17	CSTM	Constellium
18	CSX	CSX Corporation
19	D	Dominion Energy
20	DATA	GlobalData Plc
21	DRE	Duke Realty
22	DTE	DTE Energy
23	ED	Consolidated Edison
24	EFX	Equifax Inc.
25	EQR	Equity Residential
26	ES	Eversource Energy
27	ETR	Entergy Corporation
28	EVRG	Evergy
29	EXC	Exelon Corporation
30	EYEN	Eyenovia
31	FE	FirstEnergy Corp
32	FICO	Fair, Isaac and Company
33	FIS	FIS
34	G	Genpact
35	GGAL	Galicia Financial Group
36	GKOS	Glaukos Corp
37	GLDD	Great Lakes Dredge and Dock Company

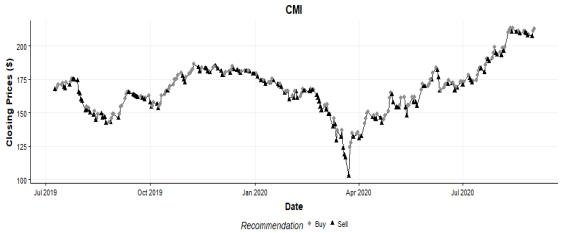
38	GLW	Corning Incorporated
39	GNW	Genworth Financial
40	GVP	GSE Systems, Inc.
41	HOLX	Hologic, Inc.
42	ICL	ICL Group Ltd.
43	ILMN	Illumina, Inc.
44	IP	The International Paper Company
45	ISRG	Intuitive Surgical, Inc.
46	KEN	Kenon Holdings
47	KIM	Kimco Realty Corporation
48	KN	Knowles
49	KOF	Coca Cola Femsa S.A.B. de C.V.
50	LANC	Lancaster Colony Corporation
51	LMRK	Landmark Infrastructure
52	LNT	Alliant Energy
53	LNTH	Lantheus Holdings
54	LPT	Liberty Property Trust
55	LSI	Life Storage, Inc.
56	LULU	Lululemon Athletica
57	MAA	Mid-America Apartment Communities
58	MAT	Mattel, Inc.
59	MDB	MongoDB Inc.
60	MESA	Mesa Air Group

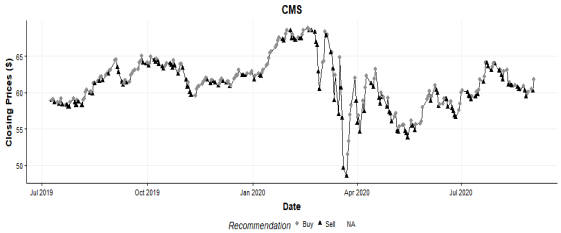
61	MSCI	MSCI Inc.
62	NEE	NextEra Energy Inc
63	NEWT	NEWTEK Business Services Corp
64	NI	NiSource Inc.
65	NKE	Nike Inc
66	NMR	Nomura Holdings Inc
67	NSA	National Storage Affiliates Trust
68	NSSC	Napco Security Technologies Inc
69	NWE	NorthWestern Corp
70	OFIX	Orthofix Medical Inc
71	PEP	PepsiCo, Inc.
72	PFGC	Performance Food Group Co
73	PG	Procter & Gamble Co
74	PLD	Prologis Inc
75	PNTR	Pantera Silver Corp
76	PNW	Pinnacle West Capital Corporation
77	PRMW	Primo Water Corp (MISSISSAUGA)
78	RDUS	Radius Health Inc
79	REDU	RISE Education Cayman Ltd
80	RWLK	Rewalk Robotics Ltd
81	SHOP	Shopify Inc
82	SNH	Steinhoff International Holdings NV
83	SPOT	Spotify Technology SA

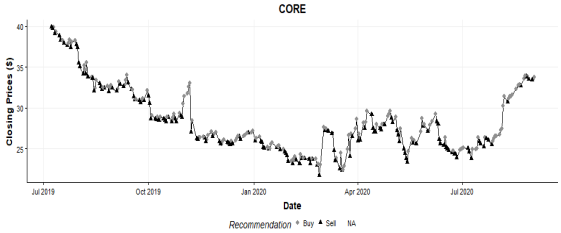
84	STAG	Stag Industrial Inc
85	TFX	Teleflex Incorporated
86	TGTX	TG Therapeutics Inc common stock
87	TRI	Thomson Reuters Corp
88	TSN	Tyson Foods, Inc.
89	UDR	UDR, Inc.
90	UIHC	United Insurance Holdings Corp
91	UNIT	Uniti Group Inc
92	VFC	VF Corp
93	VLRS	Controladora Vuela Co Avcn SA CV
94	VNET	21Vianet Group Inc - ADR
95	WEC	WEC Energy Group Inc
96	WP	WP Energy PCL
97	WRK	Westrock Co
98	XEL	Xcel Energy Inc
99	XRAY	DENTSPLY SIRONA Inc
100	ZYXI	Zynex Inc.

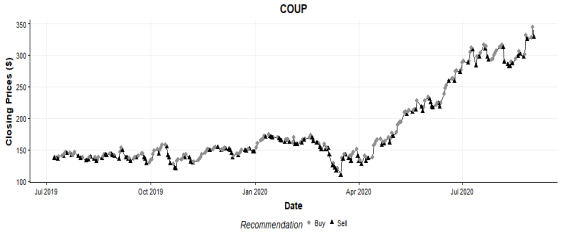
B.1.1 Plots of Closing Prices Vs. Recommendation

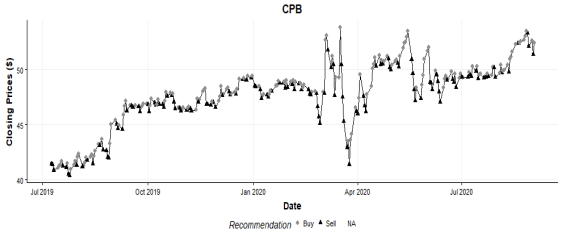


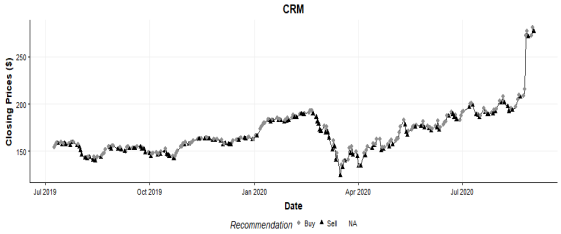

CMI
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

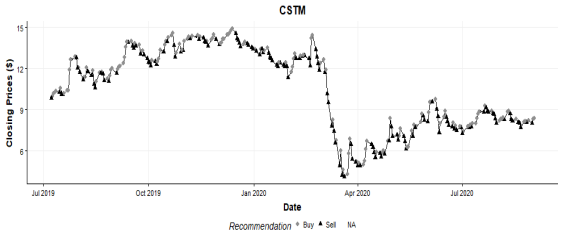

CMS
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

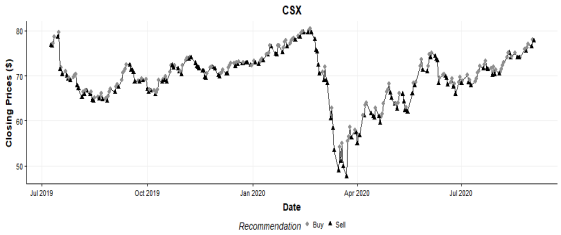

CORE
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

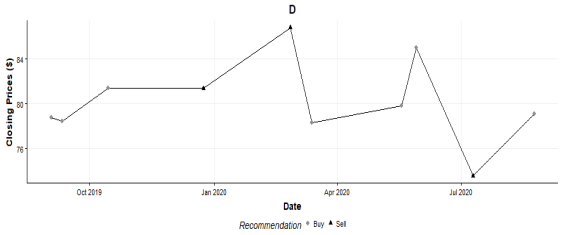

COUP
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

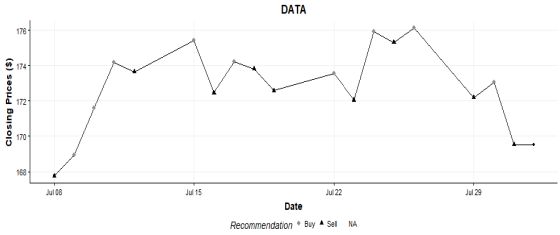

CPB
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

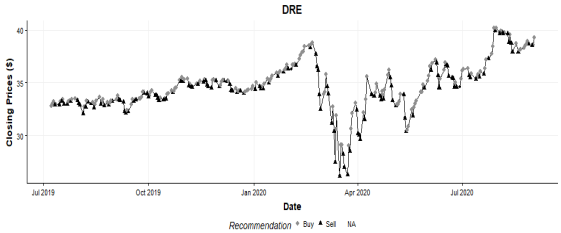

CRM
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

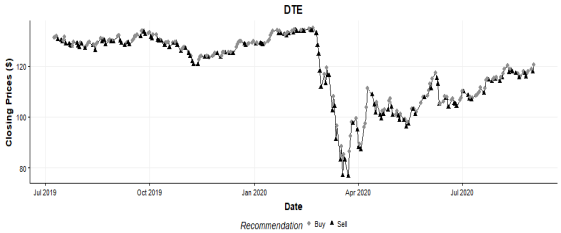

CSTM
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

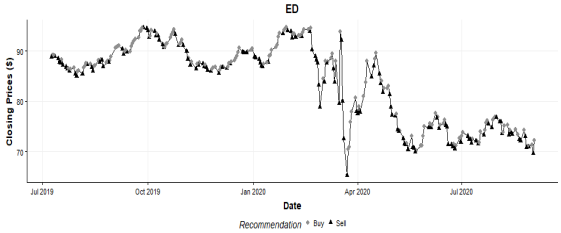

CSX
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

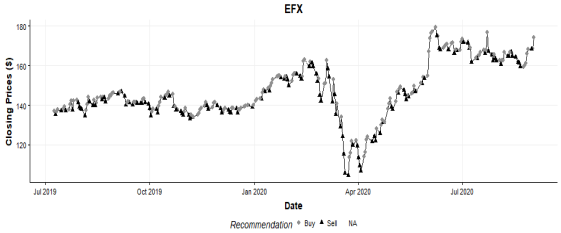

D
Closing Prices ($)
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

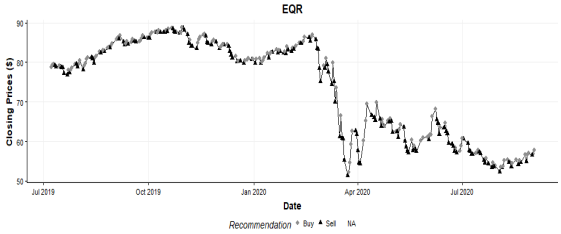

DATA
Closing Prices ($)
Jul 08
Jul 15
Jul 22
Jul 29
Date
Recommendation Buy Sell NA

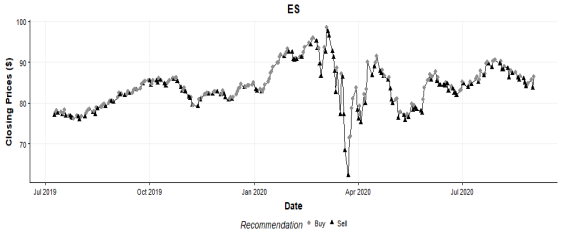

DRE
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

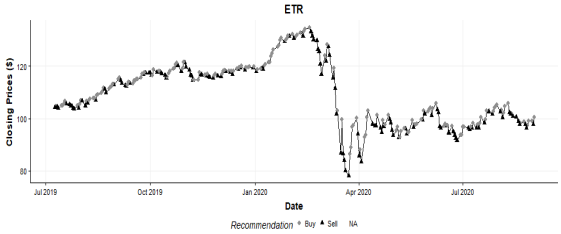

DTE
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

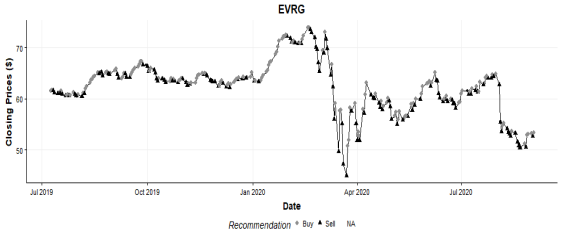

ED
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

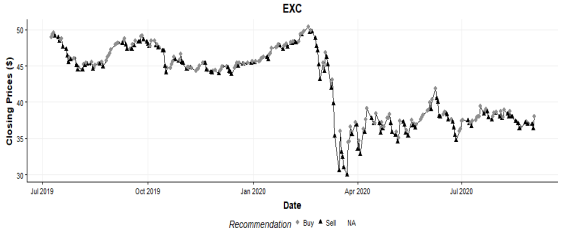

EFX
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

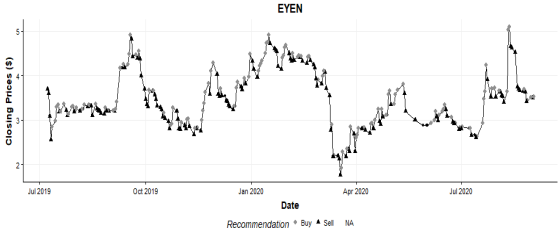

EQR
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

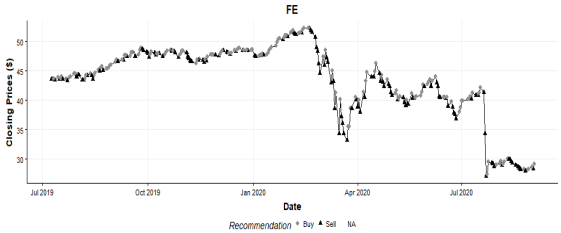

ES
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

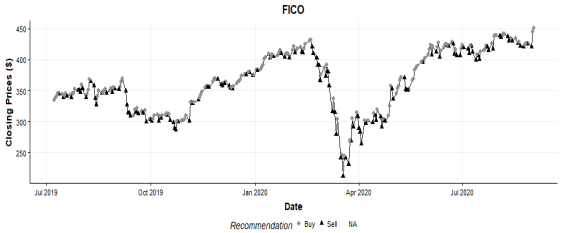

ETR
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

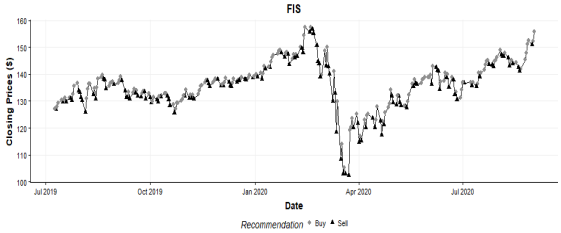

EVRG
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

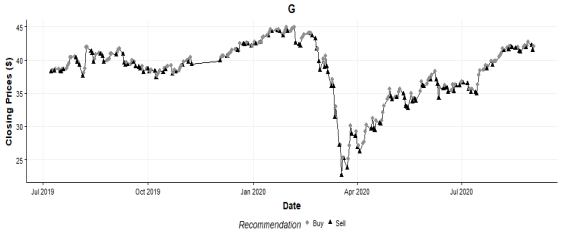

EXC
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

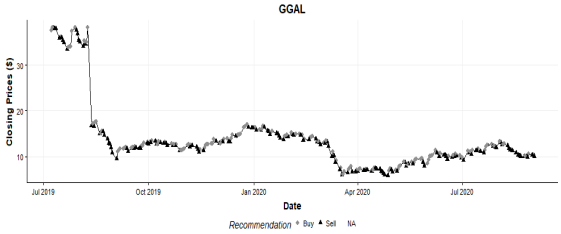

EYEN
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

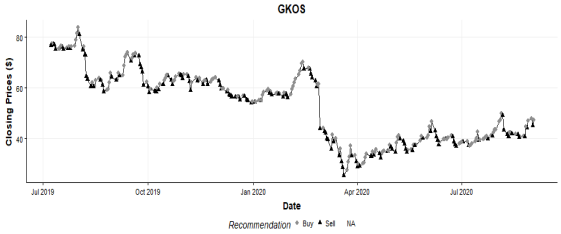

FE
Closing Prices ($)
50
45
40
35
30
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

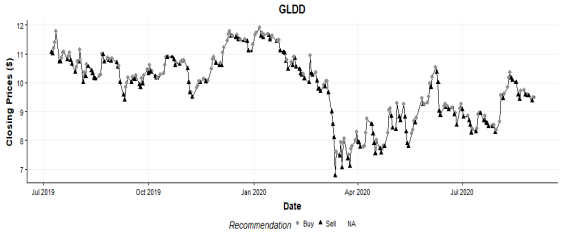

FICO
Closing Prices ($)
450
400
350
300
250
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

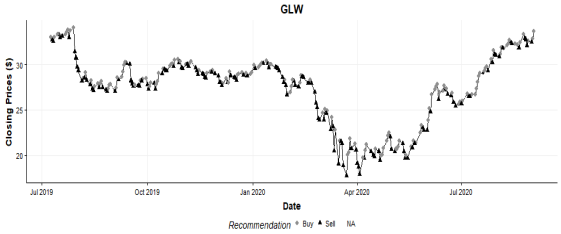

FIS
Closing Prices ($)
160
150
140
130
120
110
100
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

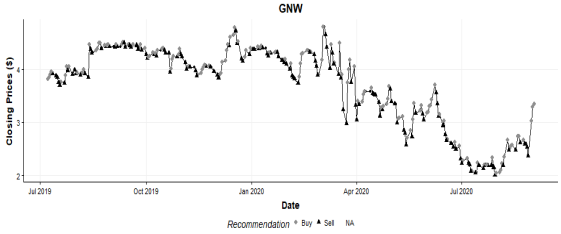

G
Closing Prices ($)
45
40
35
30
25
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

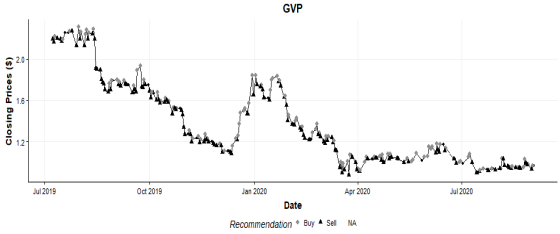

GGAL
Closing Prices ($)
30
20
10
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

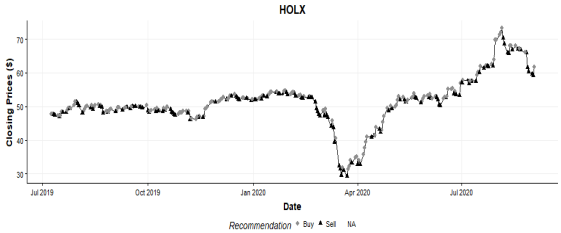

GKOS
Closing Prices ($)
80
60
40
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

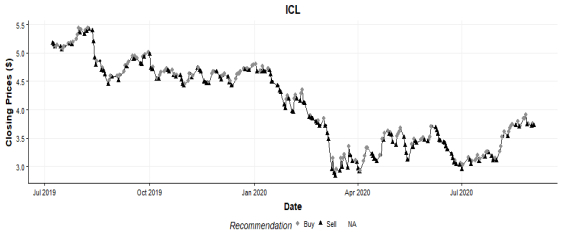

GLDD
Closing Prices ($)
12
11
10
9
8
7
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

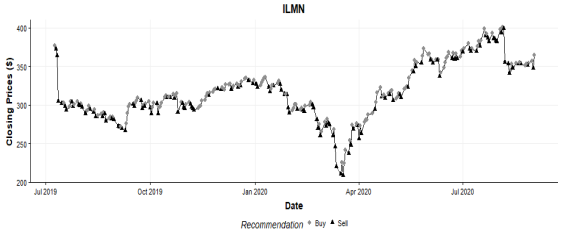

GLW
Closing Prices ($)
30
25
20
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

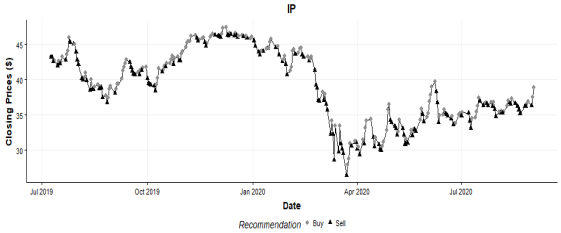

GNW
Closing Prices ($)
4
3
2
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

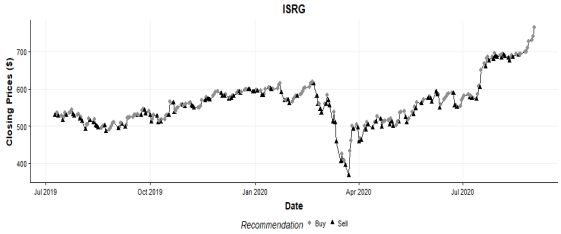

GVP
Closing Prices ($)
2.0
1.6
1.2
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

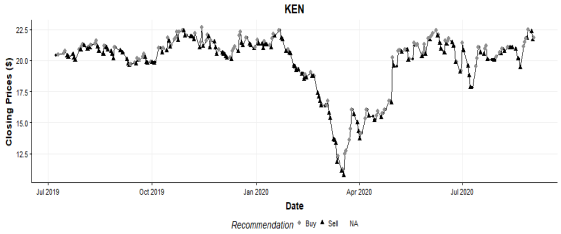

HOLX
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

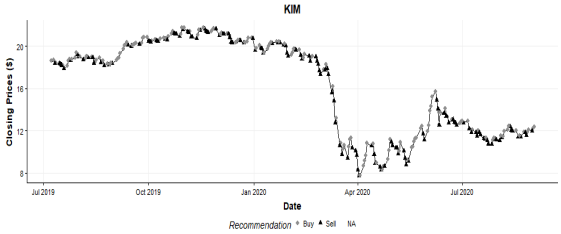

ICL
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

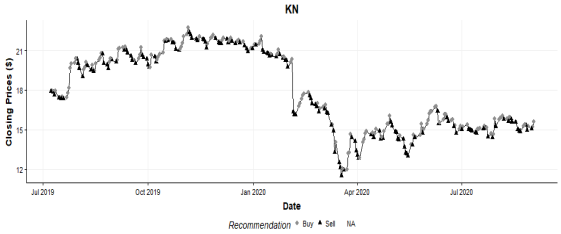

ILMN
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

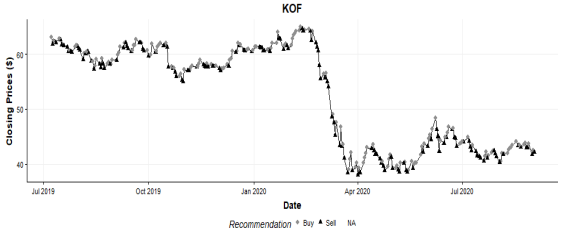

IP
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

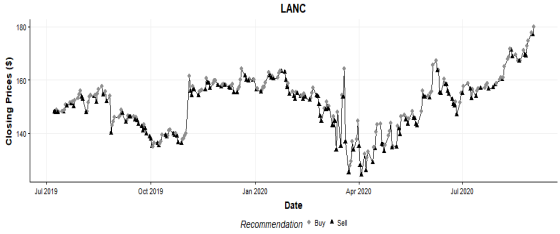

ISRG
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

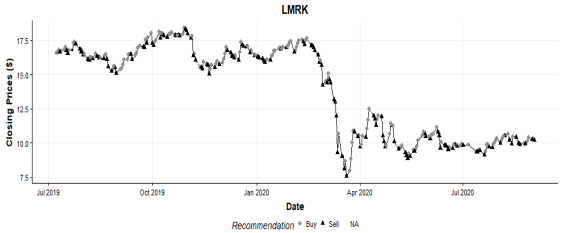

KEN
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

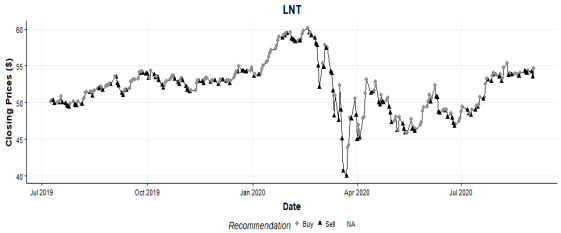

KIM
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

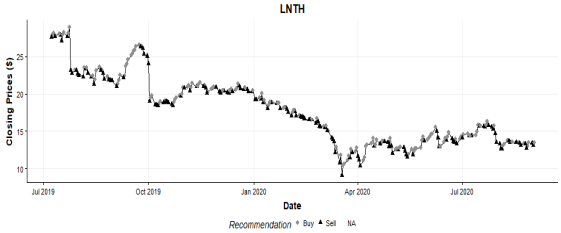

KN
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

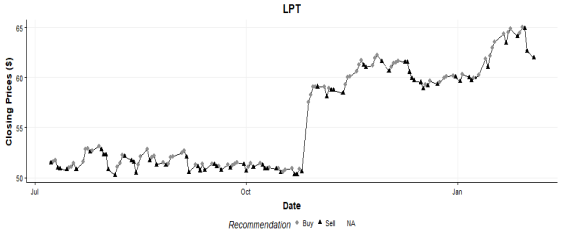

KOF
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

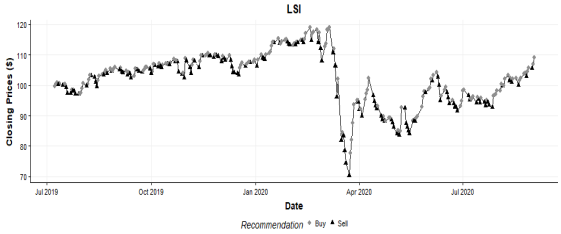

LANC
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

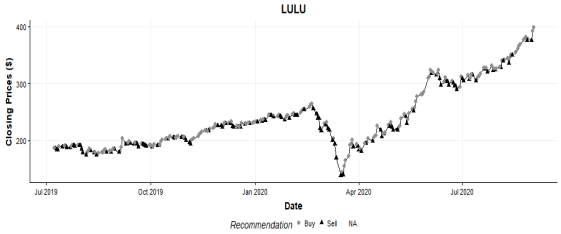

LMRK
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

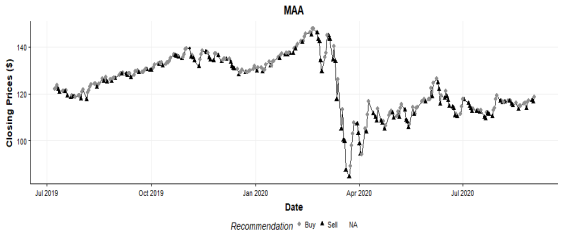

LNT
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

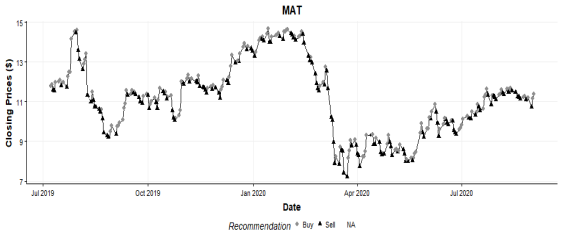

LNTH
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

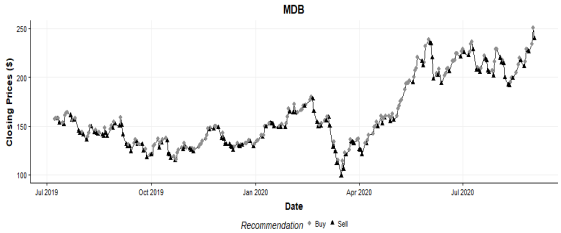

LPT
Closing Prices ($)
Jul
Oct
Jan
Date
Recommendation Buy Sell NA

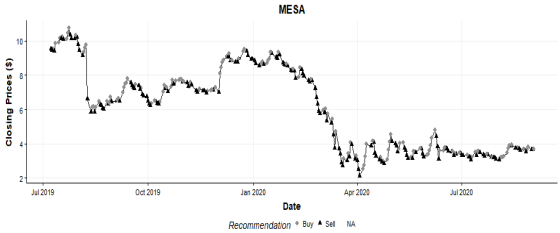

LSI
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

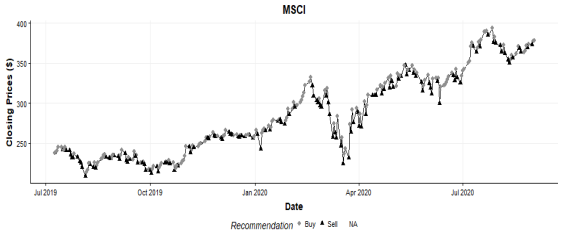

LULU
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

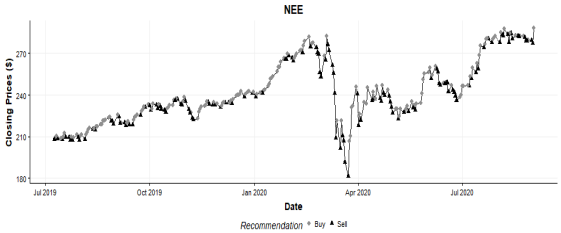

MAA
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

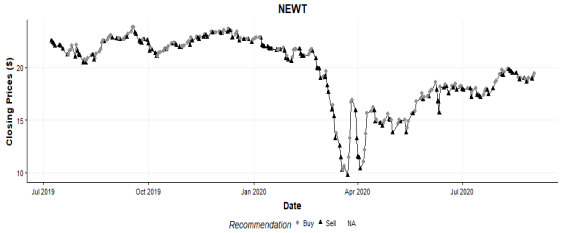

MAT
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

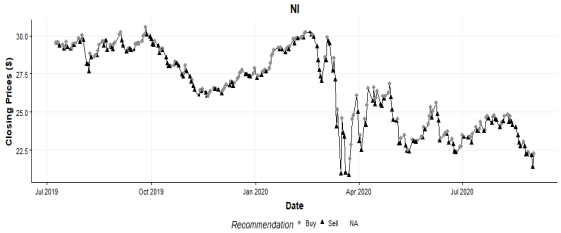

MDB
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

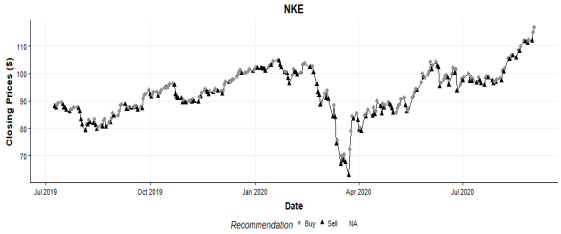

MESA
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

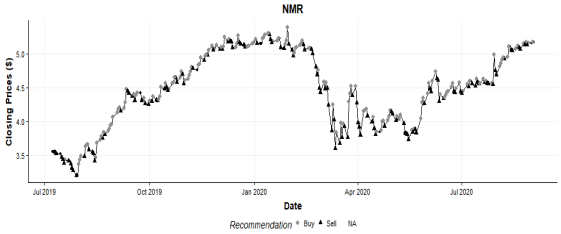

MSCI
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

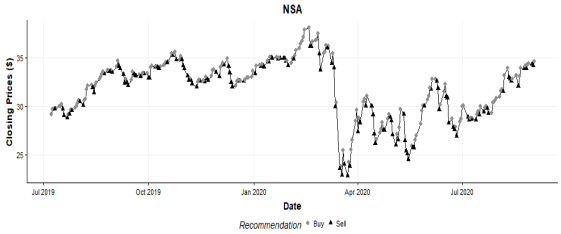

NEE
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

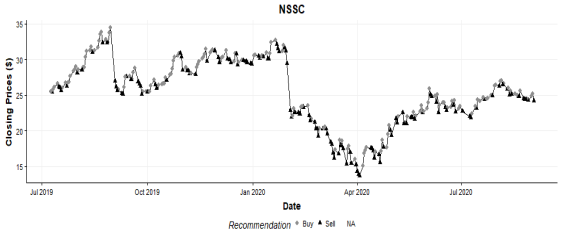

NEWT
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

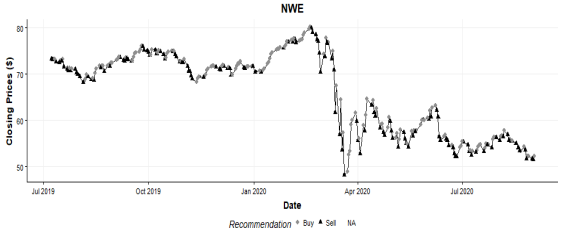

NI
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

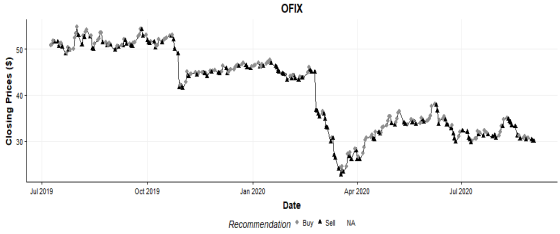

NKE
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

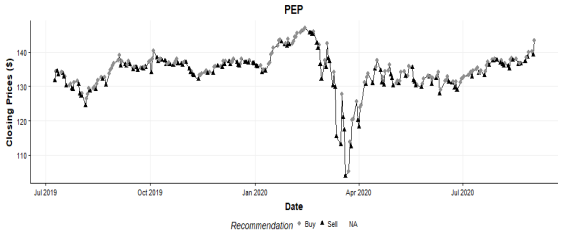

NMR
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

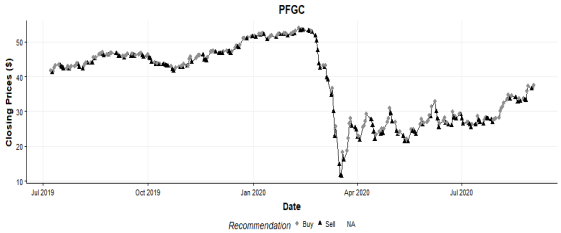

NSA
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

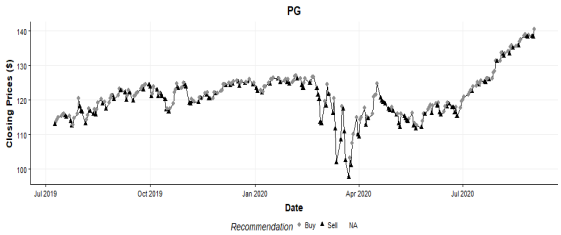

NSSC
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

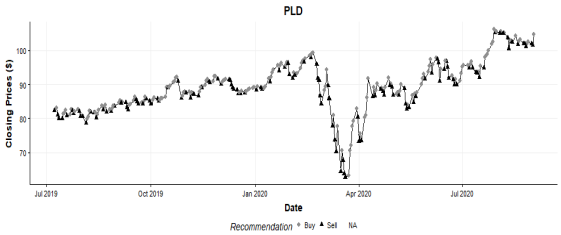

NWE
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

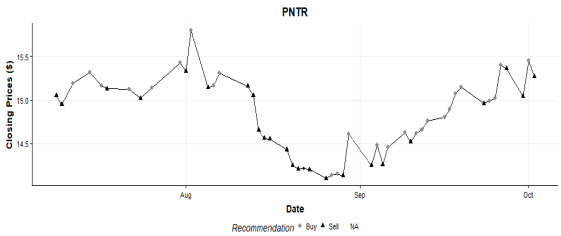

OFIX
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

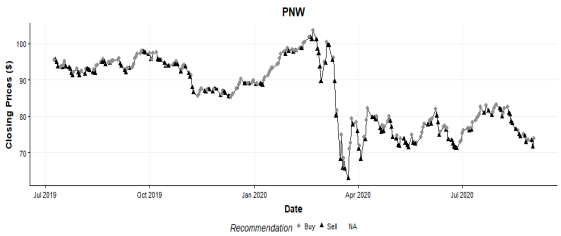

PEP
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

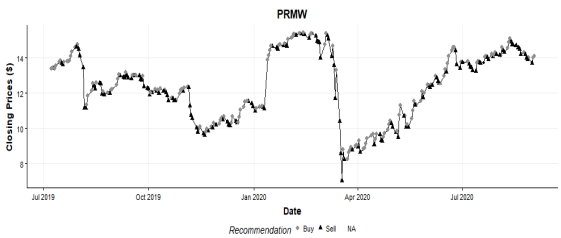

PFGC
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

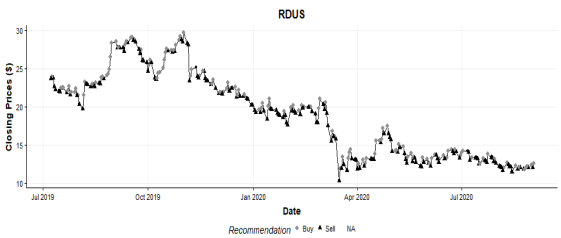

PG
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

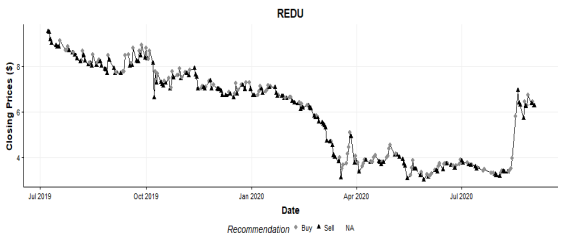

PLD
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

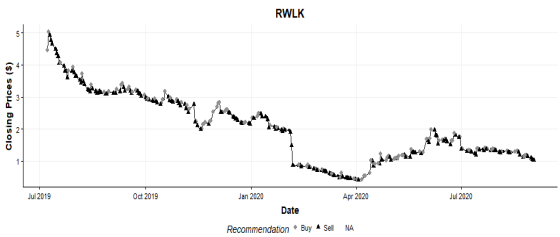

PNTR
Closing Prices ($)
Aug
Sep
Oct
Date
Recommendation Buy Sell NA

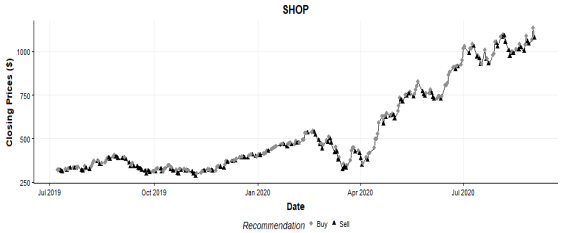

PNW
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

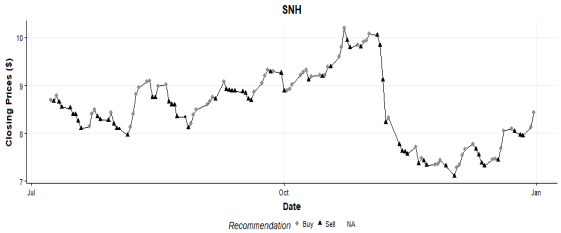

PRMW
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

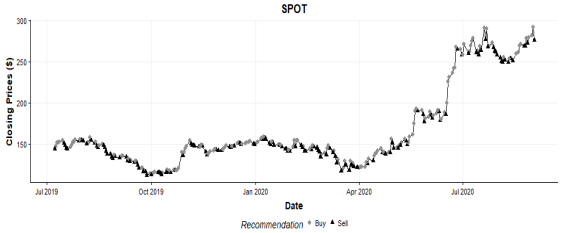

RDUS
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

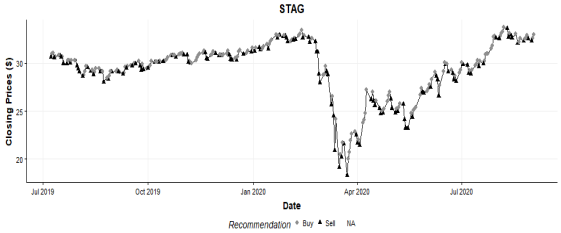

REDU
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

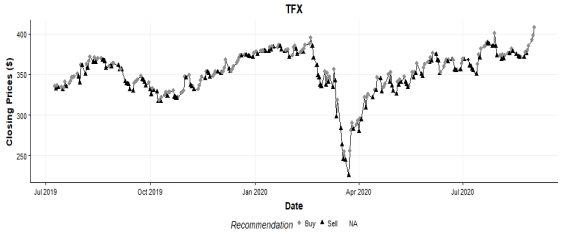

RWLK
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

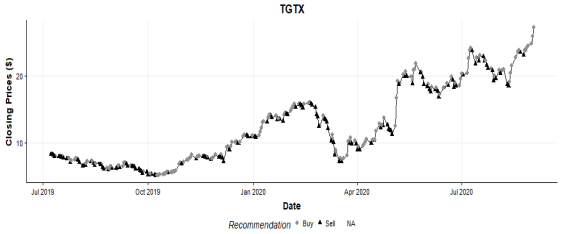

SHOP
Closing Prices ($)
1000
750
500
250
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

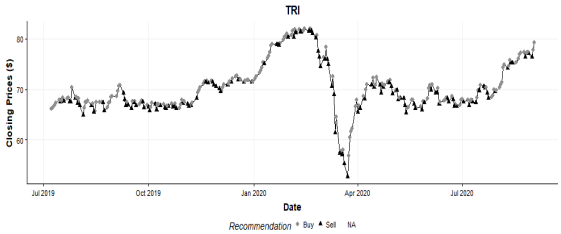

SNH
Closing Prices ($)
10
9
8
7
Jul
Oct
Jan
Date
Recommendation Buy Sell NA

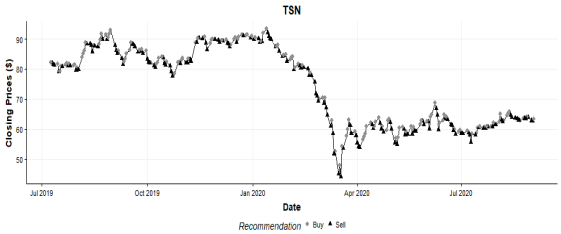

SPOT
Closing Prices ($)
300
250
200
150
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

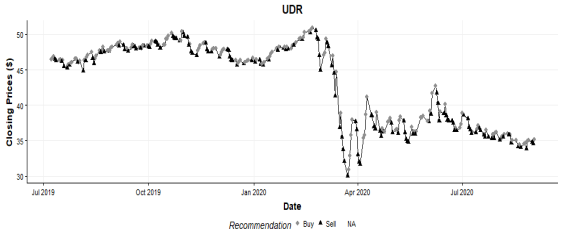

STAG
Closing Prices ($)
30
25
20
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

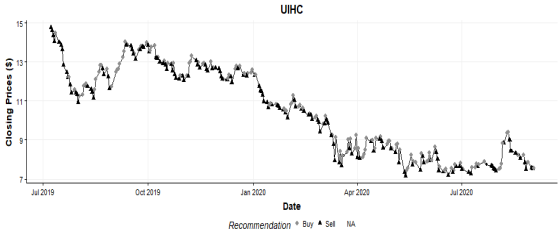

TFX
Closing Prices ($)
400
350
300
250
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

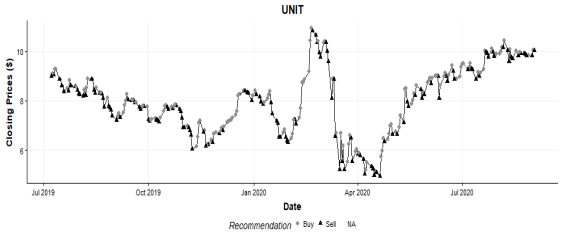

TGTX
Closing Prices ($)
20
10
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

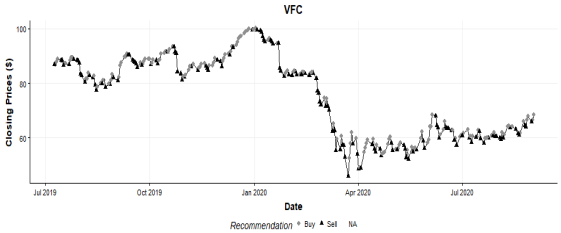

TRI
Closing Prices ($)
80
70
60
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

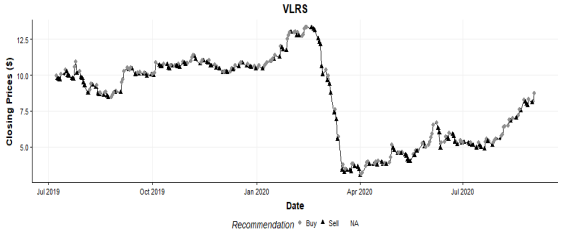

TSN
Closing Prices ($)
90
80
70
60
50
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell

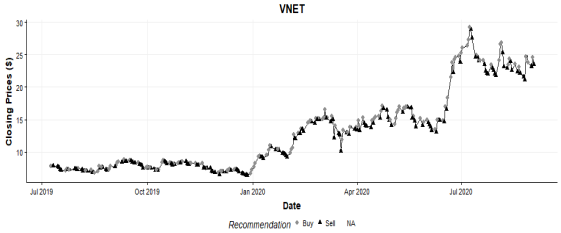

UDR
Closing Prices ($)
50
45
40
35
30
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

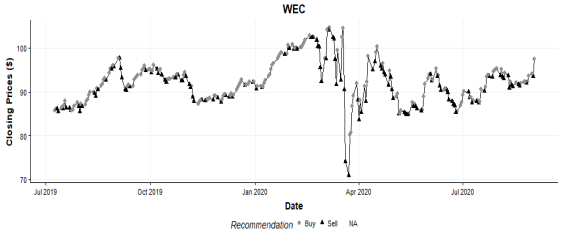

UIHC
Closing Prices ($)
15
13
11
9
7
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

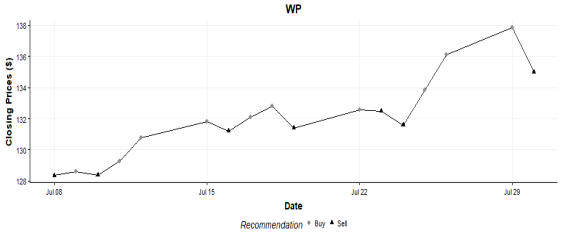

UNIT
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

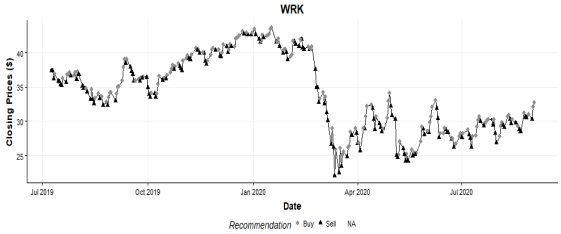

VFC
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

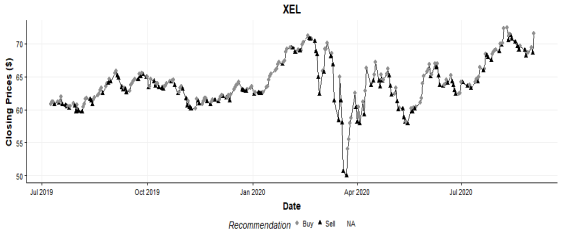

VLRS
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

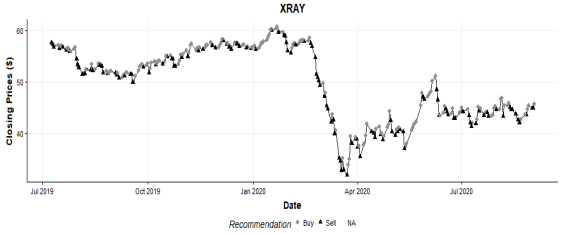

VNET
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

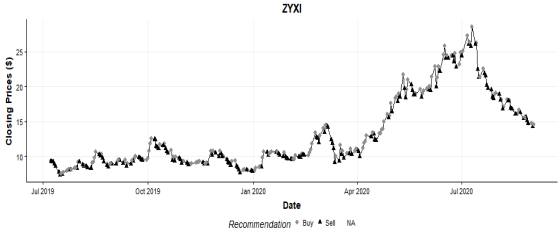

WEC
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA


WP
Closing Prices ($)
Jul 08
Jul 15
Jul 22
Jul 29
Date
Recommendation Buy Sell


WRK
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA


XEL
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA


XRAY
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA


ZYXI
Closing Prices ($)
Jul 2019
Oct 2019
Jan 2020
Apr 2020
Jul 2020
Date
Recommendation Buy Sell NA

Bibliography

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of economic literature*, 19(4):1483–1536.

Arnold, B. C. (1987). Bivariate distributions with pareto conditionals. *Stat. Probab. Lett.*, 5:263–266.

Arnold, B. C., Castillo, E., and Sarabia, J.-M. (1999). *Conditional specification of statistical models*. Springer.

Arnold, B. C., Castillo, E., and Sarabia, J. M. (2001). Conditionally specified distributions: an introduction (with comments and a rejoinder by the authors). *Statistical Science*, 16(3):249–274.

Arnold, B. C., Castillo, E., and Sarabia, J. M. (2004). Compatibility of partial or complete conditional probability specifications. *Journal of Statistical Planning and Inference*, 123(1):133 – 159.

Arnold, B. C. and Gokhale, D. (1994). On uniform marginal representation of contingency tables. *Stat. Probab. Lett.*, 21(4):311–316.

Arnold, B. C. and Gokhale, D. V. (1998). Distributions most nearly compatible with given families of conditional distributions. *Test*, 7(2):377–390.

Arnold, B. C., James, S., and Press, S. J. (1989). Compatible Conditional Distributions. *Press Source J. Am. Stat. Assoc.*, 84:152–156.

Arnold, B. C. and Strauss, D. (1991a). Pseudolikelihood Estimation: Some Examples. *Source Sankhyā Indian J. Stat. Ser. B*, 53:233–243.

Arnold, B. C. and Strauss, D. J. (1991b). Bivariate Distributions with Conditionals in Prescribed Exponential Families. *J. R. Stat. Soc. Ser. B*, 53:700–700.

Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. John Wiley Sons Ltd.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

Besag, J. (1975). Statistical Analysis of Non-Lattice Data Author (s): Julian Besag Published by : Blackwell Publishing for the Royal Statistical Society. *Stat.*, 24:179–195.

Bickel, P. J. and Doksum, K. A. (2006). *Mathematical Statistics : Basic Ideas And Selected Topics*.

Brown, L. D. (1986). Fundamentals of statistical exponential families: with applications in statistical decision theory. Ims.

Casella, G. and Lehmann, E. L. (2006). *Theory of Point Estimation.*, volume 147.

Castillo, E. and Galambos, J. (1989). Conditional distributions and the bivariate normal distribution. *Metrika*, 36(1):209–214.

Chandler, R. E. and Bate, S. (2007). Inference for Clustered Data Using the Independence Loglikelihood. 94:167–183.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B. (2021). *shiny: Web Application Framework for R*. R package version 1.6.0.

Chen, H. Y. (2010). Compatibility of conditionally specified models. *Statistics Probability Letters*, 80(7):670 – 677.

Cox, A. D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. 91:729–737.

Cramer, H. (1999). *Mathematical Methods of Statistics (PMS-9)*. Princeton University Press.

DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. Springer Science & Business Media.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-6:721–741.

Ghosh, I. and Nadarajah, S. (2016). An alternative approach for compatibility of two discrete conditional distributions. *Commun. Stat. - Theory Methods*, 45:4416–4432.

Ghosh, I. and Nadarajah, S. (2017). On the construction of a joint distribution given two discrete conditionals. *Stud. Sci. Math. Hungarica*, 54:178–204.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31(4):1208–1211.

Gourieroux, C., Laffont, J.-J., and Monfort, A. (1979). Coherency conditions in simultaneous linear equation models with endogenous switching regimes.

Grant, M. (2020). Understanding Buy, Sell, and Hold Ratings of Stock Analysts.

Hauser, G. (1993). Galton and the Study of Fingerprints. In *Sir Fr. Galton, FRS*, pages 144–157. Palgrave Macmillan UK, London.

Hosseini, M. (2017). *Statistical Modeling of Self and Proxy Responses with Gerontological Applications*. PhD thesis, Univeristy of Marylad Baltimore County.

Kuo, K.-L. and Wang, Y. (2017). Pseudo-Gibbs sampler for discrete conditional distributions. *Ann. Inst. Stat. Math.*

Kuo, K.-L. and Wang, Y. J. (2018). Simulating conditionally specified models. *Journal of Multivariate Analysis*, 167:171–180.

Kuo, K.-L. and Wang, Y. J. (2019). Pseudo-gibbs sampler for discrete conditional distributions. *Annals of the Institute of Statistical Mathematics*, 71(1):93–105.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemp. Math.*, 80:221–239.

Mardia, K. V., Hughes, G., Taylor, C. C., and Singh, H. (2008). A multivariate von mises distribution with applications to bioinformatics. *Canadian Journal of Statistics*, 36(1):99–109.

Molenberghs, G. and Verbeke, G. (2006). *Models for discrete longitudinal data*. Springer Science & Business Media.

Mosteller, F. (1968). Association and Estimation in Contingency Tables. *Journal of American*, 63:1–28.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley.

Rizzo, M. L. (2019). *Statistical computing with R*. CRC Press.

Sengupta, A. (2004). Generalized variance. *Encyclopedia of statistical sciences*.

Shardell, M., Hicks, G. E., Miller, R. R., Langenberg, P., and Magaziner, J. (2010). Pattern-mixture models for analyzing normal outcome data with proxy respondents. *Stat. Med.*

Snow, A. L., Cook, K. F., Lin, P. S., Morgan, R. O., and Magaziner, J. (2005). Proxies and other external raters: Methodological considerations. *Health Serv. Res.*, 40:1676–1693.

Varin, C. (2008). On composite marginal likelihoods. *Asta advances in statistical analysis*, 92(1):1–28.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Stat. Sin.*, 21(1):5–42.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, pages 471–494.