Composite Empirical Likelihood:
A Derivation of Multiple Non-Parametric Likelihoods

by

Adam Paul Jaeger

(Under the Direction of Nicole A. Lazar)

Abstract

The likelihood function plays a pivotal role in statistical inference; it is adaptable to a wide range of models and the resultant estimators are known to have good properties. However, these results hinge on correct specification of the true data generating mechanism. Many modern problems involve extremely complicated distribution functions, which may be difficult – if not impossible – to express explicitly. This is a serious barrier to the likelihood approach, which requires not only the specification of a distribution, but the correct distribution. Non-parametric methods are one way to avoid the problem of having to specify a particular data generating mechanism, but can be computationally intensive, reducing their accessibility for large data problems. We propose a new approach that combines multiple non-parametric likelihood-type components to build a data-driven approximation of the true function. We build on two alternative likelihood approaches, empirical and composite likelihood, taking advantage of the strengths of each. Specifically, from empirical likelihood we borrow the ability to avoid a parametric specification, and from composite likelihood we gain a decrease in computational load. We will examine the theoretical properties of this new construct, both for purposes of application and to compare properties to other established likelihood methods.

Index words: Estimating equations; Inference; Likelihood; Non-parametric; Robust

Composite Empirical Likelihood:
A Derivation of Multiple Non-Parametric Likelihoods

by

Adam Paul Jaeger

B.A., North Carolina State University, 1996

B.S., University of Georgia, 2009

M.S., University of Georgia, 2011

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2015

Composite Empirical Likelihood:

A Derivation of Multiple Non-Parametric Likelihoods

by

Adam Paul Jaeger

Major Professor: Nicole A. Lazar

Committee: Jeongyoun Ahn
Daniel B. Hall
William P. McCormick
Cheolowoo Park

Electronic Version Approved:

Julie A. Coffield
Interim Dean of the Graduate School
The University of Georgia
May 2015

Dedication

To my mother, without whom this would not have been possible.

Acknowledgments

I am truly lucky to have Nicole Lazar as my adviser. She not only guided me through the dissertation, but throughout all my years in the UGA Statistics Department, and I anticipate that guidance continuing into the future. Her advice, patience and humor have been much appreciated and will never be forgotten.

Were it not for Lynne Seymour and her guidance I would not have had half the excellent professional experiences I participated in. I am also thankful for her support of me over the years, which despite my promise to pay it forward, is a debt I doubt I will ever be able to break even on.

I would like to acknowledge the financial support provided by Dr. Albert Vexler at the State University of New York, Buffalo. This not only allowed me to focus on building the methodology for this dissertation but also allowed me the opportunity to present it.

Thank you to Jeongyoun Ahn, Dan Hall, Bill McCormick, Cheolwoo Park and Lily Wang for the time and energy you gave serving on my committee. Your comments and suggestions were most welcome and helped build this dissertation.

A special thanks to Maurice Hendon from the Department of Mathematics at University of Georgia. His constant willingness to help me walk through any math problem has taught me to appreciate the nuance of mathematics that ultimately made it enjoyable.

To Kim Love-Myers and Chris Franklin, thank you for helping me develop as a professional and reminding me that there is more to academics than publishing.

I also need to thank many of the employees at Walkers Coffee and Pub for always having a good cup of coffee and a place to sit ready for me. The relaxed environment has allowed me to complete many homework assignments, proofs, simulations and much of the writing of this dissertation.

To all the fellow students I have befriended (too many to list but you know who you are), thanks for the memories.

Finally this dissertation and my professional career would never have taken the shape they have without T.N. Sriram.

Table of Contents

Acknowledgments **v**

List of Tables **viii**

List of Figures **ix**

1 Overview of Likelihoods **1**

1.1 Introduction 1

1.2 Fisher Likelihood 3

1.3 Composite Likelihood 5

1.4 Empirical Likelihood 9

1.5 Combining Composite and Empirical Methodology 13

2 Two Component Composite Empirical Likelihood **15**

2.1 Introduction 15

2.2 Main Results 19

2.3 Estimation of a Weighted Chi Square Random Variable 36

2.4 Simulations 39

2.5 Discussion 47

3 Alternate Estimating Equation Forms for Dependent Data **49**

3.1 Introduction 49

3.2 Distribution of Test Statistic with Nuisance Parameter 51

3.3 Conditional and Marginal Estimating Equations 59

3.4 Conditional and Conditional Estimating Equations 63

3.5 Numerical Studies 65

3.6 Discussion 69

4 Composite Empirical Likelihood **72**

4.1 Introduction 72

4.2 Composite Empirical Likelihood 72

4.3 Numerical Studies Using Canadian Climate Data 83

4.4 Discussion 91

5 Summary **92**

5.1 Introduction 92

5.2 Alternate Constructions of the Composite Empirical Likelihood 92

5.3 Estimating Equations and Optimization 94

5.4 Composite Multiple Hypothesis Testing 95

5.5 Conclusion 97

Appendix **99**

A Stochastic Order Notation and Properties 99

B Matrix Inverse 100

References **101**

List of Tables

2.1 Mean and variance of parameter estimates and cumulative distribution of T at $\alpha = 0.10, 0.05, 0.01$ when data are bivariate normal. 40

2.2 Mean and variance of parameter estimates and cumulative distribution of T at $\alpha = 0.10, 0.05, 0.01$ when data are bivariate chi square. 40

2.3 Mean and variance of parameter estimates and cumulative distribution of T at $\alpha = 0.10, 0.05, 0.01$ when data are bivariate uniform. 41

2.4 Comparison of L_{CE}, L_{E1} and L_{E2} using bivariate normal with 500 replicates. 43

2.5 Comparison of L_{CE}, L_{E1} and L_{E2} using bivariate chi square with 500 replicates. . . 44

2.6 Comparison of L_{CE}, L_{E1} and L_{E2} using bivariate uniform with 500 replicates. . . . 45

4.1 95% confidence intervals of τ for each variable from the province of Alberta for June, July and August. 85

4.2 95% confidence intervals of τ for each variable from the province of Manitoba for June, July and August. 85

4.3 95% confidence intervals of τ for each variable from the province of Saskatchewan for June, July and August. 85

4.4 95% confidence intervals of τ for each variable from all three Canadian provinces for June, July and August. 87

4.5 Estimate of h for each variable from all three Canadian provinces for June, July and August using 2005 data. 90

List of Figures

2.1 Distribution of computation times of L_{CE}, L_{E1} and L_{E2} using bivariate chi-square with 500 replicates. 47

3.1 Comparison of empirical likelihood and composite empirical likelihood for behavioral and plumage characteristics of hybrid ducks data. The asterisk (in red) denotes the sample mean of the two variables, and the larger sized point at $14, 11$ denotes that there were two observations with those values. 50

3.2 Coverage of 100 confidence regions from bivariate normal with $\sigma_x = 1$, $\sigma_y = 4$, and $\rho = 0.50$. The center point is the true mean $[0, 0]$ and the ellipse is the true 95% confidence region of the mean. 66

3.3 Coverage of 100 confidence regions from bivariate normal with $\sigma_x = 4$, $\sigma_y = 1$, and $\rho = 0.50$. The center point is the true mean $[0, 0]$ and the ellipse is the true 95% confidence region of the mean. 67

3.4 Coverage of 100 confidence regions from bivariate normal with $\sigma_x = 4$, $\sigma_y = 4$, and $\rho = 0.50$. The center point is the true mean $[0, 0]$ and the ellipse is the true 95% confidence region of the mean. 67

3.5 Coverage of 100 confidence regions from bivariate normal with $\sigma_x = 2$, $\sigma_y = 2$, and $\rho = 0.10$. The center point is the true mean $[0, 0]$ and the ellipse is the true 95% confidence region of the mean. 68

3.6 Coverage of 100 confidence regions from bivariate normal with $\sigma_x = 2$, $\sigma_y = 2$, and $\rho = 0.50$. The center point is the true mean $[0, 0]$ and the ellipse is the true 95% confidence region of the mean. 68

3.7 Coverage of 100 confidence regions from bivariate normal with $\sigma_x = 2$, $\sigma_y = 2$, and $\rho = 0.90$. The center point is the true mean $[0, 0]$ and the ellipse is the true 95% confidence region of the mean. 69

3.8 Comparison of empirical and composite empirical likelihood methods for behavioral and plumage characteristics of hybrid ducks data. The asterisk (in red) denotes the sample mean of the two variables, and the larger sized point at $14, 11$ denotes that there were two observations with those values. 70

4.1 Locations of recorded Canadian weather data for climate change analysis. 84

4.2 Locations of recorded Canadian weather data for bandwidth parameter estimation. . 89

Chapter 1

Overview of Likelihoods

1.1 Introduction

The estimation of parameters that describe the characteristics of a model is a common focus of many statistical techniques. In simple cases the distribution of the estimator is easily derived, which make determining a range of likely parameter values straightforward using the exact distribution of the estimator. Sometimes the asymptotic distribution of the estimator is used, which can result in less reliable intervals and p values. In many cases it is impossible to determine any exact or recognizable distribution for the statistic, requiring an alternate approach for inference. An extremely powerful and flexible approach that does not necessitate knowledge of the distribution of the statistic is the use of a likelihood function, which is used to create a likelihood ratio (LR). The estimators generated by maximizing the likelihood function have the property of being asymptotically normal, and furthermore Wilks (1938) proved that given certain regularity conditions $-2\log LR$ asymptotically converges to a chi square distribution, allowing for hypothesis testing.

Despite the strengths and desirable properties of likelihoods, likelihood ratios and the likelihood estimators this methodology is not without limitations. First, the precision of the point estimates and the confidence intervals relies heavily on the correct specification of the distribution. Second, in complex scenarios where data are probabilistically dependent on each other or dependent on data values that are not observed, the likelihood can be difficult or impossible to define correctly. Since the true data distribution cannot be stated with total certainty, the conclusions of an analysis based on the likelihood function is potentially suspect.

One likelihood method that does not require full expression of the likelihood function is the composite likelihood (see Lindsay, 1988, for details) which divides the likelihood into smaller likelihood components that can be properly expressed, and then appropriately combines these likelihood components. Each piece of a composite likelihood is a proper likelihood itself, therefore the composite

likelihood maintains several properties present in the ordinary likelihood function, and in some cases the composite likelihood is asymptotically equivalent to the full likelihood. The performance of a composite likelihood relies on the correct specification of each likelihood component, so there is still the issue of distribution specification.

Non-parametric methods avoid the issue of proper specification of the distribution by either empirical estimation from the data or mapping the data values in such a way that the behavior of the transformed data is known. Despite the advantage of non-parametric approaches being distribution free they have their own drawbacks. First, they will not produce more efficient results than their parametric counterparts when the data distribution is correctly specified. Second, non-parametric methods are computationally intensive (especially with large amounts of data); some examples are non-parametric linear regression, monotonic regression (Conover, 1999, Chapter 5) and kernel smoothers (Hastie et al., 2009, Chapter 6).

Empirical likelihood (Owen, 1988) creates a likelihood using the properties of the empirical cumulative distribution function, which consequently eliminates the need to specify a distribution. This data driven likelihood maintains the asymptotic properties of the parametric likelihood (Owen, 2001). Empirical likelihood suffers, however, by lacking a closed form solution in most cases, causing the computational time involved with large multivariate data sets to be quite lengthy. With the advent of parallel computing, multiple computations can be carried out simultaneously decreasing computational time, but current empirical likelihood forms cannot be split in order to take advantage of parallel computing.

All likelihood methods have a very desirable characteristic; the construction is generalized allowing likelihoods to be used in almost every statistical modeling application. The parametric likelihoods express the model through the probability density functions, and empirical likelihoods define the model using moment estimators. Despite the development of composite likelihood and empirical likelihood we are still forced to choose between computational tractability and incorrect inference under model misspecification.

There are many modern problems that require a framework that relaxes or eliminates the need for distributional assumptions such as ANOVA, time series, spatial correlation, mixed modeling, longitudinal studies and estimation of certain parameters (such as quantiles). Since empirical likelihood inherits the asymptotic properties of parametric likelihood, it stands to reason that a

multiple component empirical likelihood would also exhibit the same behavior as the parametric empirical counterpart. This approach would introduce a new methodology that maintains the flexibility of likelihood, does not require a distributional assumption and allows for reduction of the computational load.

Our proposed method, the composite empirical likelihood, replaces the parametric likelihood components that comprise a composite likelihood with empirical likelihoods. In the same fashion as composite likelihoods the empirical pieces are themselves proper empirical likelihoods. Since parametric composite likelihoods share many of the asymptotic characteristics of their full likelihood equivalents, we explore if similar results hold in order to increase the understanding of likelihood theory in general. Our goal is to create a robust and flexible method of parameter estimation that is applicable to a large variety of problems.

The remainder of the chapter will introduce the relevant definitions and theoretical properties of likelihoods, composite likelihoods and empirical likelihoods.

1.2 Fisher Likelihood

In order to correctly specify the likelihood function, the exact probabilistic distribution of the data must be known. This likelihood form follows from Fisher (1922), hence we will refer to a correctly specified likelihood as the Fisher likelihood. We give a formal definition along with several relevant results (Casella and Berger, 2002, Chapters 6-8, unless otherwise referenced).

Let $\mathbf{z}_1, \ldots, \mathbf{z}_n$ be a k-variate sample from some distribution F_0 with $p \times 1$ parameter θ. Define the probability density function $f_0(\mathbf{z}; \theta)$. The Fisher likelihood function given the data is

$$\mathcal{L}_0(\theta) = \mathcal{L}_0(\theta; \mathbf{z}_1, \ldots, \mathbf{z}_n) = f_0(\mathbf{z}_1, \ldots, \mathbf{z}_n; \theta),$$

and the maximum likelihood estimator (MLE) is

$$\hat{\theta}_F = \arg\max_{\theta} \mathcal{L}_0(\theta).$$

The MLE is not always an unbiased estimator of θ. The MLE is transformation invariant, so for any function h the MLE of $h(\theta)$ is $h(\hat{\theta}_F)$.

The variance of the derivative of the likelihood function describes the information of the MLE. First define the log of the likelihood function as

$$\ell_0(\theta) = \log \mathcal{L}_0(\theta)$$

and we have

$$E\left(\partial \ell_0(\theta)/\partial\theta\right) = 0,$$

so the Fisher information matrix is

$$\begin{aligned} I(\theta) &= E(\partial \ell_0(\theta)/\partial\theta)^2 \\ &= -E(\partial^2 \ell_0(\theta)/\partial\theta\partial\theta^{\mathrm{T}}) \end{aligned}$$

with the equality by the second Bartlett identity (Ferguson, 1996, page 120). Generally the second derivative of the log likelihood is easier to compute than the square of the first derivative, making this equality especially useful.

The maximum likelihood estimator has the following asymptotic distribution;

$$\sqrt{n}(\hat{\theta}_F - \theta) \longrightarrow N(0, I(\theta)^{-1})$$

as $n \longrightarrow \infty$ given suitable regularity conditions. Additionally by Cramér-Rao Lower Bound the smallest variance any unbiased estimator of θ will have is $I(\theta)^{-1}$, making any unbiased MLE a minimum variance unbiased estimator.

We could use the asymptotic normality of $\hat{\theta}$ to derive confidence intervals and perform tests of hypotheses for θ, but a more common approach is to use the likelihood ratio test statistic to compute hypothesis tests and create confidence regions. Additionally the likelihood ratio test allows for multiple composite hypothesis testing. Define $\theta = [\phi^{\mathrm{T}}, \nu^{\mathrm{T}}]^{\mathrm{T}}$, where ϕ is $q \times 1$ and ν is $(p-q) \times 1$; ν is the set of nuisance parameters which we are not interested in for inferential purposes. Denote

$\hat{\nu}_F(\phi)$ as the value of ν that maximizes $\ell_0(\phi, \nu)$ for a fixed value of ϕ. The test statistic is

$$T_F = -2\left\{\ell_0(\phi_0, \hat{\nu}_F(\phi_0)) - \ell_0(\hat{\phi}_F, \hat{\nu}_F)\right\},$$

and given suitable regularity conditions (Casella and Berger, 2002, page 516),

$$T_F \longrightarrow \chi^2(q)$$

as $n \longrightarrow \infty$.

1.3 Composite Likelihood

A common challenge in creating a likelihood function is the ability to fully define the joint distribution of the data. There are many models where the conditional or marginal distributions can be expressed, but the joint distribution is too complex to properly define. Sometimes the product of these conditional or marginal pieces equals the true distribution or asymptotically approaches the true distribution. In cases when they are not equivalent these product forms can give enough information concerning the parameters of interest to be useful.

This approach of rewriting the true likelihood as a product of conditional and marginal likelihoods has been referred to by names such as pseudo-likelihood, approximate likelihood, split-data likelihoods and quasi-likelihoods (Varin et al., 2011). These terms are not very informative, and in the case of quasi-likelihood overlap with established alternatives such as the quasi-likelihood developed by McCullagh (1983).

Lindsay (1988) formalizes the method of rewriting a likelihood as products of multiple likelihoods by defining a composite likelihood as the weighted product of marginal or conditional likelihoods. The definition allows for a composite likelihood to be a mixture of conditional and marginal likelihood, but generally composite likelihoods are divided into the conditional version and the marginal version without any overlap. Building a likelihood that does not contain all the information may appear inefficient, but in many cases is as informative as the true likelihood (Lindsay, 1988).

Composite likelihoods can express a portion of the model or parameters (implying some aspects of the model are ignored) and result in the parameter estimators being consistent despite

the likelihood itself being incomplete. Composite likelihood methods also inherit several properties of Fisher likelihood. The Kullback-Leibler information inequality holds, so maximizing the composite likelihood will lead to a consistent estimation method, and since each component is a Fisher likelihood itself, the composite score is an unbiased estimating function for the information (Lindsay, 1988). Cox (1975) shows the large sample properties of maximum likelihood estimators apply when working with conditional likelihood if the decomposition of the nuisance parameters is done properly. Arnold and Strauss (1991) show that under certain regulatory conditions marginal likelihoods are consistent and asymptotically normal (also see Cox and Reid, 2004).

Two common assumptions to make about collected data are randomization and independence. The following definition, although being identified with the Fisher likelihood, is actually an example of a composite likelihood. Let z be a univariate sample with $z_i \overset{i.i.d}{\sim} f_0(z_i;\theta)$; we express the likelihood as

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f_0(z_i;\theta).$$

This likelihood is written as the product of distribution functions, but we can also view this as the product of n likelihood functions; let $\mathcal{L}_i(\theta) = f_0(z_i;\theta)$, then

$$\mathcal{L}_C(\theta) = \prod_{i=1}^{n} \mathcal{L}_i(\theta).$$

Since we are assuming independence the product of the marginal likelihood functions is the true probability function. Additionally the composite likelihood in this example is equivalent to the Fisher likelihood.

To demonstrate another example of a composite likelihood let X, Y have a joint distribution $f_{XY}(x, y;\theta)$. Assume that the conditional distributions $f_{X|Y}$ and $f_{Y|X}$ exist, and the marginal distributions f_X and f_Y exist. In addition to the Fisher likelihood we can construct the following composite likelihoods:

$$\mathcal{L}_{CM} = f_X(\theta;x) f_{Y|X}(\theta;y,x) = \mathcal{L}_X(\theta)\mathcal{L}_{Y|X}(\theta),$$
$$\mathcal{L}_{MM} = f_X(\theta;x) f_Y(\theta;y) = \mathcal{L}_X(\theta)\mathcal{L}_Y(\theta),$$

$$\mathcal{L}_{CC} = f_{X|Y}(\theta; x, y) f_{Y|X}(\theta; y, x) = \mathcal{L}_{X|Y}(\theta)\mathcal{L}_{Y|X}(\theta).$$

The composite likelihoods are a product of proper likelihoods derived from the data. $\mathcal{L}_{CM}$ is equal to the true likelihood but in a conditional marginal form. $\mathcal{L}_{MM}$ and $\mathcal{L}_{CC}$ are only equivalent to the true likelihood if X and Y are independent. The theoretical properties and justifications for specific forms are explored in Lindsay (1988).

The composite likelihood is formally defined as follows; let $\mathbf{z}_1, \ldots, \mathbf{z}_n$ be a k-variate sample from some distribution F_0. Assume for a given subset of the data j we can define conditional or marginal events for which the likelihood $\mathcal{L}_j(\theta)$ can be written for $j = 1, \ldots, J$ with the requirement that $\mathcal{L}_j(\theta)$ for all j must be Fisher likelihoods. The composite likelihood is

$$\mathcal{L}_C(\theta) = \prod_{j=1}^{J} \{\mathcal{L}_j(\theta)\}^{w_j}$$

where w_j is a predetermined weight. If w_j is equal for all j the weights can be ignored for purposes of maximization. The Fisher likelihood can also be viewed as a specific case of $\mathcal{L}_C(\theta)$ where $J = 1$, $w_j = 1$ and $\mathcal{L}_j(\theta) = \mathcal{L}_0(\theta)$.

An early example of what would now be classified as a composite likelihood came from the application of conditional likelihood to spatial data (Besag, 1974, 1975) where the term pseudo-likelihood is used. Besag's pseudo-likelihood consists of each likelihood piece being defined as the probability of a single point given the behavior of the other points. Modern uses of composite conditional likelihoods have expanded on Besag's work by conditioning on functions of the nearest neighbors (Liang, 1987), pooling pairwise conditional densities or full conditional densities (Mardia et al., 2008).

Composite marginal likelihoods in their simplest form ignore the dependencies between all observations. This form is sometimes referred to as the independence likelihood and it only allows for inference on the marginal parameters (Chandler and Bate, 2007). If the variables are independent, then the marginal likelihood is equal to the Fisher likelihood. Marginal likelihood can be extended to symmetric responses using a composite marginal likelihood based on pairwise differences. Instead of using each univariate likelihood, each bivariate pair is used to construct the marginal likelihood (Varin et al., 2011).

There are several features pertaining to composite likelihood that mirror Fisher likelihood (Lindsay, 1988). These results arise from the fact that a composite likelihood is composed of Fisher likelihoods. We define the log composite likelihood as

$$\begin{aligned}\ell_C(\theta) &= \log \mathcal{L}_C(\theta) \\ &= \log \prod_{j=1}^{J} \mathcal{L}_i(\theta) \\ &= \sum_{j=1}^{J} \log \mathcal{L}_i(\theta) \\ &= \sum_{j=1}^{J} \ell_j(\theta).\end{aligned}$$

The Kullback-Leibler information inequality holds for the log composite likelihood by noting that it holds for each component log likelihood since each component likelihood is itself a Fisher likelihood. Specifically from Lindsay (1988) we have

$$E_{\theta_0}(\ell_j(\theta)) \leq E_{\theta_0}(\ell_j(\theta_0)) \Rightarrow \quad \sup_{\theta} E_{\theta_0}(\ell_C(\theta)) = E_{\theta_0}(\ell_C(\theta_0)).$$

The maximum composite likelihood estimator

$$\hat{\theta}_C = \arg\max_{\theta} \mathcal{L}_C(\theta)$$

is defined in the exact same manner as the maximum likelihood estimator, and furthermore is a consistent estimator of θ given mild assumptions about the convergence of $\ell_C(\theta)$ (Lindsay, 1988).

In cases where $\mathcal{L}_C \neq \mathcal{L}_0$ the composite likelihood can be viewed as an incorrect specification of the likelihood by observing that we are assuming independence between each likelihood component $\mathcal{L}_j$ comprising the composite likelihood. As a result the second Bartlett identity is not guaranteed to hold, so it is necessary to distinguish between the sensitivity matrix

$$H(\theta) = -E\left(\partial^2 \ell_C(\theta)/\partial\theta\theta^{\mathrm{T}}\right),$$

and the variability matrix

$$J(\theta) = E\left(\partial \ell_C(\theta)/\partial\theta\right)^2.$$

The Fisher information matrix is replaced by the Godambe information matrix (Godambe, 1960)

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

If $\mathcal{L}_C$ is a true likelihood function then $G = H = I$ for all θ (Varin et al., 2011).

Under regularity, the maximum composite likelihood estimator is also asymptotically normal, specifically

$$\sqrt{n}(\hat{\theta}_C - \theta) \longrightarrow N(0, G(\theta)^{-1}).$$

The test statistic using the composite likelihood is the same form used in Fisher likelihood, but the asymptotic distribution is not a normal chi square. Define $\theta = [\phi^{\mathrm{T}}, \nu^{\mathrm{T}}]^{\mathrm{T}}$ where ϕ is $q \times 1$, ν is the $(p-q) \times 1$ set of nuisance parameters. The composite likelihood ratio statistic is

$$T_C = -2\left\{\ell_C(\phi_0, \hat{\nu}_C(\phi_0)) - \ell_C(\hat{\phi}_C, \hat{\nu}_C)\right\}.$$

where $\hat{\nu}_C(\phi_0)$ maximizes $\ell_C(\phi, \nu)$ with respect to ϕ_0. Under $H_0 : \phi = \phi_0$

$$T_C \longrightarrow \sum_{i=1}^{l} \lambda_i \chi^2(1)$$

as $n \longrightarrow \infty$. $\chi^2(1)$ for $i = 1, \ldots, l$ are independent chi square random variables with one degree of freedom, and λ_i are the non-zero eigenvalues of $(H^{\phi\phi})^{-1}G^{\phi\phi}$. The superscripts denote the submatrices of G and H pertaining to ϕ.

1.4 Empirical Likelihood

One of the earliest examples of a non-parametric likelihood appears in Thomas and Grunkemeier (1975), where a non-parametric likelihood is constructed to compute a confidence interval

for censored data. The authors show that their non-parametric likelihood ratio has an asymptotic chi square distribution, which is the same result shown by Wilks (1938) using a parametric likelihood.

Owen (1988) formalizes the framework of Thomas and Grunkemeier (1975) by developing an empirical likelihood ratio test for a single functional with emphasis on the first moment, and proves that the likelihood ratio test statistic is asymptotically chi square. The theorem governing the asymptotic convergence of the empirical likelihood for the mean is readily extended to single M estimator functionals (given certain regularity conditions of the M estimator). This extends empirical likelihood methods to quantiles and Huber's location M estimators (Huber, 1964). Using any functional that has a Fréchet derivative with an empirical likelihood results in the same asymptotic behavior seen using M estimators. A limitation of empirical likelihood is the lack of a solution when the parameter falls outside of the convex hull of the data.

Empirical likelihood methods have been extended to bivariate means problems (Owen, 1990), regression models, correlation models, ANOVA and variance modeling (Owen, 1991, 2001), generalized linear modeling (Kolaczyk, 1994), the entire class of projection pursuit models (Owen, 1992), time series modeling (Owen, 2001; Kitamura, 1997), incorporating information from multiple moments for a parameter, two sample problems with a common mean, probability measure, incomplete information problems (Qin and Lawless, 1994), ratios of parameters and logistic regression (Qin and Lawless, 1995), partially linear models (Shi and Lau, 2000), missing response problems (Qin and Zhang, 2007), finite population inference (Chen and Qin, 1993), and Bayesian settings (Lazar, 2003; Grendar and Judge, 2010).

The following outlines the derivation of an empirical likelihood. Let $\mathbf{z}_1, \ldots, \mathbf{z}_n$ be k-variate independent identically distributed observations from some distribution F_0. The empirical likelihood function is

$$L(F_0) = \prod_{i=1}^{n} dF_0(\mathbf{z}_i) = \prod_{i=1}^{n} Pr(\mathbf{Z} = \mathbf{z}_i) = \prod_{i=1}^{n} p_i,$$

which is maximized by the empirical distribution function

$$L(F_n) = \prod_{i=1}^{n} n^{-1},$$

so the empirical likelihood ratio function $R(F_0) = L(F_0)/L(F_n)$ can be written as

$$R(F_0) = \prod_{i=1}^{n} np_i.$$

The next step is estimation of a $p \times 1$ parameter θ. In order to perform inference on θ constraints in the form of $r \geq p$ unbiased estimating equations $g_j(\mathbf{z}_i, \theta)$ for $j = 1, \ldots, r$. The profile empirical likelihood ratio function is

$$R_E(\theta) = \sup_p \left(\prod_{i=1}^{n} np_i \middle| p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i g(\mathbf{z}_i, \theta) = 0 \right). \tag{1.1}$$

Provided that 0 is inside the convex hull of the points $g(\mathbf{z}_1, \theta), \ldots, g(\mathbf{z}_n, \theta)$ a unique value of Equation 1.1 exists (Owen, 1988). By definition $R_E(\theta) = 0$ for all θ not inside the convex hull.

As an example, if the parameter of interest is the mean μ for univariate data, the profile empirical likelihood ratio function is

$$R_E(\mu) = \sup_p \left(\prod_{i=1}^{n} np_i \middle| p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i (z_i - \mu) = 0 \right).$$

For simple linear regression, the estimating equations are based on the ordinary least squares solution. Let $E(Z) = \beta_0 + \beta_1 x$, then the normal equations are $X^{\mathrm{T}}(Z - X\beta) = 0$. The profile empirical likelihood ratio function for β_0, β_1 is

$$R_E(\beta_0, \beta_1) = \sup_p \left(\prod_{i=1}^{n} np_i \middle| p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i z_i(\beta) = \sum_{i=1}^{n} p_i x_i z_i(\beta) = 0 \right)$$

where

$$z_i(\beta) = z_i - \beta_0 - \beta_1 x_i.$$

An explicit expression of $R_E(\theta)$ can be derived using Lagrange multipliers; subject to the constraints that $p_i \geq 0$, $\sum_{i=1}^{n} p = 1$ and $\sum_{i=1}^{n} p_i g(\mathbf{z}_i, \theta) = 0$ the maximum value of $R_E(\theta)$ is attained when

$$p_i = n^{-1} \left\{ 1 + t^{\mathrm{T}} g(\mathbf{z}_i, \theta) \right\}^{-1}$$

where $t = t(\theta)$ is a $r \times 1$ vector based on the value of θ which satisfies

$$\sum_{i=1}^{n} \left\{1 + t^{\mathrm{T}} g(\mathbf{z}_i, \theta)\right\}^{-1} g(\mathbf{z}_i, \theta) = 0.$$

We also have an analogue to the maximum likelihood estimator. The maximum empirical likelihood estimator (MELE) is

$$\hat{\theta}_E = \arg\max_{\theta} R_E(\theta),$$

and the MELE is also asymptotically normal. We have

$$\sqrt{n}(\hat{\theta}_E - \theta_0) \longrightarrow N(0, W^{-1})$$

as $n \longrightarrow \infty$, where

$$W = E\left(\frac{\partial g}{\partial \theta}\right)^{\mathrm{T}} \left\{E(gg^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g}{\partial \theta}\right)$$

with $g = g(\mathbf{Z}, \theta_0)$.

Like parametric likelihoods, it is generally easier to work with the logarithm of the ratio function. We denote the logarithm of the empirical likelihood ratio function as

$$\ell_E(\theta) = \log\left(\prod_{i=1}^{n} np_i\right) = \sum_{i=1}^{n} \log np_i = -\sum_{i=1}^{n} \left(\log(n) + \log\left\{1 + t^{\mathrm{T}} g(\mathbf{z}_i, \theta)\right\}\right).$$

Instead of using the asymptotic distribution of $\hat{\theta}_E$ to create confidence intervals and perform tests of hypotheses, we rely on the asymptotic distribution of the likelihood ratio test statistic. Define $\theta = [\phi^{\mathrm{T}}, \nu^{\mathrm{T}}]^T$ where ϕ is a $q \times 1$ vector and ν is $(p-q) \times 1$ where ν is the set of nuisance parameters. For $H_0 : \phi = \phi_0$ the likelihood ratio statistic is

$$T_E(\phi) = -2\left\{\ell_E(\phi_0, \hat{\nu}(\phi_0)) - \ell_E(\hat{\phi}, \hat{\nu})\right\}$$

where $\hat{\nu}(\phi_0)$ maximizes $\ell_E(\phi, \nu)$ with respect to ϕ_0. Under H_0, $T_E(\phi) \longrightarrow \chi^2(q)$ as $n \longrightarrow \infty$.

Several modifications to the basic empirical likelihood are proposed in Owen (1991): Constrained empirical likelihood which allows for the creation of a conditional empirical likelihood, Euclidian likelihood which allows the confidence region to extend beyond the convex hull, and triangular array empirical likelihood which relaxes the assumption of the data being identically distributed. A variation known as the pseudo empirical likelihood allows for an additional set of weights to adjust how much influence any single data point has on the empirical likelihood function (Chen and Sitter, 1999; Wu and Rao, 2006). Kitamura (1997) proposes a blockwise empirical likelihood for weakly dependent processes by adding additional constraints on the p vector based on data blocks. Chen et al. (2008) develop an adjusted empirical likelihood which allows for inference on parameters outside of the convex hull of the data.

There have also been developments in combining empirical and parametric likelihood. This variant, the semi-empirical likelihood (Qin, 1994), is designed for problems with two independent samples where the distribution of one sample is known and the other is not specified. Qin (2000) applies this method to economic variables where the conditional distribution is parametric and the marginal is unknown.

1.5 Combining Composite and Empirical Methodology

Desirable features of any statistical method are robustness, flexibility and computational stability. A defining property of any likelihood method is model flexibility; however there is usually a trade off between robustness due to incomplete (or improper) distribution specification and computational efficiency. We proposed a composite empirical likelihood which combines the piecewise approach of the composite likelihood and the non-parametric property of the empirical likelihood. This construct does not require any distributional assumptions and maintains a high level of computational efficiency.

The following illustrative examples show several applications of composite empirical likelihood and define the relationship between the data and parameters of interest.

Example 1.1 Let $x_1, \ldots, x_{n_x}$ and $y_1, \ldots, y_{n_y}$ be two (possibly dependent) samples with a common parameter. Two empirical likelihoods are multiplied together to create the composite empirical likelihood which is then used to compute a confidence interval for the common parameter.

Example 1.2 Let $x_1, \ldots, x_{n_x}$ and $y_1, \ldots, y_{n_y}$ be two dependent samples; we are interested in estimating the marginal mean of each variable. We set up a marginal likelihood in terms of X and a conditional likelihood in terms of $Y|X$. This gives additional flexibility in describing the functional relationship between the two variables since we specify the functional form of $E(Y|X)$.

Example 1.3 Let $\mathbf{z}_1, \ldots, \mathbf{z}_n$ be a random sample with large n. To maintain computational precision and decrease computational time the sample can be split into multiple subsets. The data to define each likelihood piece can be selected using exhaustive sampling without replacement, sub sampling or bootstrapping. In this instance the composite empirical likelihood has a common set of parameters shared between all likelihood pieces.

Example 1.4 Let $\mathbf{z}$ be the set of observations from space time data with t time points and s locations. Using a similar approach to Besag (1974) each likelihood piece is expressed in terms of a time series with additional conditioning based on the surrounding neighbors, i.e. $\mathbf{z}_{\cdot j}|\mathbf{z}_{\cdot(-j)}$.

The remainder of the dissertation will proceed as follows. We explore the form and asymptotic properties of a two component composite empirical likelihood in Chapter 2. We compare the behavior of the estimator and the test statistic of the composite empirical likelihood with empirical likelihood by numerical simulations using the setup presented in Example 1.1. Chapter 3 examines defining the estimating equations in terms of marginal and conditional relationships in order to form confidence regions. Example 1.2 is used to create appropriate confidence regions which will be numerically compared to the empirical likelihood counterparts. Chapter 4 expands the two component composite empirical likelihood results from Chapters 2 and 3 to a multiple likelihood component form. We show examples with similar forms to those presented in Examples 1.3 and 1.4 using real data.

Chapter 2

Two Component Composite Empirical Likelihood

2.1 Introduction

To develop the framework for the composite empirical likelihood we start with a two component version. We define the construction and establish the asymptotic behavior of the parameter estimator and test statistic. These results are compared to the theoretical results known for other likelihood methods. We show that the two component composite empirical likelihood inherits the same asymptotic behavior as the likelihoods described in Chapter 1 given regularity conditions, suggesting that the composite empirical form is the general case of a likelihood function with other variations being special cases. We also examine how composite empirical likelihood performs in terms of accuracy and computational efficiency to empirical likelihood.

Let $X \in \mathbb{R}^{d_x}$ and $Y \in \mathbb{R}^{d_y}$ be random variables from some joint distribution f_{xy}. We assume the marginal distributions f_x and f_y exist. Let x have a sample size of n_x and y have a sample size of n_y. We also assume that (x_i, y_i) pairs are *i.i.d.* samples; however we do not assume independence of X and Y and it is possible for any given i we do not have a matched pair of x and y.

We define the composite marginal empirical likelihood function as

$$L_C(F_{xy}) = L(F_x)L(F_y) = \prod_{i=1}^{n_x} dF_x(x_i) \prod_{i=1}^{n_y} dF_y(y_i) = \prod_{i=1}^{n_x} u_i \prod_{i=1}^{n_y} v_i \tag{2.1}$$

where $dF_x(x_i) = P(X = x_i)$ and $dF_y(y_i) = P(Y = y_i)$. If no additional constraints are imposed Equation 2.1 is maximized when $u_i = 1/n_x$ and $v_i = 1/n_y$ for all i. Using these weights the composite empirical cumulative distribution function (CECDF) is

$$\widehat{F}_{n_x,n_y}(x, y) = \left(\frac{1}{n_x} \sum_{i=1}^{n_x} I(x_i < x) \right) \left(\frac{1}{n_y} \sum_{i=1}^{n_y} I(y_i < y) \right).$$

Note that the CECDF is the product of two empirical cumulative distribution functions, so any correlation between the two variables is not accounted for.

Our goal is to develop a likelihood function to perform inferential tests on parameters, so the CECDF by itself is not sufficient for this purpose. The next step incorporates the parameters in terms of a functional argument. Let θ_x be the set of parameters associated with X and denote p_x as the dimension of θ_x. Let θ_y be the set of parameters associated with Y and denote p_y as the dimension of θ_y. Finally define θ as the set of unique parameters, so the dimension of θ is $p \leq (p_x + p_y)$ with equality if $\theta_x \cap \theta_y = \emptyset$. This setup allows for parameters to be completely unique to each likelihood piece, identical in both likelihood pieces, or a combination of shared and unique parameters. The parameters are incorporated via estimating functions, which in most cases will conform to the general class of M estimates (Huber, 1964). Let g_x be the $r_x \times 1$ set of estimating equations for X and let g_y be the $r_y \times 1$ set of estimating equations for Y. We will further assume $r_x \geq p_x$ and $r_y \geq p_y$. The estimating equations must have the following property:

$$E\{g_x(X, \theta_0)\} = 0,$$
$$E\{g_y(Y, \theta_0)\} = 0.$$

In the same fashion as the parameters, the estimating equations do not have to be identical between the groups.

To demonstrate the connection between the data, parameter space and estimating equations a few basic examples are shown below.

Example 2.1 Let $x_1, \ldots, x_{n_x}$ and $y_1, \ldots, y_{n_y}$ be two *i.i.d.* samples from a common distribution. We are interested in inference on the mean μ so the estimating equations are

$$g_x(x_i, \mu) = x_i - \mu,$$
$$g_y(y_i, \mu) = y_i - \mu$$

and $\theta = \mu$.

Example 2.2 Let $x_1, \ldots, x_{n_x}$ and $y_1, \ldots, y_{n_y}$ be two samples. We assume that $E(X) = \mu_x$, $E(Y) = \mu_y$ and the variances of X and Y are equal. We are interested in inference on the variance.

The estimating equations are

$$g_x(x_i, \mu) = \begin{bmatrix} x_i - \mu_x \\ x_i^2 - \hat{mu}_x^2 - \sigma^2 \end{bmatrix}$$
$$g_y(y_i, \mu) = \begin{bmatrix} y_i - \mu_y \\ y_i^2 - \hat{mu}_y^2 - \sigma^2 \end{bmatrix}$$

where $\hat{\mu}_x^2$ and $\hat{\mu}_y^2$ are the solutions in terms of x and y (see Owen, 2001, for details). In this instance $\theta = [\phi, \nu^{\mathrm{T}}]^{\mathrm{T}} = [\sigma^2, \mu_x, \mu_y]^{\mathrm{T}}$, and $\nu = [\mu_x, \mu_y]^{\mathrm{T}}$ would be the nuisance parameters.

The composite empirical likelihood, like the composite likelihood counterpart, is the product of proper empirical likelihood functions. The two component composite empirical likelihood function is

$$L_{CE}(\theta) = \left\{ \left(\sup_u \prod_{i=1}^{n_x} u_i \left| \sum_{i=1}^{n_x} u_i g_x(x_i, \theta) = 0,\ u_i \geq 0,\ \sum_{i=1}^{n_x} u_i = 1 \right. \right) \times \left(\sup_v \prod_{i=1}^{n_y} v_i \left| \sum_{i=1}^{n_y} v_i g_y(y_i, \theta) = 0,\ v_i \geq 0,\ \sum_{i=1}^{n_y} v_i = 1 \right. \right) \right\}. \tag{2.2}$$

In order to find a solution that maximizes Equation 2.2 we follow Owen (1988, 1990) and Qin and Lawless (1994), except that now there are two sets of weights instead of one.

For a fixed θ there exists a unique maximum of Equation 2.2 provided 0 is inside of the convex hull for all points $g_x(x_i, \theta)$ and $g_y(y_i, \theta)$. Let λ_x, λ_y, $t_x = [t_{x1}, \ldots, t_{x(r_x)}]^{\mathrm{T}}$ and $t_y = [t_{y1}, \ldots, t_{y(r_y)}]^{\mathrm{T}}$ be Lagrange multipliers. Define

$$H = \sum_i^{n_x} \log u_i + \lambda_x \left(1 - \sum_i^{n_x} u_i \right) - n_x t_x^{\mathrm{T}} \sum_i^{n_x} u_i g_x(x_i, \theta) + \sum_i^{n_y} \log v_i + \lambda_y \left(1 - \sum_i^{n_y} v_i \right) - n_y t_y^{\mathrm{T}} \sum_i^{n_y} v_i g_y(y_i, \theta).$$

Taking the derivatives of H with respect to u_i and v_i we get

$$\frac{\partial H}{\partial u_i} = \frac{1}{u_i} - \lambda_x - n_x t_x^{\mathrm{T}} g_x(x_i, \theta) = 0 \tag{2.3}$$
$$\frac{\partial H}{\partial v_i} = \frac{1}{v_i} - \lambda_y - n_y t_y^{\mathrm{T}} g_y(y_i, \theta) = 0. \tag{2.4}$$

Equations 2.3 and 2.4 yield the following solutions:

$$\sum_{i=1}^{n_x} u_i \frac{\partial H}{\partial u_i} = n_x - \lambda_x = 0 \Rightarrow \lambda_x = n_x$$

and

$$\sum_{i=1}^{n_y} v_i \frac{\partial H}{\partial v_i} = n_y - \lambda_y = 0 \Rightarrow \lambda_y = n_y$$

Replacing λ_x and λ_y with n_x and n_y respectively in H yield

$$u_i = n_x^{-1}\{1 + t_x^{\mathrm{T}} g_x(x_i, \theta)\}^{-1} \tag{2.5}$$

$$v_i = n_y^{-1}\{1 + t_y^{\mathrm{T}} g_y(y_i, \theta)\}^{-1} \tag{2.6}$$

with the following restrictions

$$0 = \sum_{i=1}^{n_x} u_i g_x(x_i, \theta) = \frac{1}{n_x} \sum_i \{1 + t_x^{\mathrm{T}} g_x(x_i, \theta)\}^{-1} g_x(x_i, \theta),$$
$$0 = \sum_{i=1}^{n_y} v_i g_y(y_i, \theta) = \frac{1}{n_y} \sum_i \{1 + t_y^{\mathrm{T}} g_y(y_i, \theta)\}^{-1} g_y(y_i, \theta).$$

The values of t_x and t_y are determined based on the value of θ. Since we have the restrictions $0 \leq u_i \leq 1$ and $0 \leq v_i \leq 1$, the values of t_x, t_y and θ must satisfy $1 + t_x^{\mathrm{T}} g_x(x_i, \theta) \geq 1/n_x$ and $1 + t_y^{\mathrm{T}} g_y(y_i, \theta) \geq 1/n_y$ for each i. From this we can show t_x and t_y are both continuous differentiable functions of θ.

Lemma 2.1 *Let* $D_\theta = \left\{t_x : 1 + t_x^{\mathrm{T}} g_x(x_i, \theta) \geq 1/n_x; t_y : 1 + t_y^{\mathrm{T}} g_y(y_i, \theta) \geq 1/n_y\right\}$ *for a fixed* θ. *Then* $t_x = t_x(\theta)$ *and* $t_y = t_y(\theta)$ *are both continuous differentiable functions of* θ.

Proof. The proof follows from Qin and Lawless (1994). We have that

$$\frac{\partial}{\partial t_x}\left(\frac{1}{n_x}\sum_{i=1}^{n_x} \frac{g_x(x_i, \theta)}{1 + t_x^{\mathrm{T}} g_x(x_i, \theta)}\right) = -\frac{1}{n_x}\sum_{i=1}^{n_x} \frac{g_x(x_i, \theta) g_x^{\mathrm{T}}(x_i, \theta)}{\left(1 + t_x^{\mathrm{T}} g_x(x_i, \theta)\right)^2},$$

$$\frac{\partial}{\partial t_y}\left(\frac{1}{n_y}\sum_{i=1}^{n_y} \frac{g_y(y_i, \theta)}{1 + t_y^{\mathrm{T}} g_y(y_i, \theta)}\right) = -\frac{1}{n_y}\sum_{i=1}^{n_y} \frac{g_y(y_i, \theta) g_y^{\mathrm{T}}(y_i, \theta)}{\left(1 + t_y^{\mathrm{T}} g_y(y_i, \theta)\right)^2}$$

are both negative definite for t_x and t_y in D_θ since $\sum_{i=1}^{n_x} g_x(x_i,\theta)g_x^{\mathrm{T}}(x_i,\theta)$ and $\sum_{i=1}^{n_y} g_y(y_i,\theta)g_y^{\mathrm{T}}(y_i,\theta)$ are both positive definite. The conclusion that t_x and t_y are continuous differentiable functions of θ follow from the inverse function theorem. □

Substituting Equations 2.5 and 2.6 into Equation 2.2 yields

$$L_{CE}(\theta) = \prod_{i=1}^{n_x} {n_x}^{-1}\{1 + t_x^{\mathrm{T}}(\theta)g_x(x_i,\theta)\}^{-1} \times \prod_{i=1}^{n_y} {n_y}^{-1}\{1 + t_y^{\mathrm{T}}(\theta)g_y(y_i,\theta)\}^{-1}. \tag{2.7}$$

We define the maximum composite empirical likelihood estimate (MCELE) as the value of θ that maximizes $L_{CE}(\theta)$, specifically

$$\hat{\theta}_{CE} = \arg\max_{\theta} L_{CE}(\theta).$$

Similar to parametric likelihoods, it is easier to work with the logarithm of the ratio function. Taking the negative logarithm of Equation 2.7 and ignoring the n_x^{-1} and n_y^{-1} terms we define the composite empirical log likelihood function for X and Y as

$$\begin{aligned}\ell_{CE}(\theta) &= -\log\left(\prod_{i=1}^{n_x}\{1 + t_x^{\mathrm{T}}(\theta)g_x(x_i,\theta)\}^{-1} \times \prod_{i=1}^{n_y}\{1 + t_y^{\mathrm{T}}(\theta)g_y(y_i,\theta)\}^{-1}\right) \\ &= \sum_{i=1}^{n_x}\log\{1 + t_x^{\mathrm{T}}(\theta)g_x(x_i,\theta)\} + \sum_{i=1}^{n_y}\log\{1 + t_y^{\mathrm{T}}(\theta)g_y(x_i,\theta)\}. \end{aligned} \tag{2.8}$$

2.2 Main Results

We present the first order asymptotic properties of $t_x^{\mathrm{T}}(\hat{\theta}_{CE})$, $t_y^{\mathrm{T}}(\hat{\theta}_{CE})$, and $\hat{\theta}_{CE}$ along with asymptotic distributions of the composite empirical log likelihood ratio statistics. In addition we derive the asymptotic distribution of the MCELE and the test statistic. The proofs borrow results from Qin and Lawless (1994). We will assume that n_x and n_y increase at the same rate so that $n_x/n_y \approx 1$, and we use n to denote the sample size for both X and Y under this assumption.

We start with the following set of assumptions, and then establish the existence of a minimum finite value of $\ell_{CE}(\theta)$.

Assumption 1 Let θ_0 be the true value of θ. Then

(a) $E\{g_x(X,\theta_0)g_x^{\mathrm{T}}(X,\theta_0)\}$ and $E\{g_y(Y,\theta_0)g_y^{\mathrm{T}}(Y,\theta_0)\}$ are positive definite.

(b) $\partial g_x(X,\theta)/\partial\theta$ and $\partial g_y(Y,\theta)/\partial\theta$ are continuous in a neighborhood of the true value θ_0.

(c) $\|\partial g_x(X,\theta)/\partial\theta\|$, $\|g_x(X,\theta)\|^3$, $\|\partial g_y(Y,\theta)/\partial\theta\|$ and $\|g_y(Y,\theta)\|^3$ are all bounded by some integrable function in the same neighborhood of θ_0.

(d) The rank of $E\{\partial g_x(X,\theta)/\partial\theta\}$ is p_x and the rank of $E\{\partial g_y(Y,\theta)/\partial\theta\}$ is p_y.

Lemma 2.2 *Given the conditions from Assumption 1, $\ell_{CE}(\theta)$ attains its minimum value at some point $\hat{\theta}_{CE}$ in the interior of the ball $\|\theta-\theta_0\| \le n^{-1/3}$ with probability 1 as $n \longrightarrow \infty$.*

Furthermore $\hat{\theta}_{CE}$, $\hat{t}_x = t_x(\hat{\theta}_{CE})$, and $\hat{t}_y = t_y(\hat{\theta}_{CE})$ satisfy

$$\begin{aligned} Q_{1x}(\hat{\theta}_{CE},\hat{t}_x) &= 0, \\ Q_{1y}(\hat{\theta}_{CE},\hat{t}_y) &= 0, \\ Q_2(\hat{\theta}_{CE},\hat{t}_x,\hat{t}_y) &= 0, \end{aligned}$$

where

$$\begin{aligned} Q_{1x}(\theta,t_x) &= \frac{1}{n_x}\sum_{i=1}^{n_x}\{1+t_x^{\mathrm{T}}g_x(x_i,\theta)\}^{-1}g_x(x_i,\theta), \\ Q_{1y}(\theta,t_y) &= \frac{1}{n_y}\sum_{i=1}^{n_y}\{1+t_y^{\mathrm{T}}g_y(y_i,\theta)\}^{-1}g_y(y_i,\theta) \end{aligned}$$

and

$$\begin{aligned} Q_2(\theta,t_x,t_y) = &\frac{1}{n_x}\sum_{i=1}^{n_x}\{1+t_x^{\mathrm{T}}g_x(x_i,\theta)\}^{-1}\left(\frac{\partial g_x(x_i,\theta)}{\partial\theta}\right)^{\mathrm{T}}t_x \\ &+\frac{1}{n_y}\sum_{i=1}^{n_y}\{1+t_y^{\mathrm{T}}g_y(y_i,\theta)\}^{-1}\left(\frac{\partial g_y(y_i,\theta)}{\partial\theta}\right)^{\mathrm{T}}t_y. \end{aligned}$$

Proof. We start by showing for a fixed θ in a neighborhood of θ_0 there exists a lower bound of $\ell_{CE}(\theta)$. Denote $\theta = \theta_0 + kn^{-1/3}$ for $\theta \in \{\theta \mid \|\theta-\theta_0\| \le n^{-1/3}\}$ where $\|k\| = 1$. From Qin and Lawless (1994) we have that

$$t_x(\theta) = \left\{\frac{1}{n_x}\sum_{i=1}^{n_x}g_x(x_i,\theta)g_x^{\mathrm{T}}(x_i,\theta)\right\}^{-1}\left\{\frac{1}{n_x}\sum_{i=1}^{n_x}g_x(x_i,\theta)\right\} + o(n_x^{-1/3}) \qquad \text{(a.s.)}$$

$$= O(n_x^{-1/3}) \qquad \text{(a.s.)}$$

and

$$t_y(\theta) = \left\{ \frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta) g_y^{\mathrm{T}}(y_i,\theta) \right\}^{-1} \left\{ \frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta) \right\} + o(n_y^{-1/3}) \qquad \text{(a.s.)}$$

$$= O(n_y^{-1/3}) \qquad \text{(a.s.)}$$

uniformly about $\theta \in \{\theta \mid \|\theta - \theta_0\| \leq n^{-1/3}\}$. Using this result we can show by a Taylor expansion of $\ell_{CE}(\theta)$ uniformly for k

$$\ell_{CE}(\theta) = \sum_{i=1}^{n_x} t_x^{\mathrm{T}} g(x_i,\theta) - \frac{1}{2} \sum_{i=1}^{n_x} \left\{ t_x^{\mathrm{T}} g(x_i,\theta) \right\}^2 + \sum_{i=1}^{n_y} t_y^{\mathrm{T}} g(y_i,\theta) - \frac{1}{2} \sum_{i=1}^{n_y} \left\{ t_y^{\mathrm{T}} g(y_i,\theta) \right\}^2 + o(n^{1/3}) \qquad \text{(a.s.)}$$

$$\begin{aligned}
&= \frac{n_x}{2} \left\{ \frac{1}{n_x} \sum_{i=1}^{n_x} g_x^{\mathrm{T}}(x_i,\theta) \right\} \left\{ \frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i,\theta) g_x^{\mathrm{T}}(x_i,\theta) \right\}^{-1} \left\{ \frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i,\theta) \right\} \\
&\quad + \frac{n_y}{2} \left\{ \frac{1}{n_y} \sum_{i=1}^{n_y} g_y^{\mathrm{T}}(y_i,\theta) \right\} \left\{ \frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta) g_y^{\mathrm{T}}(y_i,\theta) \right\}^{-1} \left\{ \frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta) \right\} + o(n^{1/3}) \qquad \text{(a.s.)}
\end{aligned}$$

$$\begin{aligned}
&= \frac{n_x}{2} \left\{ \frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i,\theta_0) + \frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\theta_0)}{\partial \theta} k n_x^{-1/3} \right\}^{\mathrm{T}} \left\{ \frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i,\theta) g_x^{\mathrm{T}}(x_i,\theta) \right\}^{-1} \\
&\quad \times \left\{ \frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i,\theta_0) + \frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\theta_0)}{\partial \theta} k n_x^{-1/3} \right\} \\
&\quad + \frac{n_y}{2} \left\{ \frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta_0) + \frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\theta_0)}{\partial \theta} k n_y^{-1/3} \right\}^{\mathrm{T}} \left\{ \frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta) g_y^{\mathrm{T}}(y_i,\theta) \right\}^{-1} \\
&\quad \times \left\{ \frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta_0) + \frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\theta_0)}{\partial \theta} k n_y^{-1/3} \right\} + o(n^{1/3}) \qquad \text{(a.s.)}
\end{aligned}$$

$$\begin{aligned}
&= \frac{n_x}{2} \left\{ O(n_x^{-1/2} (\log\log n_x)^{1/2}) + E \frac{\partial g_x(X,\theta_0)}{\partial \theta} k n_x^{-1/3} \right\}^{\mathrm{T}} \left\{ E g_x(X,\theta_0) g_x^{\mathrm{T}}(X,\theta_0) \right\}^{-1} \\
&\quad \times \left\{ O(n_x^{-1/2} (\log\log n_x)^{1/2}) + E \frac{\partial g_x(X,\theta_0)}{\partial \theta} k n_x^{-1/3} \right\} \\
&\quad + \frac{n_y}{2} \left\{ O(n_y^{-1/2} (\log\log n_y)^{1/2}) + E \frac{\partial g_y(Y,\theta_0)}{\partial \theta} k n_y^{-1/3} \right\}^{\mathrm{T}} \left\{ E g_y(Y,\theta_0) g_y^{\mathrm{T}}(Y,\theta_0) \right\}^{-1} \\
&\quad \times \left\{ O(n_y^{-1/2} (\log\log n_y)^{1/2}) + E \frac{\partial g_y(Y,\theta_0)}{\partial \theta} k n_y^{-1/3} \right\} + o(n^{1/3}) \qquad \text{(a.s.)}
\end{aligned}$$

$$\geq (c_x - \varepsilon) n_x^{1/3} + (c_y - \varepsilon) n_y^{1/3} \qquad \text{(a.s.)}.$$

c_x is the smallest eigenvalue of

$$E\left(\frac{\partial g_x(X,\theta_0)}{\partial \theta}\right)^{\mathrm{T}} \left\{E g_x(X,\theta_0) g_x^{\mathrm{T}}(X,\theta_0)\right\}^{-1} E\left(\frac{\partial g_x(X,\theta_0)}{\partial \theta}\right).$$

c_y is the smallest eigenvalue of

$$E\left(\frac{\partial g_y(Y,\theta_0)}{\partial \theta}\right)^{\mathrm{T}} \left\{E g_y(Y,\theta_0) g_y^{\mathrm{T}}(Y,\theta_0)\right\}^{-1} E\left(\frac{\partial g_y(Y,\theta_0)}{\partial \theta}\right).$$

$c_x - \varepsilon > 0$ and $c_y - \varepsilon > 0$ which establishes the existence of a minimum value of ℓ_{CE} at some θ.

Next we will show that $\ell_{CE}(\theta_0)$ converges almost surely to a real finite value. When $\theta = \theta_0$ we have

$$\begin{aligned}
\ell_{CE}(\theta_0) = {} & \frac{n_x}{2}\left\{\frac{1}{n_x}\sum_{i=1}^{n_x} g_x^{\mathrm{T}}(x_i,\theta_0)\right\}\left\{\frac{1}{n_x}\sum_{i=1}^{n_x} g_x(x_i,\theta_0) g_x^{\mathrm{T}}(x_i,\theta_0)\right\}^{-1}\left\{\frac{1}{n_x}\sum_{i=1}^{n_x} g_x(x_i,\theta_0)\right\} \\
& + \frac{n_y}{2}\left\{\frac{1}{n_y}\sum_{i=1}^{n_y} g_y^{\mathrm{T}}(y_i,\theta_0)\right\}\left\{\frac{1}{n_y}\sum_{i=1}^{n_y} g_y(y_i,\theta_0) g_y^{\mathrm{T}}(y_i,\theta_0)\right\}^{-1}\left\{\frac{1}{n_y}\sum_{i=1}^{n_y} g_y(y_i,\theta_0)\right\} \\
& + o(1) \qquad \text{(a.s.)} \\
= {} & O(\log\log n) \qquad \text{(a.s.)}
\end{aligned}$$

which establishes the existence of a finite value of $\ell_{CE}(\theta_0)$.

From the restrictions placed on the weights of $\ell_{CE}(\theta)$ we have that for any θ

$$\begin{aligned}
0 &= \frac{1}{n_x}\sum_{i=1}^{n_x}\{1 + t_x^{\mathrm{T}} g_x(x_i,\theta)\}^{-1} g_x(x_i,\theta) = Q_{1x}(\theta, t_x) \\
0 &= \frac{1}{n_y}\sum_{i=1}^{n_y}\{1 + t_y^{\mathrm{T}} g_y(y_i,\theta)\}^{-1} g_y(y_i,\theta) = Q_{1y}(\theta, t_y)
\end{aligned}$$

which will also hold for $\hat{t}_x$, $\hat{t}_y$ and $\hat{\theta}_{CE}$.

Since ℓ_{CE} is a continuous function about θ and belongs to the ball $\|\theta-\theta_0\| \le n^{-1/3}$, $\ell_{CE}(\theta)$ has minimum value in the interior of the ball, and $\hat{\theta}_{CE}$ satisfies

$$\begin{aligned}
\frac{\partial \ell_{CE}(\theta)}{\partial \theta}\bigg|_{\theta=\hat{\theta}_{CE}} &= \sum_{i=1}^{n_x} \frac{(\partial t_x^{\mathrm{T}}(\theta)/\partial\theta) g_x(x_i,\theta) + (\partial g_x(x_i,\theta)/\partial\theta)^{\mathrm{T}} t_x(\theta)}{1 + t_x^{\mathrm{T}}(\theta) g_x(x_i,\theta)}\bigg|_{\theta=\hat{\theta}_{CE}} \\
&\quad + \sum_{i=1}^{n_y} \frac{(\partial t_y^{\mathrm{T}}(\theta)/\partial\theta) g_y(y_i,\theta) + (\partial g_y(y_i,\theta)/\partial\theta)^{\mathrm{T}} t_y(\theta)}{1 + t_y^{\mathrm{T}}(\theta) g_y(y_i,\theta)}\bigg|_{\theta=\hat{\theta}_{CE}} \\
&= \sum_{i=1}^{n_x} \frac{1}{1 + t_x^{\mathrm{T}}(\theta) g_x(x_i,\theta)} \left(\frac{\partial g_x(x_i,\theta)}{\partial\theta}\right)^{\mathrm{T}} t_x(\theta)\bigg|_{\theta=\hat{\theta}_{CE}} \\
&\quad + \sum_{i=1}^{n_y} \frac{1}{1 + t_y^{\mathrm{T}}(\theta) g_y(y_i,\theta)} \left(\frac{\partial g_y(y_i,\theta)}{\partial\theta}\right)^{\mathrm{T}} t_y(\theta)\bigg|_{\theta=\hat{\theta}_{CE}} \\
&= 0
\end{aligned}$$

which establishes $Q_2(\hat{\theta}_{CE}, \hat{t}_x, \hat{t}_y) = 0$ and completes the proof. □

Next we will derive the asymptotic distributions of $\hat{t}_x$, $\hat{t}_y$ and $\hat{\theta}_{CE}$ along with the covariances between $\hat{\theta}_{CE}$ and the Lagrange multipliers t_x and t_y.

Assumption 2 Let θ_0 be the true value of θ. Then

(a) The second derivatives $\partial^2 g_x(X,\theta)/\partial\theta\partial\theta^{\mathrm{T}}$ and $\partial^2 g_y(Y,\theta)/\partial\theta\partial\theta^{\mathrm{T}}$ are continuous in θ in a neighborhood of the true value θ_0.

(b) $\|\partial^2 g_x(X,\theta)/\partial\theta\partial\theta^{\mathrm{T}}\|$ and $\|\partial^2 g_y(Y,\theta)/\partial\theta\partial\theta^{\mathrm{T}}\|$ can be bounded by some integrable function in the neighborhood of θ_0.

Additionally we will denote $g_x(X,\theta)$ and $g_y(Y,\theta)$ as g_x and g_y respectively.

Theorem 2.1 *Given the conditions in Assumptions 1 and 2 we have*

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_{CE} - \theta_0) &\longrightarrow N(0, W^{-1}(W+V)W^{-1}) \\
\sqrt{n}(\hat{t}_x - 0) &\longrightarrow N(0, U_x) \\
\sqrt{n}(\hat{t}_y - 0) &\longrightarrow N(0, U_y)
\end{aligned}$$

where

$$W = \left\{ E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right) + E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right) \right\},$$

$$\begin{aligned} V = \Bigg\{ & E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E(g_x g_y^{\mathrm{T}}) \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right) + \\ & E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E(g_y g_x^{\mathrm{T}}) \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right) \Bigg\}, \end{aligned}$$

$$\begin{aligned} U_x = & \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} \left(I - E\left(\frac{\partial g_x}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} \right) \\ & + \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right) W^{-1} V W^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} \\ & - \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E(g_y g_x^{\mathrm{T}}) \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} \\ & - \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E(g_x g_y^{\mathrm{T}}) \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1}, \end{aligned}$$

$$\begin{aligned} U_y = & \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} \left(I - E\left(\frac{\partial g_y}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} \right) \\ & + \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right) W^{-1} V W^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} \\ & - \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E(g_x g_y^{\mathrm{T}}) \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} \\ & - \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E(g_y g_x^{\mathrm{T}}) \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1}. \end{aligned}$$

Additionally if $cov(g_x, g_y) = 0$ *then* $\hat{\theta}_{CE}$ *is asymptotically uncorrelated with* $\hat{t}_x$ *and* $\hat{t}_y$.

Proof. To start we use Taylor expansions and the results from Lemma 2.2 to find solutions for $\hat{t}_x - 0$, $\hat{t}_y - 0$ and $\hat{\theta}_{CE} - \theta_0$.

Expanding $Q_{1x}(\hat{\theta}_{CE}, \hat{t}_x)$ at θ,0 we get

$$\begin{aligned} 0 &= Q_{1x}(\hat{\theta}_{CE}, \hat{t}_x) \\ &= Q_{1x}(\theta_0, 0) + \frac{\partial}{\partial \theta} Q_{1x}(\theta_0, 0)(\hat{\theta}_{CE} - \theta_0) + \frac{\partial}{\partial t_x^{\mathrm{T}}} Q_{1x}(\theta_0, 0)(\hat{t}_x - 0) + o_p(\delta_x) \end{aligned}$$

where $\delta_x = \|\hat{\theta}_{CE} - \theta_0\| + \|\hat{t}_x\|$.

Expanding $Q_{1y}(\hat{\theta}_{CE}, \hat{t}_y)$ at θ,0 we get

$$\begin{aligned} 0 &= Q_{1y}(\hat{\theta}_{CE}, \hat{t}_y) \\ &= Q_{1y}(\theta_0, 0) + \frac{\partial}{\partial \theta} Q_{1y}(\theta_0, 0)(\hat{\theta}_{CE} - \theta_0) + \frac{\partial}{\partial t_y^{\mathrm{T}}} Q_{1x}(\theta_0, 0)(\hat{t}_y - 0) + o_p(\delta_y) \end{aligned}$$

where $\delta_y = \|\hat{\theta}_{CE} - \theta_0\| + \|\hat{t}_y\|$.

Expanding $Q_2(\hat{\theta}_{CE}, \hat{t}_x, \hat{t}_y)$ at θ,0,0 we get

$$\begin{aligned} 0 &= Q_2(\hat{\theta}_{CE}, \hat{t}_x, \hat{t}_y) \\ &= Q_2(\theta_0, 0, 0) + \frac{\partial}{\partial \theta} Q_2(\theta_0, 0, 0)(\hat{\theta}_{CE} - \theta_0) \\ &\quad + \frac{\partial}{\partial t_x^{\mathrm{T}}} Q_2(\theta_0, 0, 0)(\hat{t}_x - 0) + \frac{\partial}{\partial t_y^{\mathrm{T}}} Q_2(\theta_0, 0, 0)(\hat{t}_y - 0) + o_p(\delta) \end{aligned}$$

where $\delta = \|\hat{\theta}_{CE} - \theta_0\| + \|\hat{t}_x\| + \|\hat{t}_y\|$.

The derivatives of Q_{1x}, Q_{1y}, and Q_2 with respect to θ, t_x and t_y are

$$\begin{aligned} \frac{\partial Q_{1x}(\theta_0, 0)}{\partial \theta} &= \frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i, \theta_0)}{\partial \theta} \\ \frac{\partial Q_{1y}(\theta_0, 0)}{\partial \theta} &= \frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i, \theta_0)}{\partial \theta} \\ \frac{\partial Q_{1x}(\theta_0, 0)}{\partial t_x^{\mathrm{T}}} &= -\frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i, \theta_0) g_x^{\mathrm{T}}(x_i, \theta_0) \\ \frac{\partial Q_{1y}(\theta_0, 0)}{\partial t_y^{\mathrm{T}}} &= -\frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i, \theta_0) g_y^{\mathrm{T}}(y_i, \theta_0) \\ \frac{\partial Q_2(\theta_0, 0, 0)}{\partial t_x^{\mathrm{T}}} &= \frac{1}{n_x} \sum_{i=1}^{n_x} \left(\frac{\partial g_x(x_i, \theta_0)}{\partial \theta} \right)^{\mathrm{T}} \\ \frac{\partial Q_2(\theta_0, 0, 0)}{\partial t_y^{\mathrm{T}}} &= \frac{1}{n_y} \sum_{i=1}^{n_y} \left(\frac{\partial g_y(y_i, \theta_0)}{\partial \theta} \right)^{\mathrm{T}} \end{aligned}$$

and $Q_2(\theta_0, 0, 0) = \partial Q_2(\theta_0, 0, 0)/\partial \theta = 0$. A few algebraic steps yield

$$(\hat{t}_x - 0) = \left\{ -\frac{\partial Q_{1x}(\theta_0, 0)}{\partial t_x^{\mathrm{T}}} \right\}^{-1} \left\{ Q_{1x}(\theta_0, 0) + \frac{\partial}{\partial \theta} Q_{1x}(\theta_0, 0)(\hat{\theta}_{CE} - \theta_0) + o_p(\delta_x) \right\} \tag{2.9}$$

$$(\hat{t}_y - 0) = \left\{ -\frac{\partial Q_{1y}(\theta_0, 0)}{\partial t_y^{\mathrm{T}}} \right\}^{-1} \left\{ Q_{1y}(\theta_0, 0) + \frac{\partial}{\partial \theta} Q_{1y}(\theta_0, 0)(\hat{\theta}_{CE} - \theta_0) + o_p(\delta_y) \right\} \tag{2.10}$$

$$0 = \frac{\partial}{\partial t_x^{\mathrm{T}}} Q_2(\theta_0, 0, 0)\hat{t}_x + \frac{\partial}{\partial t_y^{\mathrm{T}}} Q_2(\theta_0, 0, 0)\hat{t}_y + o_p(\delta). \tag{2.11}$$

Define

$$S_n = \left[\begin{array}{ccc} -\frac{\partial Q_{1x}}{\partial t_x^{\mathrm{T}}} & 0 & \frac{\partial Q_{1x}}{\partial \theta} \\ 0 & -\frac{\partial Q_{1y}}{\partial t_y^{\mathrm{T}}} & \frac{\partial Q_{1y}}{\partial \theta} \\ \frac{\partial Q_2}{\partial t_x^{\mathrm{T}}} & \frac{\partial Q_2}{\partial t_y^{\mathrm{T}}} & 0 \end{array} \right]_{(\theta_0,0,0)},$$

and solving Equations 2.9, 2.10 and 2.11 gives us the following

$$\left[\begin{array}{c} \hat{t}_x - 0 \\ \hat{t}_y - 0 \\ \hat{\theta}_{CE} - \theta_0 \end{array} \right] = S_n^{-1} \left[\begin{array}{c} -Q_{1x}(\theta_0, 0) + o_p(\delta_x) \\ -Q_{1y}(\theta_0, 0) + o_p(\delta_y) \\ o_p(\delta) \end{array} \right].$$

The next step establishes the asymptotic distributions. As $n \longrightarrow \infty$

$$S_n \longrightarrow \left[\begin{array}{ccc} E(g_x g_x^{\mathrm{T}}) & 0 & E\left(\frac{\partial g_x}{\partial \theta}\right) \\ 0 & E(g_y g_y^{\mathrm{T}}) & E\left(\frac{\partial g_y}{\partial \theta}\right) \\ E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} & E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} & 0 \end{array} \right] \equiv \left[\begin{array}{ccc} \Sigma_{11} & 0 & \Sigma_{13} \\ 0 & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & 0 \end{array} \right].$$

By assumption $E(g_x) = 0$ and $E(g_y) = 0$, so by the weak law of large numbers

$$-\sqrt{n} Q_{1x}(\theta_0, 0) \longrightarrow N(0, E(g_x g_x^{\mathrm{T}}))$$
$$-\sqrt{n} Q_{1y}(\theta_0, 0) \longrightarrow N(0, E(g_y g_y^{\mathrm{T}})).$$

Let $\Sigma_{12} \equiv E(g_x g_y^{\mathrm{T}})$ and $\Sigma_{21} \equiv E(g_y g_x^{\mathrm{T}})$. The asymptotic distribution of $[-\sqrt{n} Q_{1x}^{\mathrm{T}}(\theta_0, 0), -\sqrt{n} Q_{1y}^{\mathrm{T}}(\theta_0^{\mathrm{T}}, 0)]$ is

$$\left[\begin{array}{c} -\sqrt{n} Q_{1x}(\theta_0, 0) \\ -\sqrt{n} Q_{1y}(\theta_0, 0) \end{array} \right] \longrightarrow N\left(\left[\begin{array}{c} 0 \\ 0 \end{array} \right], \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right] \right)$$

which establishes the asymptotic normality of $[(\hat{t}_x-0)^{\mathrm{T}},(\hat{t}_y-0)^{\mathrm{T}},(\hat{\theta}_{CE}-\theta_0)^{\mathrm{T}}]^{\mathrm{T}}$ since they are linear transformations of random normally distributed variables. By Slutsky's Theorem the covariance matrix of $[(\hat{t}_x-0)^{\mathrm{T}},(\hat{t}_y-0)^{\mathrm{T}},(\hat{\theta}_{CE}-\theta_0)^{\mathrm{T}}]^{\mathrm{T}}$ is

$$\begin{bmatrix} -\Sigma_{11}^{-1}\left(I-\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1}\right) & \Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1} \\ \Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1} & -\Sigma_{22}^{-1}\left(I-\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1}\right) \\ \Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1} & \Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \times$$
$$\begin{bmatrix} -\Sigma_{11}^{-1}\left(I-\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1}\right) & \Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1} & \Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1} \\ \Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1} & -\Sigma_{22}^{-1}\left(I-\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1}\right) & \Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1} \end{bmatrix}$$

where $\Sigma_{33.1} \equiv \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}+\Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23}$.

Multiplying through yields the following variances

$$\begin{aligned} var(\sqrt{n}(\hat{t}_x-0)) &= \Sigma_{11}^{-1}(I-\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1}) \\ &\quad +\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}(\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{23}+\Sigma_{32}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13})\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1} \\ &\quad -\Sigma_{11}^{-1}(\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1}\Sigma_{21}+\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31})\Sigma_{11}^{-1} \end{aligned}$$

$$\begin{aligned} var(\sqrt{n}(\hat{t}_y-0)) &= \Sigma_{22}^{-1}(I-\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1}) \\ &\quad +\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}(\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{23}+\Sigma_{32}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13})\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1} \\ &\quad -\Sigma_{22}^{-1}(\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{12}+\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32})\Sigma_{22}^{-1} \end{aligned}$$

$$var(\sqrt{n}(\hat{\theta}_{CE}-\theta_0)) = \Sigma_{33.1}^{-1}+\Sigma_{33.1}^{-1}(\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{23}+\Sigma_{32}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13})\Sigma_{33.1}^{-1}$$

and covariances

$$\begin{aligned} cov(\sqrt{n}(\hat{t}_x-0),\sqrt{n}(\hat{\theta}_{CE}-\theta_0)) &= \Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}(\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{23}+\Sigma_{32}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13})\Sigma_{33.1}^{-1} \\ &\quad -\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1} \end{aligned}$$

$$\begin{aligned} cov(\sqrt{n}(\hat{t}_y-0),\sqrt{n}(\hat{\theta}_{CE}-\theta_0)) &= \Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}(\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{23}+\Sigma_{32}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13})\Sigma_{33.1}^{-1} \\ &\quad -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1} \end{aligned}$$

which completes the proof. □

From the proof we can see that if we properly condition the two sets of estimating equations the correlation between $(\hat{\theta}_{CE}, \hat{t}_x)$ and $(\hat{\theta}_{CE}, \hat{t}_y)$ will be zero, and given this condition we have that $\sqrt{n}(\hat{\theta}_{CE} - \theta_0)$ is asymptotically independent of $\sqrt{n}(\hat{t}_x - 0)$ and $\sqrt{n}(\hat{t}_y - 0)$.

Theorem 2.1 can be viewed as a generalization of Theorem 1 from Qin and Lawless (1994). If $z_1, \ldots, z_n$ is an *i.i.d.* sample, we have the following property of the maximum empirical likelihood estimator

$$\sqrt{n}(\hat{\theta}_E - \theta_0) \longrightarrow N\left(0, \left\{ E\left(\frac{\partial g}{\partial \theta}\right)^{\mathrm{T}} \left\{E(gg^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g}{\partial \theta}\right) \right\}^{-1}\right).$$

If we split the sample into two parts of approximately the same size (define $n_2 = n/2$) the asymptotic behavior of the maximum composite empirical likelihood estimator is

$$\sqrt{n_2}(\hat{\theta}_{CE} - \theta_0) \longrightarrow N\left(0, \left\{ 2E\left(\frac{\partial g}{\partial \theta}\right)^{\mathrm{T}} \left\{E(gg^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g}{\partial \theta}\right) \right\}^{-1}\right)$$

$$\sqrt{n}(\hat{\theta}_{CE} - \theta_0) \longrightarrow N\left(0, \left\{ E\left(\frac{\partial g}{\partial \theta}\right)^{\mathrm{T}} \left\{E(gg^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g}{\partial \theta}\right) \right\}^{-1}\right),$$

so both the maximum empirical likelihood estimator and the maximum composite empirical likelihood estimator exhibit identical asymptotic behavior.

We can create an approximate confidence interval using the asymptotic normality of θ_{CE} shown in Theorem 2.1. The variance of $\sqrt{n}(\hat{\theta}_{CE} - \theta_0)$ can be estimated using $\widehat{W}^{-1}(\widehat{W} + \widehat{V})\widehat{W}^{-1}$ where

$$\widehat{W} = \left\{ \left(\frac{1}{n_x}\sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\hat{\theta})}{\partial \theta}\right)^{\mathrm{T}} \left(\frac{1}{n_x}\sum_{i=1}^{n_x} g_x(x_i,\hat{\theta}) g_x^{\intercal}(x_i,\hat{\theta})\right)^{-1} \left(\frac{1}{n_x}\sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\hat{\theta})}{\partial \theta}\right) + \left(\frac{1}{n_y}\sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\hat{\theta})}{\partial \theta}\right)^{\mathrm{T}} \left(\frac{1}{n_y}\sum_{i=1}^{n_y} g_y(y_i,\hat{\theta}) g_y^{\intercal}(y_i,\hat{\theta})\right)^{-1} \left(\frac{1}{n_y}\sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\hat{\theta})}{\partial \theta}\right) \right\}$$

$$\widehat{V} = \left\{ \left(\frac{1}{n_x}\sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\hat{\theta})}{\partial \theta}\right)^{\mathrm{T}} \left(\frac{1}{n_x}\sum_{i=1}^{n_x} g_x(x_i,\hat{\theta}) g_x^{\intercal}(x_i,\hat{\theta})\right)^{-1} \left(\frac{1}{n_{xy}}\sum_{i=1}^{n_{xy}} g_x(x_i,\hat{\theta}) g_y^{\intercal}(y_i,\hat{\theta})\right) \times \left(\frac{1}{n_y}\sum_{i=1}^{n_y} g_y(y_i,\hat{\theta}) g_y^{\intercal}(y_i,\hat{\theta})\right) \left(\frac{1}{n_y}\sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\hat{\theta})}{\partial \theta}\right) \right\}$$

$$+\left\{\left(\frac{1}{n_y}\sum_{i=1}^{n_y}\frac{\partial g_y(y_i,\hat{\theta})}{\partial\theta}\right)\left(\frac{1}{n_y}\sum_{i=1}^{n_y}g_y(y_i,\hat{\theta})g_y^{\intercal}(y_i,\hat{\theta})\right)^{-1}\left(\frac{1}{n_{xy}}\sum_{i=1}^{n_{xy}}g_y(y_i,\hat{\theta})g_x^{\intercal}(x_i,\hat{\theta})\right)\times\right.$$
$$\left.\left(\frac{1}{n_x}\sum_{i=1}^{n_x}g_x(x_i,\hat{\theta})g_x^{\intercal}(x_i,\hat{\theta})\right)\left(\frac{1}{n_x}\sum_{i=1}^{n_x}\frac{\partial g_x(x_i,\hat{\theta})}{\partial\theta}\right)\right\}$$

with n_{xy} denoting the number of complete data pairs (x, y) the sample (since as stated we are not assuming the data are complete). One disadvantage of using this approximation is it does not allow for asymmetric intervals, which would lead to an incorrect conclusion if the data are from a skewed distribution. The following theorem shows the distributions of the composite empirical likelihood ratio test statistics.

Theorem 2.2 *The composite empirical likelihood ratio statistic for testing $H_0 : \theta = \theta_0$ is*

$$T = 2\ell_{CE}(\theta_0) - 2\ell_{CE}(\hat{\theta}_{CE}) \tag{2.12}$$

where $\ell_{CE}(\theta)$ is given by Definition 2.8. Under Assumptions 1 and 2 we have

$$T \longrightarrow Q(\lambda)$$

as $n \longrightarrow \infty$ when H_0 is true. $Q(\lambda) = \sum_{i=1}^{l}\lambda_i\chi^2(1)$ is a weighted chi square random variable where λ_i for $i = 1, \ldots, l$ is the set of all non zero eigenvalues of

$$A\Gamma = \begin{bmatrix} A_{xx} & A_{xy} \\ A_{yx} & A_{yy} \end{bmatrix}\begin{bmatrix} E(g_x g_x^{\mathrm{T}}) & E(g_x g_y^{\mathrm{T}}) \\ E(g_y g_x^{\mathrm{T}}) & E(g_y g_y^{\mathrm{T}}) \end{bmatrix}$$

where

$$A_{xx} = \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_x}{\partial\theta}\right)W^{-1}E\left(\frac{\partial g_x}{\partial\theta}\right)^{\mathrm{T}}\left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1}$$
$$A_{yy} = \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_y}{\partial\theta}\right)W^{-1}E\left(\frac{\partial g_y}{\partial\theta}\right)^{\mathrm{T}}\left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1}$$
$$A_{xy} = \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_x}{\partial\theta}\right)W^{-1}E\left(\frac{\partial g_y}{\partial\theta}\right)^{\mathrm{T}}\left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1}$$
$$A_{yx} = \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_y}{\partial\theta}\right)W^{-1}E\left(\frac{\partial g_x}{\partial\theta}\right)^{\mathrm{T}}\left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1}$$

and W is defined in Theorem 2.1.

Proof. From Lemma 2.2 we have the following Taylor expansion

$$\begin{aligned}\ell_{CE}(\theta) = &\frac{n}{2}\left\{\frac{1}{n}\sum_{i=1}^{n} g_x(x_i,\theta)\right\}^{\mathrm{T}}\left\{\frac{1}{n}\sum_{i=1}^{n} g_x(x_i,\theta)g_x(x_i,\theta)^{\mathrm{T}}\right\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n} g_x(x_i,\theta)\right\}\\ &+\frac{n}{2}\left\{\frac{1}{n}\sum_{i=1}^{n} g_y(y_i,\theta)\right\}^{\mathrm{T}}\left\{\frac{1}{n}\sum_{i=1}^{n} g_y(y_i,\theta)g_y(y_i,\theta)^{\mathrm{T}}\right\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n} g_y(y_i,\theta)\right\}\\ &+o_p(1).\end{aligned}$$

Using the notation established in Lemma 2.2 and the results from Theorem 2.1 we have the following

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} g_x(x_i,\hat{\theta}_{CE}) &= Q_{1x}(\hat{\theta}_{CE},0)\\ &= Q_{1x}(\theta_0,0)+\partial/\partial\theta Q_{1x}(\theta_0,0)(\hat{\theta}_{CE}-\theta_0)+o_p(1),\end{aligned} \tag{2.13}$$

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} g_y(y_i,\hat{\theta}_{CE}) &= Q_{1y}(\hat{\theta}_{CE},0)\\ &= Q_{1y}(\theta_0,0)+\partial/\partial\theta Q_{1y}(\theta_0,0)(\hat{\theta}_{CE}-\theta_0)+o_p(1),\end{aligned} \tag{2.14}$$

and

$$(\hat{\theta}_{CE}-\theta_0) = -\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1}Q_{1x}(\theta_0,0)-\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1}Q_{1y}(\theta_0,0)+o_p(1). \tag{2.15}$$

Under H_0 we have

$$\ell_{CE}(\theta_0) = \frac{n}{2}Q_{1x}^{\mathrm{T}}(\theta_0,0)\Sigma_{11}^{-1}Q_{1x}(\theta_0,0)+\frac{n}{2}Q_{1y}^{\mathrm{T}}(\theta_0,0)\Sigma_{22}^{-1}Q_{1y}(\theta_0,0)+o_p(1) \tag{2.16}$$

by direct substitution. Next

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} g_x(x_i,\hat{\theta}_{CE})g_x^{\mathrm{T}}(x_i,\hat{\theta}_{CE}) &= \Sigma_{11}+o_p(1),\\ \frac{1}{n}\sum_{i=1}^{n} g_y(y_i,\hat{\theta}_{CE})g_y^{\mathrm{T}}(y_i,\hat{\theta}_{CE}) &= \Sigma_{22}+o_p(1),\end{aligned}$$

so by Slutsky's Theorem

$$\ell_{CE}(\hat{\theta}_{CE}) = \frac{n}{2} Q_{1x}^{\mathrm{T}}(\hat{\theta}_{CE}, 0)\Sigma_{11}^{-1} Q_{1x}(\hat{\theta}_{CE}, 0) + \frac{n}{2} Q_{1y}^{\mathrm{T}}(\hat{\theta}_{CE}, 0)\Sigma_{22}^{-1} Q_{1y}(\hat{\theta}_{CE}, 0) + o_p(1). \tag{2.17}$$

Equations 2.13, 2.14 and 2.15 establish

$$\begin{aligned}
Q_{1x}(\hat{\theta}_{CE}, 0) &= Q_{1x}(\theta_0, 0) - \Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1} Q_{1x}(\theta_0, 0) - \Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1} Q_{1y}(\theta_0, 0) + o_p(1), \\
Q_{1y}(\hat{\theta}_{CE}, 0) &= Q_{1y}(\theta_0, 0) - \Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1} Q_{1y}(\theta_0, 0) - \Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1} Q_{1x}(\theta_0, 0) + o_p(1),
\end{aligned}$$

which are then substituted into Equation 2.17 resulting in

$$\begin{aligned}
\ell_{CE}(\hat{\theta}_{CE}) = {} & \frac{n}{2} Q_{1x}^{\mathrm{T}}(\theta_0, 0)\Sigma_{11}^{-1} Q_{1x}(\theta_0, 0) + \frac{n}{2} Q_{1y}^{\mathrm{T}}(\theta_0, 0)\Sigma_{22}^{-1} Q_{1y}(\theta_0, 0) \\
& - \frac{n}{2} Q_{1x}^{\mathrm{T}}(\theta_0, 0)\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1} Q_{1x}(\theta_0, 0) \\
& - \frac{n}{2} Q_{1y}^{\mathrm{T}}(\theta_0, 0)\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1} Q_{1y}(\theta_0, 0) \\
& - \frac{n}{2} Q_{1x}^{\mathrm{T}}(\theta_0, 0)\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1} Q_{1y}(\theta_0, 0) \\
& - \frac{n}{2} Q_{1y}^{\mathrm{T}}(\theta_0, 0)\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1} Q_{1x}(\theta_0, 0) \\
& + o_p(1).
\end{aligned} \tag{2.18}$$

T is equal to the difference of Equations 2.16 and 2.18 so

$$\begin{aligned}
T &= 2\ell_{CE}(\theta_0) - 2\ell_{CE}(\hat{\theta}_{CE}) \\
&= nQ_{1x}^{\mathrm{T}}(\theta_0, 0)\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1} Q_{1x}(\theta_0, 0) \\
&\quad + nQ_{1y}^{\mathrm{T}}(\theta_0, 0)\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1} Q_{1y}(\theta_0, 0) \\
&\quad + nQ_{1x}^{\mathrm{T}}(\theta_0, 0)\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1} Q_{1y}(\theta_0, 0) \\
&\quad + nQ_{1y}^{\mathrm{T}}(\theta_0, 0)\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1} Q_{1x}(\theta_0, 0) \\
&\quad + o_p(1) \\
&= \sqrt{n}Q_{1x}^{\mathrm{T}}(\theta_0, 0)\left\{\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1}\right\}\sqrt{n}Q_{1x}(\theta_0, 0) \\
&\quad + \sqrt{n}Q_{1y}^{\mathrm{T}}(\theta_0, 0)\left\{\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1}\right\}\sqrt{n}Q_{1y}(\theta_0, 0)
\end{aligned}$$

$$+ \sqrt{n}Q_{1x}^{\mathrm{T}}(\theta_0, 0)\left\{\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1}\right\}\sqrt{n}Q_{1y}(\theta_0, 0)$$
$$+ \sqrt{n}Q_{1y}^{\mathrm{T}}(\theta_0, 0)\left\{\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1}\right\}\sqrt{n}Q_{1x}(\theta_0, 0)$$
$$+ o_p(1).$$

$\sqrt{n}Q_{1x}(\theta_0, 0)$ and $\sqrt{n}Q_{1y}(\theta_0, 0)$ converge to a multivariate normal with mean 0 and variance

$$\Gamma = \begin{bmatrix} E(g_x g_x^{\mathrm{T}}) & E(g_x g_y^{\mathrm{T}}) \\ E(g_y g_x^{\mathrm{T}}) & E(g_y g_y^{\mathrm{T}}) \end{bmatrix}.$$

The matrix

$$A = \begin{bmatrix} \Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1} & \Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1} \\ \Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1} & \Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1} \end{bmatrix}$$

is positive definite, and we can write T as

$$\begin{bmatrix} \sqrt{n}Q_{1x}^{\mathrm{T}}(\theta_0, 0) & \sqrt{n}Q_{1y}^{\mathrm{T}}(\theta_0, 0) \end{bmatrix} \times A \times \begin{bmatrix} \sqrt{n}Q_{1x}(\theta_0, 0) \\ \sqrt{n}Q_{1y}(\theta_0, 0) \end{bmatrix}.$$

Therefore the test statistic T converges to a weighted chi square distribution with weights determined by $A\Gamma$. □

The weighed chi square result generally occurs in cases where the distribution is incorrectly specified (Kent, 1982). In our case this result stems from the fact that the likelihood may not be accounting for the correlation between the two random variables, hence the weighted chi square would appear to be a logical result. It is important to note that if $A\Gamma$ were idempotent then there would be p eigenvalues equal to 1, and any remaining eigenvalues would be 0. This would result in a chi square with p degrees of freedom, which occurs with the empirical likelihood and when the parametric likelihood is correctly specified. We can add additional conditions to obtain this result for the composite empirical likelihood test statistic. If $Q_{1x}(\theta_0, 0)$ and $Q_{1y}(\theta_0, 0)$ have a correlation of 0 then we have asymptotic independence. This can be accomplished by defining the estimating equations so that $cov(g_x, g_y) = 0$. Addition of this condition results in the following corollary.

Corollary 2.1 *Assume that g_x and g_y are defined so that $cov(g_x, g_y) = 0$. Under Assumptions 1 and 2 the test statistic T defined in Theorem 2.2 satisfies*

$$T \longrightarrow \chi^2(p)$$

as $n \longrightarrow \infty$.

Proof. The test statistic can be rewritten as

$$\begin{aligned}
T = & \sqrt{n}Q_{1x}^{\mathrm{T}}(\theta_0,0)\Sigma_{11}^{-1/2}\left\{\Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2}\right\}\Sigma_{11}^{-1/2}\sqrt{n}Q_{1x}(\theta_0,0) \\
& + \sqrt{n}Q_{1y}^{\mathrm{T}}(\theta_0,0)\Sigma_{22}^{-1/2}\left\{\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2}\right\}\Sigma_{22}^{-1/2}\sqrt{n}Q_{1y}(\theta_0,0) \\
& + \sqrt{n}Q_{1x}^{\mathrm{T}}(\theta_0,0)\Sigma_{11}^{-1/2}\left\{\Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2}\right\}\Sigma_{22}^{-1/2}\sqrt{n}Q_{1y}(\theta_0,0) \\
& + \sqrt{n}Q_{1y}^{\mathrm{T}}(\theta_0,0)\Sigma_{22}^{-1/2}\left\{\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2}\right\}\Sigma_{11}^{-1/2}\sqrt{n}Q_{1x}(\theta_0,0) \\
& + o_p(1) \\
= & \left[\begin{array}{cc} \sqrt{n}Q_{1x}^{\mathrm{T}}(\theta_0,0)^{\mathrm{T}}\Sigma_{11}^{-1/2} & \sqrt{n}Q_{1y}^{\mathrm{T}}(\theta_0,0)^{\mathrm{T}}\Sigma_{22}^{-1/2} \end{array}\right] \times A \times \left[\begin{array}{c} \Sigma_{11}^{-1/2}\sqrt{n}Q_{1x}(\theta_0,0) \\ \Sigma_{22}^{-1/2}\sqrt{n}Q_{1y}(\theta_0,0) \end{array}\right] \\
& + o_p(1)
\end{aligned}$$

where

$$A = \left[\begin{array}{cc} \Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2} & \Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2} \\ \Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2} & \Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2} \end{array}\right].$$

Now

$$\begin{aligned}
A^2 = & \left[\begin{array}{cc} \Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2} & \Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2} \\ \Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2} & \Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2} \end{array}\right] \\
& \times \left[\begin{array}{cc} \Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2} & \Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2} \\ \Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2} & \Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2} \end{array}\right]
\end{aligned}$$

$$
= \left[\begin{array}{c:c}
\begin{array}{c}
\Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2} \\
+\Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2}
\end{array} &
\begin{array}{c}
\Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2} \\
+\Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2}
\end{array} \\ \hdashline
\begin{array}{c}
\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2} \\
+\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2}
\end{array} &
\begin{array}{c}
\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2} \\
+\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2}
\end{array}
\end{array}\right]
$$

$$
= \left[\begin{array}{c:c}
\begin{array}{c}
\Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1} \\
\times\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13} + \Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23}\times \\
\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2}
\end{array} &
\begin{array}{c}
\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1} \\
\times\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13} + \Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23}\times \\
\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2}
\end{array} \\ \hdashline
\begin{array}{c}
\Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1} \\
\times\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13} + \Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23}\times \\
\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2}
\end{array} &
\begin{array}{c}
\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1} \\
\times\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13} + \Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23}\times \\
\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2}
\end{array}
\end{array}\right]
$$

$$
= \left[\begin{array}{c:c}
\begin{array}{c}
\Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1} \\
\times\Sigma_{33.1}\times \\
\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2}
\end{array} &
\begin{array}{c}
\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1} \\
\times\Sigma_{33.1}\times \\
\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2}
\end{array} \\ \hdashline
\begin{array}{c}
\Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1} \\
\times\Sigma_{33.1}\times \\
\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2}
\end{array} &
\begin{array}{c}
\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1} \\
\times\Sigma_{33.1}\times \\
\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2}
\end{array}
\end{array}\right]
$$

$$
= \begin{bmatrix}
\Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2} & \Sigma_{11}^{-1/2}\Sigma_{13}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2} \\
\Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{31}\Sigma_{11}^{-1/2} & \Sigma_{22}^{-1/2}\Sigma_{23}\Sigma_{33.1}^{-1}\Sigma_{32}\Sigma_{22}^{-1/2}
\end{bmatrix}
$$

so A is idempotent. Additionally $\Sigma_{11}^{-1/2}\sqrt{n}Q_{1x}(\theta_0, 0)$ and $\Sigma_{22}^{-1/2}\sqrt{n}Q_{1y}(\theta_0, 0)$ are independent standard multivariate normal random variables. A is symmetric with trace of p so T has a chi square distribution with p degrees of freedom. □

One application of the composite empirical likelihood is reduction of computation time by splitting up large data sets into smaller pieces. In this instance the estimating equations and parameters of interest are identical in both groups, and by assumption of an *i.i.d.* sample the two data groups are independent. The following corollary shows the asymptotic distribution of the test statistic under these conditions.

Corollary 2.2 *Let z be an i.i.d univariate sample from some distribution F_0 and randomly separate z into two mutually exclusive subsets x and y. Let the parameters and estimating equations for both groups be identical, so $g_x = g_y = g$ and $\theta_x = \theta_y = \theta$. Define p as the number of parameters. Under Assumptions 1 and 2 the test statistic T defined in Theorem 2.2 satisfies*

$$T \longrightarrow \chi^2(p)$$

as $n \longrightarrow \infty$.

Although the result follows directly from Corollary 2.1 we also give the following alternate proof.

Proof. Since x and y are from the same distribution and the estimating equations are identical we have the following simplifications

$$\begin{aligned}
\Sigma_{13} &= \Sigma_{23} = E\left(\frac{\partial g}{\partial \theta}\right) \\
\Sigma_{11} &= \Sigma_{22} = E(gg^{\mathrm{T}}) \\
W &= \left\{\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13} + \Sigma_{32}\Sigma_{22}^{-1}\Sigma_{23}\right\} = 2\left\{E\left(\frac{\partial g}{\partial \theta}\right)^{\mathrm{T}}\left\{Egg^{\mathrm{T}}\right\}^{-1}E\left(\frac{\partial g}{\partial \theta}\right)\right\}
\end{aligned}$$

so we can rewrite the test statistic T as

$$\begin{aligned}
T &= 2\ell_{CE}(\theta_0) - 2\ell_{CE}(\hat{\theta}_{CE}) \\
&= \sqrt{n}Q_{1x}^{\mathrm{T}}(\theta_0, 0)\left\{E(gg^{\mathrm{T}})\right\}^{-1/2}(1/2)A_0\left\{E(gg^{\mathrm{T}})\right\}^{-1/2}\sqrt{n}Q_{1x}(\theta_0, 0) \\
&\quad + \sqrt{n}Q_{1y}^{\mathrm{T}}(\theta_0, 0)\left\{E(gg^{\mathrm{T}})\right\}^{-1/2}(1/2)A_0\left\{E(gg^{\mathrm{T}})\right\}^{-1/2}\sqrt{n}Q_{1y}(\theta_0, 0) \\
&\quad + \sqrt{n}Q_{1x}^{\mathrm{T}}(\theta_0, 0)\left\{E(gg^{\mathrm{T}})\right\}^{-1/2}(1/2)A_0\left\{E(gg^{\mathrm{T}})\right\}^{-1/2}\sqrt{n}Q_{1y}(\theta_0, 0) \\
&\quad + \sqrt{n}Q_{1y}^{\mathrm{T}}(\theta_0, 0)\left\{E(gg^{\mathrm{T}})\right\}^{-1/2}(1/2)A_0\left\{E(gg^{\mathrm{T}})\right\}^{-1/2}\sqrt{n}Q_{1x}(\theta_0, 0)
\end{aligned}$$

$$+ o_p(1)$$

where

$$A_0 \equiv \left\{E(gg^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g}{\partial \theta}\right)^{\mathrm{T}} \left\{E(gg^{\mathrm{T}})\right\}^{-1/2}.$$

$\left\{Egg^{\mathrm{T}}\right\}^{-1/2} \sqrt{n} Q_{1x}(\theta_0, 0)$ and $\left\{Egg^{\mathrm{T}}\right\}^{-1/2} \sqrt{n} Q_{1y}(\theta_0, 0)$ are both independent standard normal variables. The matrix A_0 is symmetric and idempotent with a rank of p, therefore has p eigenvalues equal to 1. The matrix

$$A\Gamma = AI = A = 1/2 \begin{bmatrix} A_0 & A_0 \\ A_0 & A_0 \end{bmatrix}$$

can be rewritten as

$$(1/2) \left\{E(gg^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g}{\partial \theta}\right)^{\mathrm{T}} \left\{E(gg^{\mathrm{T}})\right\}^{-1/2} \otimes \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The matrix $[11^{\mathrm{T}}]$ has a single eigenvalue of 2. The eigenvalues of A are equal to the eigenvalues of A_0 times 2 times $1/2$ which results in p eigenvalues of 1. Therefore T is asymptotically chi square with p degrees of freedom. □

The result from Corollary 2.2 indicates that given correct assumptions, the likelihood ratio test statistic for the composite empirical likelihood has the same asymptotic behavior as the Fisher likelihood. This indicates the composite empirical likelihood, much like the parametric composite marginal likelihood, is an equivalent form of the standard empirical likelihood given independence of the data.

2.3 Estimation of a Weighted Chi Square Random Variable

Working with a weighted chi square presents some challenges. The probability density function of a weighted chi square is an incomplete gamma function (Moschopoulos and Canada, 1984), and in our case the distribution requires the eigenvalues of the transformation matrix, which itself is a

function of the parameters. Methods of approximating weighted chi squared variables have been explored by Welch (1938), Satterthwaite (1946) and Box (1954) by adjusting a standard chi square in order to match either the first or first and second moments of the weighted chi square. We will approximate the weighted chi square using the correction of the first two moments.

Let Z be from a multivariate normal distribution with mean 0 and variance Γ, and let A be a non negative definite matrix. Then

$$Q = Z'AZ = \sum_{i=1}^{l} \lambda_i {\chi^2}(1)$$

where λ_i for $i = 1, \ldots, l$ are the eigenvalues of $A\Gamma$. The distribution of Q can be approximated by

$$Q \sim a{\chi^2}(b)$$

where

$$a = \frac{\sum_{i=1}^{l} \lambda_i^2}{\sum_{i=1}^{l} \lambda_i},$$
$$b = \frac{\left(\sum_{i=1}^{l} \lambda_i\right)^2}{\sum_{i=1}^{l} \lambda_i^2}.$$

The eigenvalues can be estimated with a consistent estimator of $A\Gamma$. Γ can be consistently estimated with

$$\widehat{\Gamma}(\theta) = \begin{bmatrix} \frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i, \theta) g_x^{\intercal}(x_i, \theta) & \frac{1}{n_{xy}} \sum_i g_x(x_i, \theta) g_y^{\mathrm{T}}(y_i, \theta) \\ \frac{1}{n_{xy}} \sum_i g_y(y_i, \theta) g_x^{\mathrm{T}}(x_i, \theta) & \frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i, \theta) g_y^{\intercal}(y_i, \theta) \end{bmatrix}$$

with n_{xy} denoting the number of observed matched pairs of (x, y). The matrix A for the test statistic T in Theorem 2.2 can be consistently estimated with

$$\widehat{W}(\theta) = \left\{ \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i, \theta)}{\partial \theta} \right)^{\mathrm{T}} \left(\frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i, \theta) g_x^{\intercal}(x_i, \theta) \right)^{-1} \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i, \theta)}{\partial \theta} \right) + \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i, \theta)}{\partial \theta} \right)^{\mathrm{T}} \left(\frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i, \theta) g_y^{\intercal}(y_i, \theta) \right)^{-1} \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i, \theta)}{\partial \theta} \right) \right\}$$

$$\widehat{A}_{11}(\theta) = \left(\frac{1}{n_x}\sum_{i=1}^{n_x} g_x(x_i,\theta)g_x^{\intercal}(x_i,\theta)\right)^{-1}\left(\frac{1}{n_x}\sum_{i=1}^{n_x}\frac{\partial g_x(x_i,\theta)}{\partial\theta}\right)\widehat{W}(\theta)^{-1}$$
$$\times\left(\frac{1}{n_x}\sum_{i=1}^{n_x}\frac{\partial g_x(x_i,\theta)}{\partial\theta}\right)^{\mathrm{T}}\left(\frac{1}{n_x}\sum_{i=1}^{n_x} g_x(x_i,\theta)g_x^{\intercal}(x_i,\theta)\right)^{-1}$$

$$\widehat{A}_{22}(\theta) = \left(\frac{1}{n_y}\sum_{i=1}^{n_y} g_y(y_i,\theta)g_y^{\intercal}(y_i,\theta)\right)^{-1}\left(\frac{1}{n_y}\sum_{i=1}^{n_y}\frac{\partial g_y(y_i,\theta)}{\partial\theta}\right)\widehat{W}(\theta)^{-1}$$
$$\times\left(\frac{1}{n_y}\sum_{i=1}^{n_y}\frac{\partial g_y(y_i,\theta)}{\partial\theta}\right)^{\mathrm{T}}\left(\frac{1}{n_y}\sum_{i=1}^{n_y} g_y(y_i,\theta)g_y^{\intercal}(y_i,\theta)\right)^{-1}$$

$$\widehat{A}_{12}(\theta) = \left(\frac{1}{n_x}\sum_{i=1}^{n_x} g_x(x_i,\theta)g_x^{\intercal}(x_i,\theta)\right)^{-1}\left(\frac{1}{n_x}\sum_{i=1}^{n_x}\frac{\partial g_x(x_i,\theta)}{\partial\theta}\right)\widehat{W}(\theta)^{-1}$$
$$\times\left(\frac{1}{n_y}\sum_{i=1}^{n_y}\frac{\partial g_y(y_i,\theta)}{\partial\theta}\right)^{\mathrm{T}}\left(\frac{1}{n_y}\sum_{i=1}^{n_y} g_y(y_i,\theta)g_y^{\intercal}(y_i,\theta)\right)^{-1}$$

$$\widehat{A}_{21}(\theta) = \left(\frac{1}{n_y}\sum_{i=1}^{n_y} g_y(y_i,\theta)g_y^{\intercal}(y_i,\theta)\right)^{-1}\left(\frac{1}{n_y}\sum_{i=1}^{n_y}\frac{\partial g_y(y_i,\theta)}{\partial\theta}\right)\widehat{W}(\theta)^{-1}$$
$$\times\left(\frac{1}{n_x}\sum_{i=1}^{n_x}\frac{\partial g_x(x_i,\theta)}{\partial\theta}\right)^{\mathrm{T}}\left(\frac{1}{n_x}\sum_{i=1}^{n_x} g_x(x_i,\theta)g_x^{\intercal}(x_i,\theta)\right)^{-1}.$$

The value of θ can be set to the null hypothesis value θ_0 for hypothesis testing, or to simplify optimization for confidence intervals the MCELE $\hat{\theta}_{CE}$ can be used. We may know (or wish to assume) the covariance between the two variables, in which case the values can be substituted directly.

Finally by using the property that the trace of a matrix is equal to the sum of the eigenvalues, the estimators of a and b are

$$\hat{a} = tr[(\widehat{A}(\theta)\widehat{\Gamma}(\theta))^2]/tr[\widehat{A}(\theta)\widehat{\Gamma}(\theta)],$$
$$\hat{b} = tr[\widehat{A}(\theta)\widehat{\Gamma}(\theta)]^2/tr[(\widehat{A}(\theta)\widehat{\Gamma}(\theta))^2].$$

2.4 Simulations

2.4.1 Numerical Study of Theoretical Results

We numerically examine the mean and variance of $\hat{\theta}_{CE}$ and the Type I error using T for hypothesis tests. To accomplish this we assume that $E(g_x g_x^{\mathrm{T}})$, $E(g_y g_y^{\mathrm{T}})$ and $E(g_x g_y^{\mathrm{T}})$ are known, as are $E(\partial g_x/\partial\theta)$ and $E(\partial g_y/\partial\theta)$. Given this the true value of $A\Gamma$ is known, and as a result the exact distribution of the test statistic T is known. We will examine cases to confirm the results of Theorems 2.1, 2.2 and Corollary 2.1.

We examine three data distributions: bivariate normal, bivariate chi square, and bivariate uniform. We use 500 replicates of varying sample size n and correlation ρ. The bivariate normal random variables are generated with parameter values of $\mu_x = \mu_y = 1$ and $\sigma_x^2 = \sigma_y^2 = 2$. The bivariate chi square random variables are generated with parameter values $df_x = df_y = 1$. The bivariate uniform random variables are generated with parameter values $a_x = a_y = 1 - \sqrt{6}$ and $b_x = b_y = 1 + \sqrt{6}$. All data distributions have expected values of $\theta = 1$, variances of 2 and correlation of ρ. The chi square values are generated by squaring bivariate standard normal random variable with correlation of $\sqrt{\rho}$, and the uniform are generated using the inverse probability density functions of bivariate standard normal random variables with correlation ρ. The mean is the parameter of interest, so $g_x(x_i, \theta) = x_i - \theta$ and $g_y(y_i, \theta) = y_i - \theta$.

Working from Theorem 2.1 the estimator will have an asymptotic mean of 1 and an asymptotic variance of $n^{-1}(1+\rho)$. We use the sample mean and sample standard deviation to empirically estimate these values from the simulations, and we show the number of false rejections out of the 500 simulations at $\alpha = 0.10, 0.05, 0.01$. The weighted chi square distribution for the test statistic is estimated using the approach shown in Section 2.3, but since we know all the necessary values we have

$$T \longrightarrow (1+\rho)\chi^2(1)$$

as $n \longrightarrow \infty$.

Table 2.1: Mean and variance of parameter estimates and cumulative distribution of T at $\alpha = 0.10, 0.05, 0.01$ when data are bivariate normal.

ρ	n	Observed Mean	Observed Variance	$N_{p<0.10}$	$N_{p<0.05}$	$N_{p<0.01}$
0.00	10	1.010	0.115	104	72	34
	50	1.012	0.022	62	35	8
	100	1.002	0.011	60	32	6
0.10	10	1.009	0.124	101	70	33
	50	1.013	0.024	63	33	8
	100	1.002	0.012	61	30	7
0.50	10	1.006	0.155	97	62	26
	50	1.016	0.031	55	28	7
	100	1.001	0.016	53	31	8
0.90	10	1.005	0.186	85	56	24
	50	1.018	0.038	58	25	8
	100	1.001	0.020	55	28	6

Table 2.2: Mean and variance of parameter estimates and cumulative distribution of T at $\alpha = 0.10, 0.05, 0.01$ when data are bivariate chi square.

ρ	n	Observed Mean	Observed Variance	$N_{p<0.10}$	$N_{p<0.05}$	$N_{p<0.01}$
0.00	10	0.923	0.102	157	126	78
	50	0.978	0.018	62	39	11
	100	0.998	0.009	46	30	10
0.10	10	0.931	0.114	158	127	80
	50	0.981	0.021	74	44	12
	100	0.997	0.011	51	31	9
0.50	10	0.962	0.148	134	102	63
	50	0.988	0.030	75	40	13
	100	0.997	0.015	51	28	7
0.90	10	0.990	0.174	125	99	53
	50	0.991	0.037	66	40	14
	100	0.996	0.018	51	25	7

Table 2.3: Mean and variance of parameter estimates and cumulative distribution of T at $\alpha = 0.10, 0.05, 0.01$ when data are bivariate uniform.

ρ	n	Observed Mean	Observed Variance	$N_{p<0.10}$	$N_{p<0.05}$	$N_{p<0.01}$
0.00	10	1.037	0.113	81	55	22
	50	1.016	0.020	50	29	11
	100	1.010	0.010	43	27	7
0.10	10	1.038	0.119	79	53	22
	50	1.017	0.022	52	26	11
	100	1.010	0.011	43	26	7
0.50	10	1.038	0.144	70	41	15
	50	1.020	0.029	44	20	7
	100	1.010	0.014	43	29	9
0.90	10	1.033	0.174	62	30	14
	50	1.022	0.035	40	17	6
	100	1.008	0.019	53	29	4

Table 2.1 shows when the data are normal the observed mean of $\hat{\theta}_{CE}$ matches very closely with the theoretical expectations even at small sample sizes. Additionally the number of false rejections is very close to theoretical value when the sample size is 50 and 100. Table 2.2 shows when the data are chi square the false rejection percent is much higher than expected with small sample sizes, but as the sample size increases the false rejection rates approach the theoretical values. This is not unexpected given the skew of the chi square distribution. Table 2.3 shows when the data are uniform the mean and variance of the estimator are very close to theoretical expectations at small sample sizes. The Type I errors using T are much higher when the sample size is small, but still approach the theoretical expectations as the sample size increases. All three simulations confirm the theoretical results shown in Theorems 2.1, 2.2 and Corollary 2.1.

2.4.2 Comparison to Empirical Likelihood

We now examine the case where we have to estimate $A\Gamma$ from the data, and compare the composite empirical likelihood to two empirical likelihood setups. As in the previous section, we focus on the mean parameter for ease of demonstration.

We consider the following three functions

$$L_{CE}(\theta) = \left\{ \sup_u \left(\prod_{i=1}^n u_i | \sum_{i=1}^n u_i x_i = \theta,\ u_i \geq 0,\ \sum_{i=1}^n u_i = 1 \right) \times \right.$$
$$\left. \sup_v \left(\prod_{i=1}^n v_i | \sum_{i=1}^n v_i y_i = \theta,\ v_i \geq 0,\ \sum_{i=1}^n v_i = 1 \right) \right\}$$
$$L_{E1}(\theta) = \sup_p \left(\prod_{i=1}^{2n} p_i | \sum_{i=1}^n p_i x_i + \sum_{i=1}^n p_{n+i} y_i = \theta,\ p_i \geq 0,\ \sum_{i=1}^{2n} p_i = 1 \right)$$
$$L_{E2}(\theta) = \sup_p \left(\prod_{i=1}^n p_i | \sum_{i=1}^n p_i x_i = \theta,\ \sum_{i=1}^n p_i y_i = \theta,\ p_i \geq 0,\ \sum_{i=1}^n p_i = 1 \right).$$

$L_{E1}(\theta)$ is the empirical likelihood form used by Owen (1988) if we combine the two data vectors into a single vector. $L_{E2}(\theta)$ is the form explored in Owen (2001) when two variables have a shared mean.

Since we do not assume knowledge of the covariance and expected value of the derivatives of g_x and g_y we estimate A and Γ given a value of θ using the sample versions shown in Section 2.3. Similar to parametric tests we will use $\theta = \hat{\theta}_{CE}$ when working with confidence intervals and $\theta = \theta_0$ for tests of hypothesis in order to estimate a and b. Theorem 2.2 indicates that $A\Gamma$ is dependent on our null hypothesis value, so for confidence intervals we should use the two values of θ that define the endpoints in order to compute $\hat{a}$ and $\hat{b}$. However this is problematic since the weighted chi square that determine the endpoints are dependent on the values of θ, requiring additional computations during the optimization step to compute the confidence intervals. Instead we can use $\hat{\theta}$, since it is a consistent estimator of θ, to determine the weighted chi square (see Barndorff-Nielsen and Cox, 1994, page 91). For the hypothesis test we use $\theta_0 = 1$ in place of $\hat{\theta}$ to estimate a and b.

To examine how the three functions compare we generate 500 realizations from bivariate normal, bivariate chi square and bivariate uniform all with varying correlation values. All three distributions use the same parameter values from the simulations in Section 2.4.1, so all the variables have a mean of 1 and variance of 2. We compare the mean and variance of the estimator, the length of the 95% confidence interval, how many times the lower and upper endpoints do not cover the true value $\theta = 1$ and how many times the test of $H_0 : \theta = 1$ is rejected at $\alpha = 0.05$.

Table 2.4: Comparison of L_{CE}, L_{E1} and L_{E2} using bivariate normal with 500 replicates.

ρ	n	Method	Observed Mean	Observed Variance	Average Length	$N_{L>1}$	$N_{U<1}$	$N_{p<0.05}$
0.00	25	L_{CE}	1.005	0.044	0.761	23	22	40
		L_{E1}	1.006	0.043	0.786	22	12	34
		L_{E2}	1.005	0.045	1.481	20	14	46
	100	L_{CE}	0.993	0.011	0.390	18	17	33
		L_{E1}	0.993	0.011	0.392	16	20	36
		L_{E2}	0.993	0.011	2.193	13	10	35
0.10	25	L_{CE}	0.995	0.047	0.791	19	22	40
		L_{E1}	0.997	0.046	0.783	20	18	38
		L_{E2}	0.995	0.048	1.376	20	19	43
	100	L_{CE}	1.000	0.011	0.409	16	14	29
		L_{E1}	1.000	0.011	0.392	20	14	34
		L_{E2}	1.000	0.011	1.815	8	8	28
0.50	25	L_{CE}	1.006	0.062	0.921	18	19	34
		L_{E1}	1.010	0.061	0.772	33	29	62
		L_{E2}	1.001	0.067	1.017	16	21	37
	100	L_{CE}	0.995	0.015	0.479	15	11	26
		L_{E1}	0.995	0.015	0.392	27	27	54
		L_{E2}	0.995	0.015	0.739	15	11	27
0.90	25	L_{CE}	0.982	0.084	1.058	21	19	40
		L_{E1}	0.981	0.084	0.767	40	53	93
		L_{E2}	0.984	0.087	1.099	22	21	44
	100	L_{CE}	1.005	0.018	0.537	14	11	25
		L_{E1}	1.005	0.018	0.388	35	33	68
		L_{E2}	1.005	0.018	0.566	13	11	24

Table 2.5: Comparison of L_{CE}, L_{E1} and L_{E2} using bivariate chi square with 500 replicates.

ρ	n	Method	Observed Mean	Observed Variance	Average Length	$N_{L>1}$	$N_{U<1}$	$N_{p<0.05}$
0.00	25	L_{CE}	0.976	0.044	0.684	14	49	61
		L_{E1}	0.999	0.041	0.783	10	26	36
		L_{E2}	0.974	0.044	1.981	13	39	66
	100	L_{CE}	0.997	0.009	0.382	9	25	33
		L_{E1}	1.001	0.009	0.394	7	19	26
		L_{E2}	0.997	0.009	4.123	7	8	32
0.10	25	L_{CE}	0.986	0.046	0.715	9	54	58
		L_{E1}	1.006	0.044	0.775	15	31	46
		L_{E2}	0.983	0.048	1.840	8	45	64
	100	L_{CE}	1.000	0.012	0.403	13	22	33
		L_{E1}	1.002	0.012	0.395	16	21	37
		L_{E2}	1.000	0.012	3.644	15	8	36
0.50	25	L_{CE}	1.012	0.060	0.883	12	36	45
		L_{E1}	1.022	0.061	0.782	24	40	64
		L_{E2}	0.998	0.060	1.601	11	37	51
	100	L_{CE}	1.000	0.015	0.478	9	14	23
		L_{E1}	1.002	0.015	0.396	24	26	50
		L_{E2}	0.999	0.015	2.369	10	13	26
0.90	25	L_{CE}	1.006	0.076	1.029	10	40	48
		L_{E1}	1.009	0.076	0.757	41	55	96
		L_{E2}	0.986	0.082	1.886	13	44	71
	100	L_{CE}	0.989	0.018	0.535	12	20	31
		L_{E1}	0.990	0.018	0.386	29	55	84
		L_{E2}	0.984	0.018	1.998	12	23	39

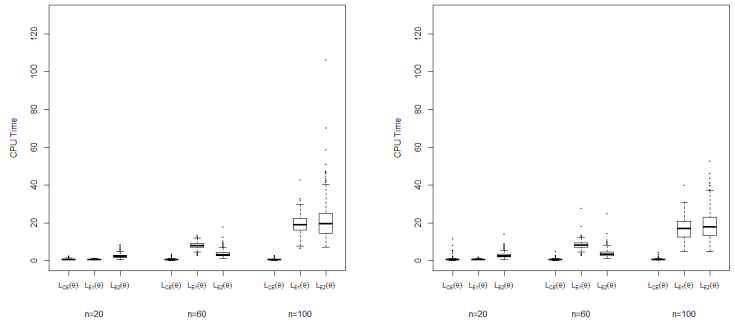
Table 2.6: Comparison of L_{CE}, L_{E1} and L_{E2} using bivariate uniform with 500 replicates.

ρ	n	Method	Observed Mean	Observed Variance	Average Length	$N_{L>1}$	$N_{U<1}$	$N_{p<0.05}$
0.00	25	L_{CE}	0.999	0.040	0.762	18	11	27
		L_{E1}	1.001	0.038	0.771	12	9	21
		L_{E2}	0.999	0.040	0.857	11	10	30
	100	L_{CE}	1.005	0.009	0.388	8	14	21
		L_{E1}	1.005	0.009	0.390	10	12	22
		L_{E2}	1.005	0.009	0.390	8	13	21
0.10	25	L_{CE}	0.992	0.042	0.792	10	16	23
		L_{E1}	0.992	0.041	0.772	12	15	27
		L_{E2}	0.993	0.043	0.874	9	8	27
	100	L_{CE}	1.010	0.012	0.408	24	12	33
		L_{E1}	1.010	0.012	0.390	27	15	42
		L_{E2}	1.010	0.012	0.409	23	12	35
0.50	25	L_{CE}	0.983	0.057	0.916	9	16	20
		L_{E1}	0.983	0.056	0.763	24	33	57
		L_{E2}	0.982	0.058	0.922	7	17	24
	100	L_{CE}	0.992	0.015	0.475	11	13	23
		L_{E1}	0.991	0.015	0.390	25	32	57
		L_{E2}	0.992	0.015	0.477	11	12	23
0.90	25	L_{CE}	0.985	0.070	1.051	9	11	20
		L_{E1}	0.985	0.071	0.769	32	39	71
		L_{E2}	0.985	0.072	1.050	10	13	23
	100	L_{CE}	1.003	0.021	0.536	21	8	29
		L_{E1}	1.003	0.021	0.390	47	41	88
		L_{E2}	1.003	0.021	0.537	21	7	28

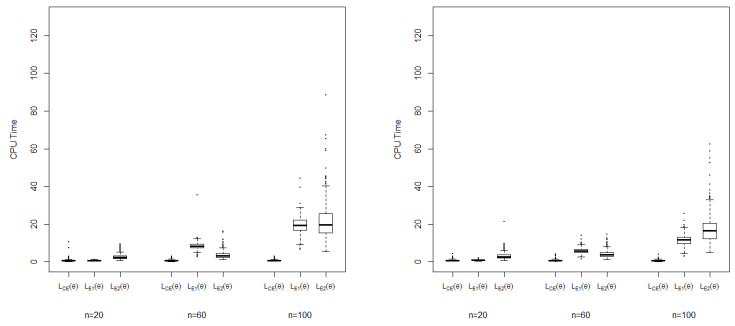
Tables 2.4, 2.5 and 2.6 show similar patterns. When there is no correlation L_{E1} performs better both in terms of standard deviation of the estimator and number of false rejections for the confidence interval and hypothesis test. This is not unexpected since in this case the bivariate data can be treated as one univariate sample due to the independence between the two variables. As the correlation increases L_{CE} performs better than L_{E1} and L_{E2} in terms of false rejections of both the confidence intervals and hypothesis tests. The length of the confidence interval of L_{E2} varies depending on the data generating mechanism and the correlation. Specifically when the data are from the uniform distribution larger sample size decreases the length, while the length increases as sample size increases when the data are generated from the chi square. In the case of normal data the length of the confidence interval increases as sample size increases when the correlation 0.00 and 0.10, but deceases as the sample size increases when the correlation is 0.50 and 0.90. L_{CE} outperforms L_{E1} and performs comparably to L_{E2} when there is data correlation. Given these results it does not appear that estimation of $A\Gamma$ has a negative impact on the performance of L_{CE}.

2.4.3 Computational Time

Another question of interest is computational gains using composite empirical likelihood versus other empirical likelihoods. To demonstrate this we examine the CPU time to find the value $\hat{\theta}_{CE}$ using L_{CE} in parallel using two local workers, and compare to the CPU time using L_{E1} and L_{E2}. The `fmincon` function in MatLab is used for optimization, and a starting value of 1 is used for all three likelihood functions. The data are independent bivariate chi square random variables with 1 degree of freedom and varying sample sizes and correlations.



$\rho = 0.00$ $\rho = 0.10$



$\rho = 0.50$ $\rho = 0.90$

Figure 2.1: Distribution of computation times of L_{CE}, L_{E1} and L_{E2} using bivariate chi-square with 500 replicates.

Figure 2.1 shows that as the sample size increases the CPU times of both L_{E1} and L_{E2} noticeably increase, while the CPU time using composite empirical likelihood does not appreciably increase. Also the variation in computation time is consistent using L_{CE} regardless of the sample size whereas the other methods have an increased variation of computation time as the sample size goes up.

2.5 Discussion

The theoretical results can be viewed as a generalization of likelihood theory since we define the composite empirical likelihood with no specification of data distribution and separation of functions. As Corollary 2.1 shows, given necessary conditions the composite empirical likelihood exhibits the asymptotic properties seen in Fisher likelihood, empirical likelihood and several cases of composite likelihood. The composite empirical likelihood with a single piece is equivalent to the standard empirical likelihood which in turn can be viewed as a general case of the Fisher likelihood.

Analogous to the common practice of redefining bivariate distributions in terms of marginal components to simplify computation, we have demonstrated an equivalent method with empirical likelihood. The computational simplification is not seen in the form of the equation, but rather in the ability to calculate each piece in separate steps. In a parallel computing environment this feature of the composite empirical likelihood gives it a computational time advantage over the current forms.

Chapter 3

Alternate Estimating Equation Forms for Dependent Data

3.1 Introduction

A confidence region is a high dimensional equivalent of a confidence interval; it contains all values of the parameters simultaneously at a specified $(1-\alpha)\%$ level of confidence. Specifically confidence regions account for both multiple testing (since there is more than one parameter) and data correlation. Creating a region using sets of univariate confidence intervals would cover a larger area than a properly constructed confidence region, implying that any inference using multiple intervals simultaneously would not be at the claimed level of confidence. Confidence regions can be constructed for any number of parameters provided the sample size is sufficiently larger than the dimension of the parameter space, but due to visualization constraints we are usually limited to bivariate confidence regions.

In Chapter 2 we establish the distribution of the test statistic T for a two component composite empirical likelihood. This result suggests that a confidence region for two mean parameters consists of all the values of μ_x' and μ_y' such that

$$2\ell_{CE}(\mu_x', \mu_y') - 2\ell_{CE}(\hat{\mu}_x, \hat{\mu}_y) \leq Q(\lambda)_{1-\alpha}$$

with the boundary being at the equality. However the construction of the composite empirical likelihood does not inherently account for dependency between the two random variables, and as a result the confidence region may not accurately reflect the area pertaining to the stated level of confidence. If we define the estimating equations as

$$g_x = x_i - \mu_x$$

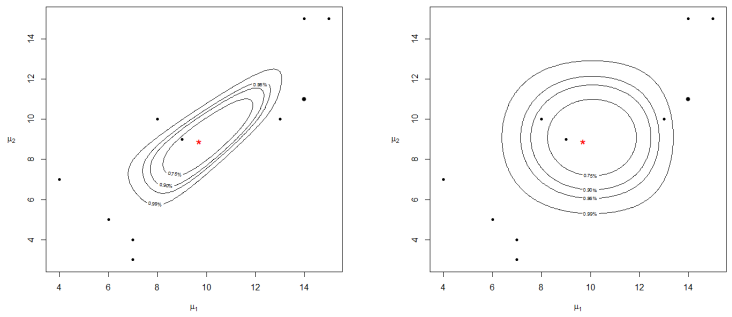
$$g_y = y_i - \mu_y$$

then we have for a fixed μ_y^0

$$\begin{aligned} T &= 2\ell_{CE}(\mu_x', \mu_y^0) - 2\ell_{CE}(\hat{\mu}_x, \hat{\mu}_y) \\ &= 2\ell_E(\mu_x') - 2\ell_E(\hat{\mu}_x) + 2(\ell_E(\mu_y^0) - 2\ell_E(\hat{\mu}_y)) \\ &= 2\ell_E(\mu_x') - 2\ell_E(\hat{\mu}_x) + c \end{aligned}$$

where c is a constant. So for any μ_y^0 the length of the interval in terms of μ_x will change, but the center of the interval remains fixed at $\hat{\mu}_x$, resulting in the region potentially failing to cover the proper $(1-\alpha)\%$ area.

We demonstrate this limitation using data from Larsen and Marx (1986), from which a confidence region for the means is explored in Owen (1990). Eleven male second generation crosses between a mallard and pintail duck were examined, with plumage rated on a scale from 0 (completely mallard like) to 20 (completely pintail like). Their behavior was also rated on a scale of 0 to 15, with 0 indicating completely mallard like and 15 indicating completely pintail like. We compare confidence regions using empirical likelihood from Owen (1990) (which we denote as L_E) against those created using composite empirical likelihood (L_{CE}) utilizing the estimating equations shown above. The plots show the regions at the following confidence levels: 75%, 90%, 95% and 99%. We determine the boundary of the empirical likelihood with a chi square with two degrees of freedom, and for the composite empirical likelihood we estimate the distribution of T using the approximation shown in Section 2.3.



L_E L_{CE}

Figure 3.1: Comparison of empirical likelihood and composite empirical likelihood for behavioral and plumage characteristics of hybrid ducks data. The asterisk (in red) denotes the sample mean of the two variables, and the larger sized point at $14, 11$ denotes that there were two observations with those values.

Figure 3.1 shows that the composite empirical likelihood does not account for the correlation structure between the two variables, and as a result the confidence region is approximately circular (which is what we would see if the data were independent). Even though we cannot introduce information about the correlation directly into the composite empirical likelihood, we can define the estimating equations in a conditional-marginal or conditional-conditional form. Each form makes different assumptions about the data generating mechanism, but both connect the two likelihood components in order to capture the dependency between the variables. We will explore both of these constructions in the context of a linear relationship of the means between two variables.

3.2 Distribution of Test Statistic with Nuisance Parameter

Describing conditional moment relationships generally requires the addition of nuisance parameters. In Chapter 2 we examined the distribution of the test statistic when inference was on the entire parameter set θ, so we start by exploring the general case of the composite empirical likelihood ratio test statistic when we incorporate nuisance parameters.

Theorem 3.1 *Let $\theta = [\phi^{\mathrm{T}}, \nu^{\mathrm{T}}]^{\mathrm{T}}$ be a p dimensional vector where ϕ is a $q \times 1$ vector and ν is a $(p-q) \times 1$ vector. Under Assumptions 1 and 2 from Section 2.2 the profile composite empirical likelihood ratio test statistic to test $H_0 : \phi = \phi_0$ is*

$$T_2 = 2\ell_{CE}(\phi_0, \hat{\nu}_{CE}(\phi_0)) - 2\ell_{CE}(\hat{\phi}_{CE}, \hat{\nu}_{CE}) \tag{3.1}$$

where $\hat{\nu}_{CE}(\phi_0)$ minimizes $\ell_{CE}(\phi_0, \nu)$ with respect to ν. Under H_0 we have

$$T_2 \longrightarrow Q(\lambda)$$

as $n \longrightarrow \infty$. $Q(\lambda) = \sum_{i=1}^{l} \lambda_i {\chi^2}(1)$ is a weighted chi square random variable where λ_i for $i = 1, \ldots, l$ is the set of all non zero eigenvalues of

$$A_2\Gamma = \begin{bmatrix} A_{xx2} & A_{xy2} \\ A_{yx2} & A_{yy2} \end{bmatrix} \begin{bmatrix} E(g_x g_x^{\mathrm{T}}) & E(g_x g_y^{\mathrm{T}}) \\ E(g_y g_x^{\mathrm{T}}) & E(g_y g_y^{\mathrm{T}}) \end{bmatrix}$$

where

$$A_{2xx} = \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1}\left\{E\left(\frac{\partial g_x}{\partial\theta}\right)W^{-1}E\left(\frac{\partial g_x}{\partial\theta}\right)^{\mathrm{T}} - E\left(\frac{\partial g_x}{\partial\nu}\right)W_2^{-1}E\left(\frac{\partial g_x}{\partial\nu}\right)^{\mathrm{T}}\right\}\left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1}$$

$$A_{2yy} = \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1}\left\{E\left(\frac{\partial g_y}{\partial\theta}\right)W^{-1}E\left(\frac{\partial g_y}{\partial\theta}\right)^{\mathrm{T}} - E\left(\frac{\partial g_y}{\partial\nu}\right)W_2^{-1}E\left(\frac{\partial g_y}{\partial\nu}\right)^{\mathrm{T}}\right\}\left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1}$$

$$A_{2xy} = \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1}\left\{E\left(\frac{\partial g_x}{\partial\theta}\right)W^{-1}E\left(\frac{\partial g_y}{\partial\theta}\right)^{\mathrm{T}} - E\left(\frac{\partial g_x}{\partial\nu}\right)W_2^{-1}E\left(\frac{\partial g_y}{\partial\nu}\right)^{\mathrm{T}}\right\}\left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1}$$

$$A_{2yx} = \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1}\left\{E\left(\frac{\partial g_y}{\partial\theta}\right)W^{-1}E\left(\frac{\partial g_x}{\partial\theta}\right)^{\mathrm{T}} - E\left(\frac{\partial g_y}{\partial\nu}\right)W_2^{-1}E\left(\frac{\partial g_x}{\partial\nu}\right)^{\mathrm{T}}\right\}\left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1}$$

$$W_2 = \left\{E\left(\frac{\partial g_x}{\partial\nu}\right)^{\mathrm{T}}\left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1}E\left(\frac{\partial g_x}{\partial\nu}\right) + E\left(\frac{\partial g_y}{\partial\nu}\right)^{\mathrm{T}}\left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1}E\left(\frac{\partial g_y}{\partial\nu}\right)\right\}$$

and W is as defined in Theorem 2.1.

Proof. In terms of ϕ and ν, Q_{1x} and Q_{1y} are

$$Q_{1x}(\phi,\nu,t_x) = \frac{1}{n_x}\sum_{i=1}^{n_x}(1+t_x^{\mathrm{T}}g_x(x_i,\phi,\nu))^{-1}g_x(x_i,\phi,\nu),$$

$$Q_{1y}(\phi,\nu,t_y) = \frac{1}{n_y}\sum_{i=1}^{n_y}(1+t_y^{\mathrm{T}}g_y(y_i,\phi,\nu))^{-1}g_y(y_i,\phi,\nu).$$

Using the same steps as in Lemma 2.2, it follows that for a fixed value ϕ there is a $\hat{t}_x$, $\hat{t}_y$ and $\hat{\nu}_{CE}(\phi)$ such that $\ell_{CE}(\phi,\hat{\nu}_{CE}(\phi))$ obtains a minimum value. Furthermore $\hat{t}_x$, $\hat{t}_y$, and $\hat{\nu}_{CE}(\phi)$ must satisfy

$$Q_{1x}(\phi,\hat{\nu}_{CE}(\phi),\hat{t}_x) = 0$$
$$Q_{1y}(\phi,\hat{\nu}_{CE}(\phi),\hat{t}_y) = 0.$$

Redefine Q_2 as

$$Q_2(\phi,\nu,t_x,t_y) = \frac{\partial\ell_{CE}(\phi,\nu)}{\partial\nu}$$

and we have when $\nu = \hat{\nu}_{CE}(\phi)$

$$\begin{aligned}
\left.\frac{\partial \ell_{CE}(\phi,\nu)}{\partial \nu}\right|_{\nu=\hat{\nu}_{CE}(\phi)} &= \left.\sum_{i=1}^{n_x} \frac{(\partial t_x^{\mathrm{T}}(\nu(\phi))/\partial\nu) g_x(x_i,\phi,\nu) + (\partial g_x(x_i,\phi,\nu)/\partial\nu)^{\mathrm{T}} t_x(\nu(\phi))}{1 + t_x^{\mathrm{T}}(\nu(\phi)) g_x(x_i,\phi,\nu)}\right|_{\nu=\hat{\nu}_{CE}(\phi)} \\
&\quad + \left.\sum_{i=1}^{n_y} \frac{(\partial t_y^{\mathrm{T}}(\nu(\phi))/\partial\nu) g_y(y_i,\phi,\nu) + (\partial g_y(y_i,\phi,\nu)/\partial\nu)^{\mathrm{T}} t_y(\nu(\phi))}{1 + t_y^{\mathrm{T}}(\nu(\phi)) g_y(y_i,\phi,\nu)}\right|_{\nu=\hat{\nu}_{CE}(\phi)} \\
&= \left.\sum_{i=1}^{n_x} \frac{1}{1 + t_x^{\mathrm{T}}(\nu(\phi)) g_x(x_i,\phi,\nu)} \left(\frac{\partial g_x(x_i,\phi,\nu)}{\partial\nu}\right)^{\mathrm{T}} t_x(\nu(\phi))\right|_{\nu=\hat{\nu}_{CE}(\phi)} \\
&\quad + \left.\sum_{i=1}^{n_y} \frac{1}{1 + t_y^{\mathrm{T}}(\nu(\phi)) g_y(y_i,\phi,\nu)} \left(\frac{\partial g_y(y_i,\phi,\nu)}{\partial\nu}\right)^{\mathrm{T}} t_y(\nu(\phi))\right|_{\nu=\hat{\nu}_{CE}(\phi)} \\
&= 0.
\end{aligned}$$

Following the proof of Theorem 2.1 we can show

$$\begin{aligned}
(\hat{\nu}_{CE}(\phi) - \nu_0) = &\left\{ E\left(\frac{\partial g_x}{\partial\nu}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_x}{\partial\nu}\right) + E\left(\frac{\partial g_y}{\partial\nu}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_y}{\partial\nu}\right)\right\}^{-1} \\
&\times \left\{ - E\left(\frac{\partial g_x}{\partial\nu}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} Q_{1x}(\theta_0, 0) - E\left(\frac{\partial g_y}{\partial\nu}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} Q_{1y}(\theta_0, 0)\right\} \\
&+ o_p(1). \qquad (3.2)
\end{aligned}$$

Using a Taylor expansion gives

$$\begin{aligned}
Q_{1x}(\phi_0, \hat{\nu}_{CE}(\phi), 0) = &\, Q_{1x}(\theta_0, 0) \\
&- E\left(\frac{\partial g_x}{\partial\nu}\right) W_2^{-1} E\left(\frac{\partial g_x}{\partial\nu}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} Q_{1x}(\theta_0, 0) \\
&- E\left(\frac{\partial g_x}{\partial\nu}\right) W_2^{-1} E\left(\frac{\partial g_y}{\partial\nu}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} Q_{1y}(\theta_0, 0) \\
&+ o_p(1) \qquad (3.3)
\end{aligned}$$

and

$$Q_{1y}(\phi_0, \hat{\nu}_{CE}(\phi), 0) = Q_{1y}(\theta_0, 0)$$

$$
\begin{aligned}
&- E\left(\frac{\partial g_y}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_y}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} Q_{1y}(\theta_0, 0) \\
&- E\left(\frac{\partial g_y}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_x}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} Q_{1x}(\theta_0, 0) \\
&+ o_p(1), \qquad (3.4)
\end{aligned}
$$

where W_2 is as defined previously, and $\theta_0 = [\phi_0^{\mathrm{T}}, \nu_0^{\mathrm{T}}]^{\mathrm{T}}$.

Equations 3.2, 3.3 and 3.4 allow us to establish

$$
\begin{aligned}
\ell_{CE}(\phi_0, \hat{\nu}_{CE}(\phi)) = &\frac{n}{2} Q_{1x}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} Q_{1x}(\theta_0, 0) + \frac{n}{2} Q_{1y}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} Q_{1y}(\theta_0, 0) \\
&- \frac{n}{2} Q_{1x}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_x}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_x}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} Q_{1x}(\theta_0, 0) \\
&- \frac{n}{2} Q_{1y}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_y}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_x}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} Q_{1y}(\theta_0, 0) \\
&- \frac{n}{2} Q_{1x}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_x}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_y}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} Q_{1y}(\theta_0, 0) \\
&- \frac{n}{2} Q_{1y}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_y}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_x}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} Q_{1x}(\theta_0, 0) \\
&+ o_p(1).
\end{aligned}
$$

Theorem 2.2 established

$$
\begin{aligned}
\ell_{CE}(\hat{\phi}_{CE}, \hat{\nu}_{CE}) &= \ell_{CE}(\hat{\theta}_{CE}) \\
&= \frac{n}{2} Q_{1x}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} Q_{1x}(\theta_0, 0) + \frac{n}{2} Q_{1y}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} Q_{1y}(\theta_0, 0) \\
&\quad - \frac{n}{2} Q_{1x}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} Q_{1x}(\theta_0, 0) \\
&\quad - \frac{n}{2} Q_{1y}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} Q_{1y}(\theta_0, 0) \\
&\quad - \frac{n}{2} Q_{1x}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} Q_{1y}(\theta_0, 0) \\
&\quad - \frac{n}{2} Q_{1y}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1} Q_{1x}(\theta_0, 0) \\
&\quad + o_p(1).
\end{aligned}
$$

The remainder of the proof follows from the same arguments used in Theorem 2.2. □

Like Theorem 2.2, the asymptotic distribution of the test statistic requires that we know the quadratic form matrix A_2. The approximation of A_2 for the test statistic T_2 in Theorem 3.1 can be consistently estimated with

$$\widehat{W}(\theta) = \left\{ \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\theta)}{\partial \theta} \right)^{\mathrm{T}} \left(\frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i,\theta) g_x^{\intercal}(x_i,\theta) \right)^{-1} \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\theta)}{\partial \theta} \right) + \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\theta)}{\partial \theta} \right)^{\mathrm{T}} \left(\frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta) g_y^{\intercal}(y_i,\theta) \right)^{-1} \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\theta)}{\partial \theta} \right) \right\}$$

$$\widehat{W}_2(\theta) = \left\{ \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\nu)}{\partial \nu} \right)^{\mathrm{T}} \left(\frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i,\theta) g_x^{\intercal}(x_i,\theta) \right)^{-1} \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\nu)}{\partial \nu} \right) + \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\nu)}{\partial \nu} \right)^{\mathrm{T}} \left(\frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta) g_y^{\intercal}(y_i,\theta) \right)^{-1} \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\nu)}{\partial \nu} \right) \right\}$$

$$\begin{aligned} \widehat{A}_{11}(\theta) = & \left(\frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i,\theta) g_x^{\intercal}(x_i,\theta) \right)^{-1} \\ & \times \left\{ \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\theta)}{\partial \theta} \right) \widehat{W}(\theta)^{-1} \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\theta)}{\partial \theta} \right)^{\mathrm{T}} - \right. \\ & \left. \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\nu)}{\partial \nu} \right) \widehat{W}_2(\theta)^{-1} \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\nu)}{\partial \nu} \right)^{\mathrm{T}} \right\} \\ & \times \left(\frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i,\theta) g_x^{\intercal}(x_i,\theta) \right)^{-1} \end{aligned}$$

$$\begin{aligned} \widehat{A}_{22}(\theta) = & \left(\frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta) g_y^{\intercal}(y_i,\theta) \right)^{-1} \\ & \times \left\{ \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\theta)}{\partial \theta} \right) \widehat{W}(\theta)^{-1} \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\theta)}{\partial \theta} \right)^{\mathrm{T}} - \right. \\ & \left. \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\nu)}{\partial \nu} \right) \widehat{W}_2(\theta)^{-1} \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\nu)}{\partial \nu} \right)^{\mathrm{T}} \right\} \\ & \times \left(\frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta) g_y^{\intercal}(y_i,\theta) \right)^{-1} \end{aligned}$$

$$\widehat{A}_{12}(\theta) = \left(\frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i,\theta) g_x^{\intercal}(x_i,\theta) \right)^{-1}$$

$$\times \left\{ \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\theta)}{\partial \theta} \right) \widehat{W}(\theta)^{-1} \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\theta)}{\partial \theta} \right)^{\mathrm{T}} - \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\nu)}{\partial \nu} \right) \widehat{W}_2^{-1}(\theta) \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\nu)}{\partial \nu} \right)^{\mathrm{T}} \right\}$$
$$\times \left(\frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta) g_y^{\intercal}(y_i,\theta) \right)^{-1}$$

$$\widehat{A}_{21}(\theta) = \left(\frac{1}{n_y} \sum_{i=1}^{n_y} g_y(y_i,\theta) g_y^{\intercal}(y_i,\theta) \right)^{-1}$$
$$\times \left\{ \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\theta)}{\partial \theta} \right) \widehat{W}(\theta)^{-1} \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\theta)}{\partial \theta} \right)^{\mathrm{T}} - \left(\frac{1}{n_y} \sum_{i=1}^{n_y} \frac{\partial g_y(y_i,\nu)}{\partial \nu} \right) \widehat{W}_2(\theta)^{-1} \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \frac{\partial g_x(x_i,\nu)}{\partial \nu} \right)^{\mathrm{T}} \right\}$$
$$\times \left(\frac{1}{n_x} \sum_{i=1}^{n_x} g_x(x_i,\theta) g_x^{\intercal}(x_i,\theta) \right)^{-1}.$$

The covariance matrix and the weighted chi square are estimated using the formulation shown in Section 2.3. If the parameter values are unknown we can use $\hat{\theta}_{CE} = [\hat{\phi}^{\mathrm{T}}, \hat{\nu}^{\mathrm{T}}]$, or in the case of hypothesis testing ϕ_0 and $\hat{\nu}_{CE}(\phi_0)$.

As we saw in Chapter 2, if the estimating functions are not correlated the asymptotic distribution of the test statistic is a standard chi square with the degrees of freedom being the number of parameters of interest. The following corollary establishes when the estimating functions are uncorrelated the test statistic to test $H_0 : \phi = \phi_0$ is asymptotically chi square.

Corollary 3.1 *Given Assumptions 1 and 2 from Section 2.2 and* $cov(g_x, g_y) = 0$

$$T_2 \longrightarrow \chi^2(q)$$

as $n \to \infty$ *under* H_0. T_2 *is given by Equation 3.1.*

Proof. Rewrite the test statistic as

$$T_2 = \sqrt{n} Q_{1x}^{\mathrm{T}}(\theta_0, 0) \left\{ E(g_x g_x^{\mathrm{T}}) \right\}^{-1/2}$$
$$\times \left\{ E(g_x g_x^{\mathrm{T}}) \right\}^{-1/2} \left\{ E\left(\frac{\partial g_x}{\partial \theta} \right) W^{-1} E\left(\frac{\partial g_x}{\partial \theta} \right)^{\mathrm{T}} - E\left(\frac{\partial g_x}{\partial \nu} \right) W_2^{-1} E\left(\frac{\partial g_x}{\partial \nu} \right)^{\mathrm{T}} \right\} \left\{ E(g_x g_x^{\mathrm{T}}) \right\}^{-1/2}$$

$$
\begin{aligned}
&\times \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} \sqrt{n} Q_{1x}(\theta_0, 0)^{\mathrm{T}} \\
&+ \sqrt{n} Q_{1y}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} \\
&\times \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} \left\{ E\left(\frac{\partial g_y}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} - E\left(\frac{\partial g_y}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_y}{\partial \nu}\right)^{\mathrm{T}} \right\} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} \\
&\times \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} \sqrt{n} Q_{1y}(\theta_0, 0) \\
&+ \sqrt{n} Q_{1x}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} \\
&\times \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} \left\{ E\left(\frac{\partial g_x}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} - E\left(\frac{\partial g_x}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_y}{\partial \nu}\right)^{\mathrm{T}} \right\} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} \\
&\times \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} \sqrt{n} Q_{1y}(\theta_0, 0) \\
&+ \sqrt{n} Q_{1y}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} \\
&\times \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} \left\{ E\left(\frac{\partial g_y}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} - E\left(\frac{\partial g_y}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_x}{\partial \nu}\right)^{\mathrm{T}} \right\} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} \\
&\times \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} \sqrt{n} Q_{1x}(\theta_0, 0) \\
&+ o_p(1).
\end{aligned}
$$

$\left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} \sqrt{n} Q_{1x}(\theta_0, 0)$ and $\left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} \sqrt{n} Q_{1y}(\theta_0, 0)$ are asymptotically independent standard multivariate normal random variables. The quadratic form matrix is

$$
A_2 = A - B
$$

where

$$
A = \left[\begin{array}{c:c}
\begin{array}{c} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g_x}{\partial \theta}\right) \\ \times W^{-1} \times \\ E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} \end{array} &
\begin{array}{c} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g_x}{\partial \theta}\right) \\ \times W^{-1} \times \\ E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} \end{array} \\
\hdashline
\begin{array}{c} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g_y}{\partial \theta}\right) \\ \times W^{-1} \times \\ E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} \end{array} &
\begin{array}{c} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g_y}{\partial \theta}\right) \\ \times W^{-1} \times \\ E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} \end{array}
\end{array}\right],
$$

$$B = \left[\begin{array}{c|c} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g_x}{\partial \nu}\right) & \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g_x}{\partial \nu}\right) \\ \times W_2^{-1} \times & \times W_2^{-1} \times \\ E\left(\frac{\partial g_x}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} & E\left(\frac{\partial g_y}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} \\ \hline \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g_y}{\partial \nu}\right) & \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g_y}{\partial \nu}\right) \\ \times W_2^{-1} \times & \times W_2^{-1} \times \\ E\left(\frac{\partial g_x}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_x g_x^{\mathrm{T}})\right\}^{-1/2} & E\left(\frac{\partial g_y}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_y g_y^{\mathrm{T}})\right\}^{-1/2} \end{array}\right]$$

so if $A - B$ is idempotent then T_2 is asymptotically chi square with degrees of freedom equal to the trace of $A - B$.

Corollary 2.1 establishes that A is idempotent. We can show B is idempotent using the same steps. For $A - B$ to be idempotent we only need establish that $A - B$ is non negative definite (see Rao, 1973, page 187). In fact

$$\begin{aligned} A_{xx} &\propto E\left(\frac{\partial g_x}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} \\ &\geq \left[E\left(\frac{\partial g_x}{\partial \phi}\right), E\left(\frac{\partial g_x}{\partial \nu}\right)\right] \left[\begin{array}{cc} 0 & 0 \\ 0 & W_2^{-1} \end{array}\right] \left[\begin{array}{c} E\left(\frac{\partial g_x}{\partial \phi}\right)^{\mathrm{T}} \\ E\left(\frac{\partial g_x}{\partial \nu}\right)^{\mathrm{T}} \end{array}\right] \\ &= E\left(\frac{\partial g_x}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_x}{\partial \nu}\right)^{\mathrm{T}} \end{aligned}$$

$$\begin{aligned} A_{yy} &\propto E\left(\frac{\partial g_y}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} \\ &\geq \left[E\left(\frac{\partial g_y}{\partial \phi}\right), E\left(\frac{\partial g_y}{\partial \nu}\right)\right] \left[\begin{array}{cc} 0 & 0 \\ 0 & W_2^{-1} \end{array}\right] \left[\begin{array}{c} E\left(\frac{\partial g_y}{\partial \phi}\right)^{\mathrm{T}} \\ E\left(\frac{\partial g_y}{\partial \nu}\right)^{\mathrm{T}} \end{array}\right] \\ &= E\left(\frac{\partial g_y}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_y}{\partial \nu}\right)^{\mathrm{T}} \end{aligned}$$

$$\begin{aligned} A_{xy} &\propto E\left(\frac{\partial g_x}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_y}{\partial \theta}\right)^{\mathrm{T}} \\ &\geq \left[E\left(\frac{\partial g_x}{\partial \phi}\right), E\left(\frac{\partial g_x}{\partial \nu}\right)\right] \left[\begin{array}{cc} 0 & 0 \\ 0 & W_2^{-1} \end{array}\right] \left[\begin{array}{c} E\left(\frac{\partial g_y}{\partial \phi}\right)^{\mathrm{T}} \\ E\left(\frac{\partial g_y}{\partial \nu}\right)^{\mathrm{T}} \end{array}\right] \end{aligned}$$

$$= E\left(\frac{\partial g_x}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_y}{\partial \nu}\right)^{\mathrm{T}}$$

$$\begin{aligned} A_{yx} &\propto E\left(\frac{\partial g_y}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_x}{\partial \theta}\right)^{\mathrm{T}} \\ &\geq \left[E\left(\frac{\partial g_y}{\partial \phi}\right), E\left(\frac{\partial g_y}{\partial \nu}\right)\right] \begin{bmatrix} 0 & 0 \\ 0 & W_2^{-1} \end{bmatrix} \begin{bmatrix} E\left(\frac{\partial g_x}{\partial \phi}\right)^{\mathrm{T}} \\ E\left(\frac{\partial g_x}{\partial \nu}\right)^{\mathrm{T}} \end{bmatrix} \\ &= E\left(\frac{\partial g_y}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_x}{\partial \nu}\right)^{\mathrm{T}} \end{aligned}$$

which establishes that $A - B$ is non negative definite, therefore $A - B$ is idempotent. Finally

$$tr(A - B) = tr(A) - tr(B) = p - (p - q) = q$$

which completes the proof. □

3.3 Conditional and Marginal Estimating Equations

Let $f(X, Y; \theta)$ be a joint probability distribution. This can be expressed in terms of the conditional distribution times the marginal distribution, i.e.

$$f(X, Y; \theta) = f(X; \theta) f(Y|X; \theta). \tag{3.5}$$

Creating the likelihood using the left hand side of Equation 3.5 would be the Fisher likelihood, while using the right hand side of Equation 3.5 results in the composite likelihood. Because of the equality shown in Equation 3.5 the Fisher likelihood and composite likelihood are equal for all θ. We can take a similar approach with the composite empirical likelihood, except we express the conditional and marginal components through the estimating equations. If we define the conditional moments correctly through the estimating equations, we can create valid confidence regions using the composite empirical likelihood.

Given any set of estimating equations, we can use the results from Theorem 2.2 or Theorem 3.1 (depending on the presence of nuisance parameters) to generate confidence regions. There are some

special cases where we can use the results from Corollary 3.1, eliminating the need to estimate the distribution of the test statistic. We explore one of these cases, and establish the necessary and sufficient conditions for the estimating equations to be uncorrelated.

Let $X \in \mathbb{R}^{d_x}$ and $Y \in \mathbb{R}^{d_y}$ be random variables from some joint distribution $F_{X,Y}$. For simplicity we will work with cases where $d_x = d_y = 1$. Assume that the conditional distribution $F_{Y|X}$ and the marginal distribution F_X exist. Define the marginal estimating equations as g_x and the conditional estimating equations as $g_{y|x}$, and we have that $E(g_x(X, \theta_0)) = 0$ and $E(g_{y|x}(Y|X, \theta_0)) = 0$.

To demonstrate we create the confidence region for the means of two random variables. We consider the case when the conditional mean can be expressed in terms of a linear function of the marginal random variable and a mean function m. The mean function will generally correspond to the inverse of a link function, with the exact form of the link function depending on the assumptions concerning the range of the parameters and random variables.

Define $m'(x)$ as the first derivative of the mean function. The marginal estimating equation is

$$g_x(x_i, \theta) = (x_i - m(\mu_x))\, m'(\mu_x) \tag{3.6}$$

and the estimating equation for the conditional mean is

$$g_{y|x}(y_i, x_i^*, \theta) = \begin{bmatrix} \left(y_i - m\left(\mu_y + \beta x_i^*\right)\right) m'\left(\mu_y + \beta x_i^*\right) \\ \left\{x_i^*\left(y_i - m\left(\mu_y + \beta x_i^*\right)\right)\right\} m'\left(\mu_y + \beta x_i^*\right) \end{bmatrix} \tag{3.7}$$

where

$$x_i^* = x_i - m(\hat{\mu})_x$$

and $\hat{\mu}_x$ is the estimator of μ_x from Equation 3.6, specifically the value of μ that solves

$$\sum_{i=1}^{n} u_i \left(x_i - m(\mu_x)\right) m'(\mu_x) = 0.$$

The simplest case is when the mean function is the identity, so $m(\theta) = \theta$. Then the marginal first moment of X is

$$E(X) = \mu_x$$

and the conditional first moment of $Y|X$ is

$$E(Y|X) = \mu_y + \beta(X - \mu_x).$$

In this context the nuisance parameter β can be expressed in terms of the marginal variances and correlation. The estimating equations $g_{y|x}$ results in the maximum composite empirical likelihood estimator of β being equal to the ordinary least squares fit, so

$$E\left(\hat{\beta}\right) = \beta = \rho\sigma_y/\sigma_x.$$

We only need one condition for the estimating equations to be uncorrelated as shown in the following Lemma.

Lemma 3.1 *Let g_x be as defined in Equation 3.6 and let $g_{y|x}$ be as defined in Equation 3.8. Furthermore let $m(\theta) = \theta$. Then a necessary and sufficient condition for $cov(g_x, g_{y|x}) = 0$ is*

$$E\left((X - \mu_x)^2(Y - \mu_y)\right) = \rho\sigma_x\sigma_y E\left((X - \mu_x)^3\right).$$

Proof. Without loss of generality let $\mu_x = 0$ and $\mu_y = 0$. By the strong law of large numbers

$$\hat{\mu}_x \longrightarrow \mu = 0 \quad (a.s.) \Rightarrow$$
$$E(X - \hat{\mu}_x) \longrightarrow E(X - \mu_x) = E(X).$$

First we will show that the correlation between g_x and the first estimating equation of $g_{y|x}$ is always 0 given the ordinary least squares assumption.

$$E\left((X)(Y - X\hat{\beta})\right) = E(XY) - \beta E(X^2)$$

$$= \rho\sigma_x\sigma_y - (\rho\sigma_y/\sigma_x)\sigma_x^2$$
$$= \rho\sigma_x\sigma_y - \rho\sigma_x\sigma_y$$
$$= 0.$$

Using the second estimating equation of $g_{y|x}$ we have

$$E\left((X)(X(Y - X\hat{\beta}))\right) = E(X^2Y) - \beta E(X^3)$$
$$= E(X^2Y) - \rho\sigma_x\sigma_y E(X^3)$$

which completes the proof. □

When X and Y are normally distributed the moment condition from Lemma 3.1 holds, but this condition may not be exclusive to the normal distribution. In general the condition is difficult to check, especially given that we are not specifying a data distribution. If the condition of Lemma 3.1 is not met the distribution of the test statistic can be estimated using the formulation shown in Section 3.2.

When the mean function is no longer the identity the marginal and conditional means are

$$E(X) = m\left(\mu_x\right),$$
$$E(Y|X) = m\left(\mu_y + \beta\left(X - m\left(\mu_x\right)\right)\right).$$

Although the marginal mean of X is $m(\mu_x)$ it is not necessarily the case that $E(Y) = m(\mu_y)$. We can create confidence regions for μ_x and μ_y, but we may not be able to examine the marginal mean of Y itself. In addition establishing that the estimating equations are uncorrelated is further complicated when m is not the identity function.

This construction is not limited to having the same mean function for both random variables. For example we could use m_x as the mean function of X and a mean function m_y for Y where $m_x \neq m_y$, in which case the estimating equations are

$$g_x(x_i, \theta) = \left(x_i - m_x(\mu_x)\right) m_x'(\mu_x)$$

and

$$g_{y|x}(y_i, x_i^*, \theta) = \begin{bmatrix} \left(y_i - m_y\left(\mu_y + \beta x_i^*\right)\right) m_y'\left(\mu_y + \beta x_i^*\right) \\ \left\{x_i^*\left(y_i - m_y\left(\mu_y + \beta x_i^*\right)\right)\right\} m_y'\left(\mu_y + \beta x_i^*\right) \end{bmatrix} \tag{3.8}$$

where

$$x_i^* = x_i - m_x(\hat{\mu}_x).$$

Another consideration is the use of the mean function for g_x. The addition of the mean function is to ensure that the predicted value for the mean parameter does not extend beyond a certain range. The linear component of the conditional mean $(\mu_y + \beta(x_i - \hat{\mu}_x))$ can be any real number, but the conditional mean of Y may not lie on $(-\infty, \infty)$. In contrast the estimator for the marginal mean of X will always lie between $\min(x_1, \ldots, x_n)$ and $\max(x_1, \ldots, x_n)$, and consequently will not give an implausible value for the marginal mean. To demonstrate this if we compute the MCELE $m(\hat{\mu}_1)$ using the the following estimating function

$$g_1 = (x_i - m(\mu_1))\, m'(\mu_1)$$

and then using the same data compute the MCELE $\hat{\mu}_2$ using

$$g_2 = x_i - \mu_2$$

we would have $m(\hat{\mu}_1) = \hat{\mu}_2$, making the use of the mean function unnecessary for the marginal component.

3.4 Conditional and Conditional Estimating Equations

Another approach when creating composite likelihoods is to condition each likelihood component using information from some (or all) of the remaining variables. This is the approach used by Besag (1975) where the likelihood of each variable is based on the conditional probability given the nearest points. This and similar composite likelihood methods usually involve a large number of conditional

likelihood components, but we will explore this approach for a bivariate region to establish the basic properties and methodology.

For ease of demonstration we will again work with inference of the bivariate mean. Instead of defining one of the means conditionally on the other random variable, both means are expressed conditionally. If we assume that the conditional means are linear in terms of the marginal variable and the mean function is the identity, we would have

$$E(X|Y=y) = \mu_x + \beta_x(y-\mu_y)$$
$$E(Y|X=x) = \mu_y + \beta_y(x-\mu_x).$$

The estimating equations are

$$g_{x|y}(x_i, y_i, \theta) = \begin{bmatrix} \left(x_i - \{\mu_x + \beta_x(y_i - \hat{\mu}_y)\}\right) \\ (y_i - \hat{\mu}_y)\left(x_i - \{\mu_x + \beta_x(y_i - \hat{\mu}_y)\}\right) \end{bmatrix} \tag{3.9}$$

and

$$g_{y|x}(y_i, x_i, \theta) = \begin{bmatrix} \left(y_i - \{\mu_y + \beta_y(x_i - \hat{\mu}_x)\}\right) \\ (x_i - \hat{\mu}_x)\left(y_i - \{\mu_y + \beta_y(x_i - \hat{\mu}_x)\}\right) \end{bmatrix} \tag{3.10}$$

where $\hat{\mu}_x$ and $\hat{\mu}_y$ are the estimators of μ_x and μ_y respectively.

Again the estimating equations are based on the ordinary least squares solution for simple linear regression, so we have

$$\beta_x = \rho\sigma_x/\sigma_y,$$
$$\beta_y = \rho\sigma_y/\sigma_x.$$

In order to ensure that $cov(g_{x|y}g_{y|x}) = 0$ we must have

$$E\left(\left(X - \{\mu_x + \beta_x(Y - \hat{\mu}_y)\}\right)\left(Y - \{\mu_y + \beta_y(X - \hat{\mu}_x)\}\right)\right) = 0$$
$$E\left(\left(X - \{\mu_x + \beta_x(Y - \hat{\mu}_y)\}\right)(X - \hat{\mu}_x)\left(Y - \{\mu_y + \beta_y(X - \hat{\mu}_x)\}\right)\right) = 0$$
$$E\left((Y - \hat{\mu}_y)\left(X - \{\mu_x + \beta_x(X - \hat{\mu}_y)\}\right)\left(Y - \{\mu_y + \beta_y(X - \hat{\mu}_x)\}\right)\right) = 0$$

$$E\left((Y-\hat{\mu}_y)\left(X-\{\mu_x+\beta_x(X-\hat{\mu}_y)\}\right)(X-\hat{\mu}_x)\left(Y-\{\mu_y+\beta_y(X-\hat{\mu}_x)\}\right)\right)=0.$$

If we assume $\mu_x = \mu_y = 0$ we have the four following conditions:

1. $\rho^3 = \rho$
2. $(1+\rho^2)E(X^2Y) = \rho(\sigma_y E(X^3)/\sigma_x + \sigma_x E(X^2Y)/\sigma_y)$
3. $(1+\rho^2)E(XY^2) = \rho(\sigma_x E(Y^3)/\sigma_y + \sigma_y E(XY^2)/\sigma_x)$
4. $(1+\rho^2)E(X^2Y^2) = \rho(\sigma_x E(XY^3)/\sigma_y + \sigma_y E(X^3Y)/\sigma_x)$

Condition 1 indicates that it is not possible for $cov(g_{x|y}, g_{y|x}) = 0$ unless ρ is 0, 1, or -1 when creating a composite empirical likelihood using Equations 3.9 and 3.10. Even if condition 1 is met, the remaining conditions are not simple properties to confirm even if the data distribution is known. Also we should note that if condition 1 holds either we would not have to make a confidence region (since variables are uncorrelated) or the confidence region would be a line. The addition of a mean function further complicates confirming these conditions.

Given the issues with checking $cov(g_{x|y}, g_{y|x}) = 0$ (and the implications given by condition 1), it is easier to utilize the result from Theorem 3.1. Even though this setup of the estimating equations requires estimation of several quantities in order to use the test statistic, these conditional forms are still very useful. We will further explore this use of conditional estimating equations for all the likelihood components in Chapter 4.

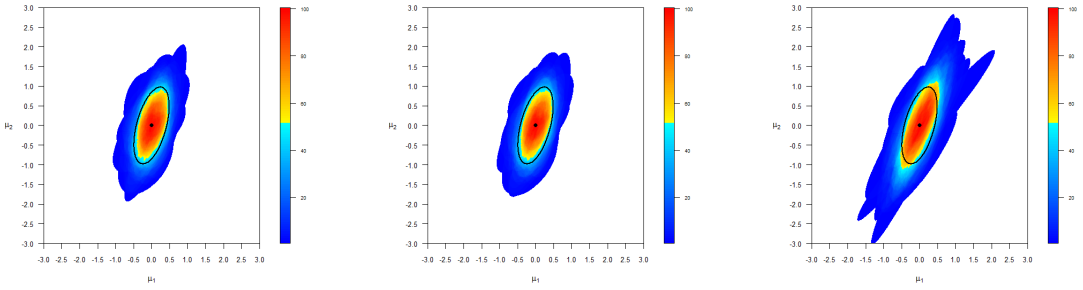
3.5 Numerical Studies

To examine the precision of the composite empirical likelihood confidence bound for both the conditional-marginal (L_{CM}) and conditional-conditional (L_{CC}) estimating equation setup, we examine the coverage of 100 confidence regions for randomly generated bivariate normal data with a sample size of 25. We use normal data since the necessary conditions are met so that the test statistic using the conditional-marginal form to have an asymptotic distribution of chi square with two degrees of freedom. Additionally the conditional mean of a normal random variable is $\mu_y + \beta(x - \mu_x)$, so $g_{y|x}$ is correctly describing the conditional mean. We compare the conditional-marginal and conditional-conditional methods to the empirical likelihood described by Owen (1990)

(L_E) using a chi square with two degrees of freedom to determine the boundary of the confidence region.

Theorem 3.1 and Corollary 3.1 indicate that the value of the nuisance parameter used in creation of a confidence region is the value of ν which maximizes the likelihood function at our null hypothesis ($\hat{\nu}_{CE}(\phi_0)$). This means that ν is determined by the value of the two means defining the border of the $(1-\alpha)\%$ confidence region. In the case of the conditional-conditional equation form this would require constant recalculation of the weighted chi square distribution in the optimization step, making the boundary region almost impossible to determine. Even in the conditional-marginal setup, numerically deriving $\hat{\nu}(\phi_0)$ adds additional computational strain, so we will use the MCELE $\hat{\nu}$ to calculate both the test statistic and the weighted chi square. Theorem 2.1 shows that both $\hat{\phi}$ and $\hat{\nu}$ are consistent estimators, therefore justifying our approach (Barndorff-Nielsen and Cox, 1994, page 91).

The data have a mean of $[0, 0]$. Three different sets of variances are used to confirm the coverage does not change based on choice of which variable uses the marginal estimating equations and which uses the conditional set. The marginal estimating equations are used with the random variable X and the conditional estimating equations with Y. The plots show the coverage percentage of the confidence regions.

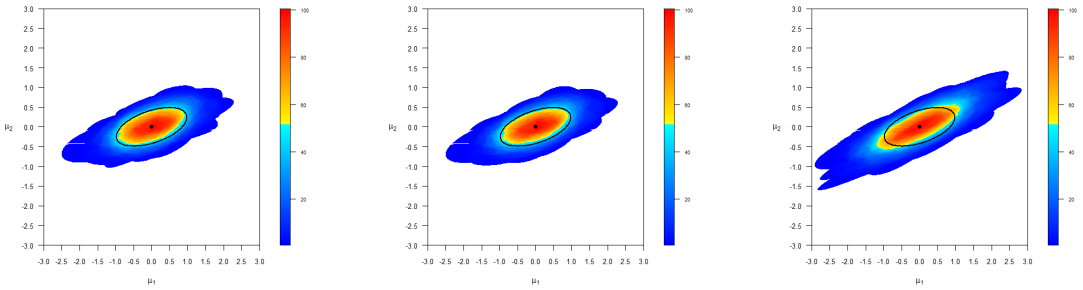


L_E
$[0, 0]$ captured 93%

L_{CM}
$[0, 0]$ captured 91%

L_{CC}
$[0, 0]$ captured 87%

Figure 3.2: Coverage of 100 confidence regions from bivariate normal with $\sigma_x = 1$, $\sigma_y = 4$, and $\rho = 0.50$. The center point is the true mean $[0, 0]$ and the ellipse is the true 95% confidence region of the mean.

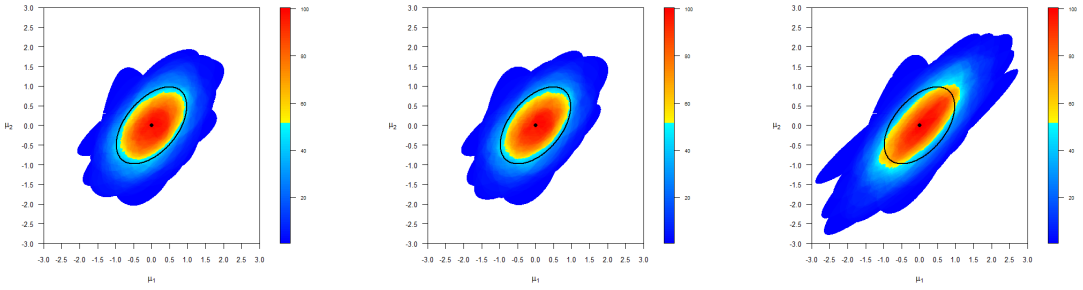


L_E
$[0, 0]$ captured 90%

L_{CM}
$[0, 0]$ captured 87%

L_{CC}
$[0, 0]$ captured 92%

Figure 3.3: Coverage of 100 confidence regions from bivariate normal with $\sigma_x = 4$, $\sigma_y = 1$, and $\rho = 0.50$. The center point is the true mean $[0, 0]$ and the ellipse is the true 95% confidence region of the mean.



L_E
$[0, 0]$ captured 94%

L_{CM}
$[0, 0]$ captured 91%

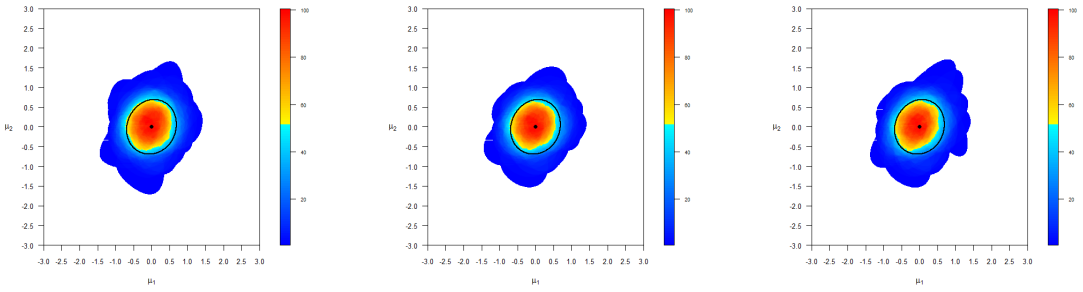
L_{CC}
$[0, 0]$ captured 96%

Figure 3.4: Coverage of 100 confidence regions from bivariate normal with $\sigma_x = 4$, $\sigma_y = 4$, and $\rho = 0.50$. The center point is the true mean $[0, 0]$ and the ellipse is the true 95% confidence region of the mean.

Figures 3.2, 3.3 and 3.4 confirm that the choice of which variable is assigned to the marginal and conditional equation does not affect coverage of region. L_{CC} does not accurately capture the coverage area, and appears to be over estimating the correlation given the elongation of the regions. Given that L_{CC} is not properly defining the relationship between the two variables (since it is describing a different data generating mechanism), the inaccuracy of the intervals is not surprising. Interestingly though the percentage of times the true mean is captured using L_{CC} is actually higher than the percentage captured by L_{CM} or L_E in two of the three simulations (see Figures 3.3 and 3.4), with L_{CM} having the lowest percentage capture of the mean overall.

Additionally we see that the regions created using L_{CM} and L_E show a high percentage of the simulations falling within the theoretical bound, and the number of times L_{CM} captures the true mean is 2% to 3% lower than L_E. It appears that the pattern of L_E over the 100 simulations is almost identical to L_{CM}, suggesting that the region generated for each simulation is almost identical.

The next simulations examine how different correlation values affect the coverage of the methods. Since the previous set of simulations indicates that variability is not a factor in assignment of estimating equation, we use the same marginal variances for both random variables.

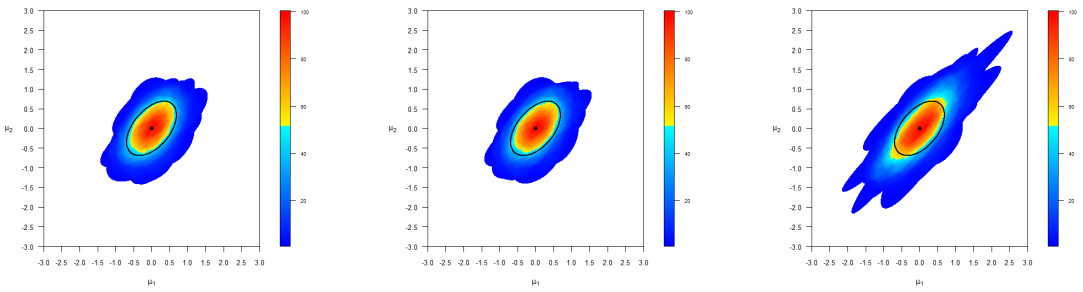


L_E
$[0, 0]$ captured 92%

L_{CM}
$[0, 0]$ captured 92%

L_{CC}
$[0, 0]$ captured 90%

Figure 3.5: Coverage of 100 confidence regions from bivariate normal with $\sigma_x = 2$, $\sigma_y = 2$, and $\rho = 0.10$. The center point is the true mean $[0, 0]$ and the ellipse is the true 95% confidence region of the mean.

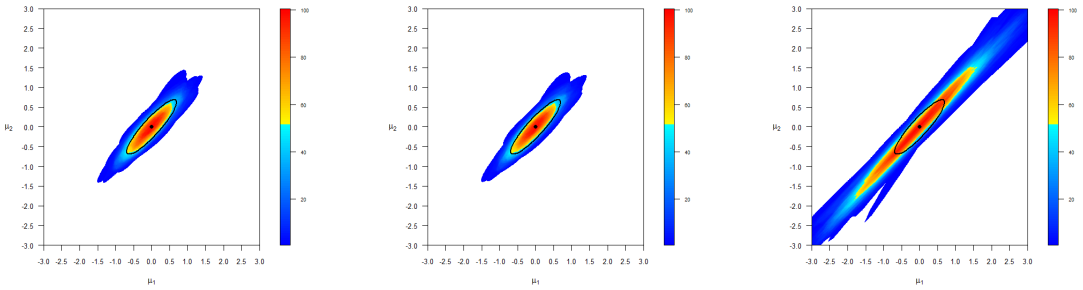


L_E
$[0, 0]$ captured 92%

L_{CM}
$[0, 0]$ captured 89%

L_{CC}
$[0, 0]$ captured 85%

Figure 3.6: Coverage of 100 confidence regions from bivariate normal with $\sigma_x = 2$, $\sigma_y = 2$, and $\rho = 0.50$. The center point is the true mean $[0, 0]$ and the ellipse is the true 95% confidence region of the mean.



L_E	L_{CM}	L_{CC}
$[0,0]$ captured 92%	$[0,0]$ captured 90%	$[0,0]$ captured 92%

Figure 3.7: Coverage of 100 confidence regions from bivariate normal with $\sigma_x = 2$, $\sigma_y = 2$, and $\rho = 0.90$. The center point is the true mean $[0,0]$ and the ellipse is the true 95% confidence region of the mean.

Again we see that the coverage percentage of L_{CM} is almost indistinguishable from L_E. An interesting characteristic we see from Figure 3.5 is that when the correlation is small L_{CC} shows very similar coverage to what we see with the other two methods. However Figures 3.6 and 3.7 show as the correlation increases the region of L_{CC} begins to elongate, suggesting that this method is over estimating the correlation and variability. A possible explanation for this behavior follows from the derivation of the information when working with conditional-conditional likelihoods. Lindsay (1988) shows that if both variables are marginally normal than the conditional-conditional form is only fully informative when the correlation between the two variables is 0. This result may give an explanation as to the behavior we are seeing of the conditional-conditional composite empirical likelihood.

The coverage of the true mean is the same for all three simulations using L_E (92%) while L_{CM} is 2% to 3% lower than L_E. L_{CC} shows an unusual pattern with the capture percentage of $[0,0]$ decreasing when the correlation is at 0.5 but then increases again when the correlation is 0.9.

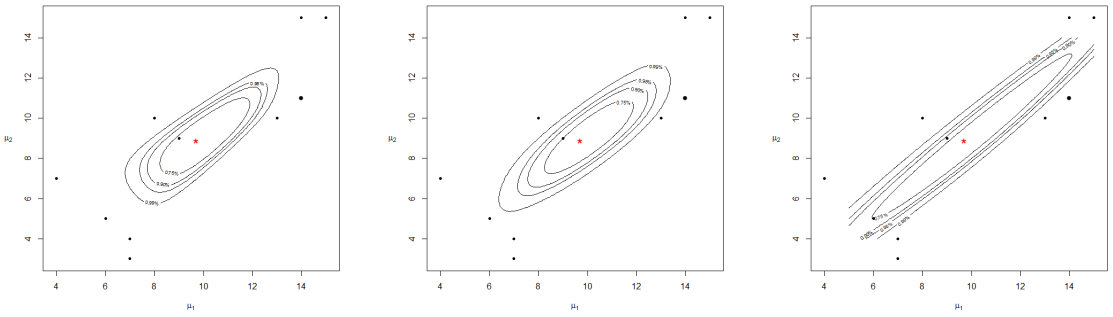
3.6 Discussion

The conditional-marginal equation setup appears to be a better method compared to the conditional-conditional equation setup when creating confidence regions for the mean parameters if we consider the overall pattern of the regions, but if we look solely at the coverage of $[0,0]$ the conditional-marginal equation setup in most of the cases we examine has a slightly lower capture

percentage. The reduction may be a result of the necessity of estimating β for the conditional-marginal forms, whereas the empirical likelihood does not require estimation of any additional information. The simulations show that given correct moment assumptions the conditional-marginal empirical likelihood results in almost identical regions as we would see with empirical likelihood. The conditional-conditional approach does not accurately capture the region, but this can be attributed to the fact that the specification of both moments does not correctly describe the data generating mechanism.

We now examine L_{CM} and L_{CC} with the plumage characteristics data. Since we cannot confirm the moment conditions from Lemma 3.1 we will use the result from Theorem 3.1 for the test statistic of L_{CM}.

Figure 3.8 shows that L_{CM} results in a similar region to what we observed in Figure 3.1 using empirical likelihood. The conditional-marginal form covers a larger area than L_E, and has a more uniform shape. We also see from Figure 3.8 that L_{CC} appears to overestimate the correlation in the data, as seen by the stretching of the interval. This is not unexpected given the simulation results observed; the sample correlation of this data is 0.8250.



L_E L_{CM} L_{CC}

Figure 3.8: Comparison of empirical and composite empirical likelihood methods for behavioral and plumage characteristics of hybrid ducks data. The asterisk (in red) denotes the sample mean of the two variables, and the larger sized point at $14, 11$ denotes that there were two observations with those values.

Despite the apparent equivalence between empirical likelihood and conditional-marginal composite empirical likelihood, we do have the additional issue of proper specification of the moments between the two random variables. In many cases working with a linear relationship is sound, since the measure of correlation describes linear dependency. There are many situations where the

dependency is not linear, which would put our method at a disadvantage since either the linear assumption would be wrong or the relationship between variables would have to be known in order to properly define the estimating equations. Despite this, we see our approach as being advantageous when working with high dimensional inference and large sample sizes. In these cases having to make additional assumptions may outweigh the computational issues that would arise if working with empirical likelihood.

Although the conditional-conditional estimating equation setup does not result in valid coverage the use of conditional equations is more common when working with higher dimensional problems, such as spatial fields. The composite likelihoods created using a high dimensional conditional setup have been studied and result in consistent estimators (Besag, 1974, 1975).

The examples explored in this chapter are not designed to replace empirical likelihood as a method to create confidence regions, but rather to demonstrate methodology and proof of concept. We establish that if the moment conditions are properly specified then a conditional-marginal equation approach results in almost identical intervals seen with empirical likelihood. This follows from the theoretical property that a joint distribution is equivalent to the product of the marginal and conditional distribution.

Chapter 4

Composite Empirical Likelihood

4.1 Introduction

Advances in data collection, management and storage have increased available sample sizes by multiple orders of magnitude. This increase of high throughput data has opened up many statistical modeling questions due to the availability of data, but at the same time have dramatically increased the computational load for standard methodologies. In particular methods that require numerical optimization in order to find solutions are greatly affected when sample sizes are large.

In order to fully utilize the composite empirical likelihood method, we need to extend the theoretical results beyond two likelihood components. Using multiple components to define the composite empirical likelihood can dramatically reduce the sample size required for each empirical likelihood component, thus reducing the impact large samples and high dimensional data have on numerical optimization. Furthermore, this high dimensional composite empirical likelihood is applicable to problems relying on complex conditional moments, such as those seen in spatial statistics.

This chapter will expand on the theorems and corollaries shown in Chapters 2 and 3 by expressing the results in terms of a general number of likelihood components.

4.2 Composite Empirical Likelihood

Let $\mathbf{Z} \in \mathbb{R}^d$ be from some distribution f_0. Define $Z_{.j}$ as a subset of $\mathbf{Z}$ for $j = 1, \ldots, J$, and we will assume that $Z_{.j} \sim f_j$. Define g_j for $j = 1, \ldots, J$ as the estimating equations for each subset and define the parameters as θ_j having dimension $p_j \times 1$. The dimension of the parameter θ is $p \leq \sum_{j=1}^{J} p_j$, and finally assume $r_j \geq p_j$ for all j, where r_j is dimension of g_j. For each subset j

we have the component empirical likelihood

$$L_{CE}^{(j)}(\theta) = \left(\sup_{p_{\cdot j}} \prod_{i=1}^{n_j} p_{i,j} \left| \sum_{i=1}^{n_j} p_{i,j} g_j(z_{i,j}, \theta) = 0,\ p_{i,j} \geq 0,\ \sum_{i=1}^{n_j} p_{i,j} = 1 \right. \right),$$

and denote the negative log of the empirical likelihood for each subset as $\ell_{CE}^{(j)}(\theta)$. The composite empirical likelihood function is

$$\begin{aligned} L_{CE}(\theta) &= \prod_{j=1}^{J} L_E^{(j)}(\theta) \\ &= \prod_{j=1}^{J} \left(\sup_{p_{\cdot j}} \prod_{i=1}^{n_j} p_{i,j} \left| \sum_{i=1}^{n_j} p_{i,j} g_j(z_{i,j}, \theta) = 0,\ p_{i,j} \geq 0,\ \sum_{i=1}^{n_j} p_{i,j} = 1 \right. \right), \end{aligned}$$

and the negative log of the general composite empirical likelihood is

$$\begin{aligned} \ell_{CE}(\theta) &= \sum_{j=1}^{J} -\log(L_E^{(j)}(\theta)) \\ &= \sum_{j=1}^{J} \ell_E^{(j)}(\theta). \end{aligned} \tag{4.1}$$

As seen in Chapter 2 the maximum composite empirical likelihood estimator is

$$\hat{\theta}_{CE} = \arg\max_{\theta} L_{CE}(\theta).$$

The following assumptions, lemmas, theorems and corollaries are extensions of the results from Chapter 2 to apply for a general J case of composite empirical likelihood.

Assumption 3 Let θ_0 be the true value of θ. Then for all j

(a) $E\{g_j(Z_j, \theta_0) g_j^{\mathrm{T}}(Z_j, \theta_0)\}$ is positive definite.

(b) $\partial g_j(Z_j, \theta)/\partial\theta$ is continuous in a neighborhood of the true value θ_0.

(c) $\|\partial g_j(Z_j, \theta)/\partial\theta\|$ and $\|g_j(Z_j, \theta)\|^3$ are both bounded by some integrable function in the same neighborhood of θ_0.

(d) The rank of $E\{\partial g_j(Z_j, \theta)/\partial\theta\}$ is p_j.

(e) The second derivative $\partial^2 g_j(Z_j, \theta)/\partial\theta\partial\theta^{\mathrm{T}}$ is continuous in θ in a neighborhood of the true value

θ_0.

(f) $\|\partial^2 g_j(Z_j,\theta)/\partial\theta\partial\theta^{\mathrm{T}}\|$ can be bounded by some integrable function in the neighborhood of θ_0.

We assume Assumption 3 holds throughout. Additionally assume the sample sizes n_j for all j are increasing at approximately the same rate.

Lemma 4.1 *$\ell_{CE}(\theta)$ attains its minimum value at some point $\hat{\theta}_{CE}$ in the interior of the ball $\|\theta-\theta_0\| \le n^{-1/3}$ with probability 1 as $n \longrightarrow \infty$.*

Furthermore $\hat{\theta}_{CE}$ and $\hat{t}_j = t_j(\hat{\theta}_{CE})$ for all j satisfy

$$Q_{1j}(\hat{\theta}_{CE}, \hat{t}_j) = 0$$

where

$$Q_{1j}(\theta, t_j) = \frac{1}{n_j}\sum_{i=1}^{n_j}\{1 + t_j^{\mathrm{T}} g_j(z_{i,j},\theta)\}^{-1} g_j(z_{i,j},\theta)$$

and

$$Q_2(\hat{\theta}_{CE}, \hat{t}_1, \ldots, \hat{t}_J) = 0$$

where

$$Q_2(\theta, t_1, \ldots, t_J) = \sum_{j=1}^{J}\left\{\frac{1}{n_j}\sum_{i=1}^{n_j}\{1 + t_j^{\mathrm{T}} g_j(z_{i,j},\theta)\}^{-1}\left(\frac{\partial g_j(z_{i,j},\theta)}{\partial\theta}\right)^{\mathrm{T}} t_j\right\}.$$

The proof follows directly from Qin and Lawless (1994, Lemma 1) and Lemma 2.2.

Theorem 4.1 *We have*

$$\sqrt{n}(\hat{\theta}_{CE} - \theta_0) \longrightarrow N(0, W^{-1}(W+V)W^{-1})$$

where

$$W = \sum_{j=1}^{J} E\left(\frac{\partial g_j}{\partial\theta}\right)^{\mathrm{T}}\left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial\theta}\right)$$

$$V = \sum_{j \neq k}^{J} E\left(\frac{\partial g_j}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E(g_j g_k^{\mathrm{T}}) \left\{E(g_k g_k^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_k}{\partial \theta}\right).$$

Proof. First by Taylor expansion of $Q_{1j}(\hat{\theta}_{CE}, \hat{t}_j)$ around θ_0 and 0

$$\begin{aligned} 0 &= Q_{1j}(\hat{\theta}_{CE}, \hat{t}_j) \\ &= Q_{1j}(\theta_0, 0) + \frac{\partial}{\partial \theta} Q_{1j}(\theta_0, 0)(\hat{\theta}_{CE} - \theta_0) + \frac{\partial}{\partial t_j^{\mathrm{T}}} Q_{1j}(\theta_0, 0)(\hat{t}_j - 0) + o_p(\delta_j) \end{aligned}$$

where $\delta_j = \|\hat{\theta}_{CE} - \theta_0\| + \|\hat{t}_j\|$. Solving for $\hat{t}_j$ yields

$$(\hat{t}_j - 0) = \left\{-\frac{\partial Q_{1j}(\theta_0, 0)}{\partial t_j^{\mathrm{T}}}\right\}^{-1} \left\{Q_{1j}(\theta_0, 0) + \frac{\partial}{\partial \theta} Q_{1j}(\theta_0, 0)(\hat{\theta}_{CE} - \theta_0) + o_p(\delta_j)\right\} \tag{4.2}$$

for all j. Now by Taylor expansion of $Q_2(\hat{\theta}_{CE}, \hat{t}_1, \ldots, \hat{t}_J)$ around $\theta_0, 0, \ldots, 0$

$$\begin{aligned} 0 &= Q_2(\hat{\theta}_{CE}, \hat{t}_1, \ldots, \hat{t}_J) \\ &= Q_2(\theta_0, 0, \ldots, 0) + \frac{\partial}{\partial \theta} Q_2(\theta_0, 0, \ldots, 0)(\hat{\theta}_{CE} - \theta_0) \\ &\quad + \sum_{j=1}^{J} \frac{\partial}{\partial t_j^{\mathrm{T}}} Q_2(\theta_0, 0, \ldots, 0)(\hat{t}_j - 0) + o_p(\delta) \end{aligned}$$

where $\delta = \|\hat{\theta}_{CE} - \theta_0\| + \sum \|\hat{t}_j\|$. $Q_2(\theta_0, 0, \ldots, 0) = 0$ and $\partial Q_2(\theta_0, 0, \ldots, 0)/\partial \theta = 0$ so

$$0 = \sum_{j=1}^{J} \frac{\partial}{\partial t_j^{\mathrm{T}}} Q_2(\theta_0, 0, \ldots, 0)(\hat{t}_j - 0) + o_p(\delta). \tag{4.3}$$

Substituting Equation 4.2 into Equation 4.3 gives

$$\begin{aligned} 0 = \sum_{j=1}^{J} &\left(\frac{\partial}{\partial t_j^{\mathrm{T}}} Q_2(\theta_0, 0, \ldots, 0) \left\{-\frac{\partial Q_{1j}(\theta_0, 0)}{\partial t_j^{\mathrm{T}}}\right\}^{-1} \left\{Q_{1j}(\theta_0, 0) + \frac{\partial}{\partial \theta} Q_{1j}(\theta_0, 0)(\hat{\theta}_{CE} - \theta_0)\right\} \right) \\ &+ o_p(\max(\delta_j, \delta)). \end{aligned}$$

The derivatives of Q_{1j} and Q_2 with respect to θ and t_j are

$$\begin{aligned}
\frac{\partial Q_{1j}(\theta_0,0)}{\partial \theta} &= \frac{1}{n_j}\sum_{i=1}^{n_j}\frac{\partial g_j(z_{i,j},\theta_0)}{\partial \theta} \longrightarrow E\left(\frac{\partial g_j}{\partial \theta}\right) \\
\frac{\partial Q_{1j}(\theta_0,0)}{\partial t_j^{\mathrm{T}}} &= -\frac{1}{n_j}\sum_{i=1}^{n_j} g_j(z_{i,j},\theta_0)g_j^{\mathrm{T}}(z_{i,j},\theta_0) \longrightarrow -E(g_j g_j^{\mathrm{T}}) \\
\frac{\partial Q_2(\theta_0,0,\ldots,0)}{\partial t_j^{\mathrm{T}}} &= \frac{1}{n_j}\sum_{i=1}^{n_j}\left(\frac{\partial g_x(x_i,\theta_0)}{\partial \theta}\right)^{\mathrm{T}} \longrightarrow E\left(\frac{\partial g_j}{\partial \theta}\right)^{\mathrm{T}}
\end{aligned}$$

so

$$\begin{aligned}
0 = &\sum_{j=1}^{J}\left(E\left(\frac{\partial g_j}{\partial \theta}\right)^{\mathrm{T}}\left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1}\left\{-Q_{1j}(\theta_0,0)\right\}\right) \\
&-\sum_{j=1}^{J}\left(E\left(\frac{\partial g_j}{\partial \theta}\right)^{\mathrm{T}}\left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1}E\left(\frac{\partial g_j}{\partial \theta}\right)(\hat{\theta}_{CE}-\theta_0)\right) \\
&+o_p(1) \qquad \Rightarrow
\end{aligned}$$

$$(\hat{\theta}_{CE}-\theta_0) = W^{-1}\sum_{j=1}^{J}E\left(\frac{\partial g_j}{\partial \theta}\right)^{\mathrm{T}}\left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1}\left\{-Q_{1j}(\theta_0,0)\right\}+o_p(1)$$

where

$$W = \sum_{j=1}^{J}\left\{E\left(\frac{\partial g_j}{\partial \theta}\right)^{\mathrm{T}}\left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1}E\left(\frac{\partial g_j}{\partial \theta}\right)\right\}.$$

For all j

$$-\sqrt{n}Q_{1j}(\theta_0,0) \longrightarrow N(0,E(g_j g_j^{\mathrm{T}})),$$

so the variance of $\sqrt{n}(\hat{\theta}_{CE}-\theta_0)$ is

$$\left[\begin{array}{cccc} W^{-1}E\left(\frac{\partial g_1}{\partial \theta}\right)^{\mathrm{T}}\left\{E(g_1 g_1^{\mathrm{T}})\right\}^{-1} & \cdots & W^{-1}E\left(\frac{\partial g_J}{\partial \theta}\right)^{\mathrm{T}}\left\{E(g_J g_J^{\mathrm{T}})\right\}^{-1} \end{array}\right]\times$$

$$\begin{bmatrix} E(g_1g_1^{\mathrm{T}}) & E(g_1g_2^{\mathrm{T}}) & \cdots & E(g_1g_J^{\mathrm{T}}) \\ E(g_2g_1^{\mathrm{T}}) & E(g_2g_2^{\mathrm{T}}) & \cdots & E(g_2g_J^{\mathrm{T}}) \\ \vdots & \vdots & \ddots & \vdots \\ E(g_Jg_1^{\mathrm{T}}) & E(g_Jg_2^{\mathrm{T}}) & \cdots & E(g_Jg_J^{\mathrm{T}}) \end{bmatrix} \times \begin{bmatrix} \left\{E(g_1g_1^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_1}{\partial\theta}\right) W^{-1} \\ \vdots \\ \left\{E(g_Jg_J^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_J}{\partial\theta}\right) W^{-1} \end{bmatrix}$$

$$= \sum_{k=1}^{J}\sum_{j=1}^{J} W^{-1} E\left(\frac{\partial g_j}{\partial\theta}\right)^{\mathrm{T}} \left\{E(g_jg_j^{\mathrm{T}})\right\}^{-1} E(g_jg_k^{\mathrm{T}}) \left\{E(g_kg_k^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_k}{\partial\theta}\right) W^{-1}$$

$$= W^{-1}\left\{\sum_{j=1}^{J} E\left(\frac{\partial g_j}{\partial\theta}\right)^{\mathrm{T}} \left\{E(g_jg_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial\theta}\right)\right\} W^{-1}$$
$$+ W^{-1}\left\{\sum_{j\neq k}^{J} E\left(\frac{\partial g_j}{\partial\theta}\right)^{\mathrm{T}} \left\{E(g_jg_j^{\mathrm{T}})\right\}^{-1} E(g_jg_k^{\mathrm{T}}) \left\{E(g_kg_k^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_k}{\partial\theta}\right)\right\} W^{-1}$$

$$= W^{-1}(W+V)W^{-1}$$

which completes the proof. □

Theorem 4.2 *The composite empirical likelihood ratio statistic for testing $H_0 : \theta = \theta_0$ is*

$$T = 2\ell_{CE}(\theta_0) - 2\ell_{CE}(\hat{\theta}_{CE})$$

where ℓ_{CE} is given by Equation 4.1. We have

$$T \longrightarrow Q(\lambda)$$

as $n \longrightarrow \infty$ when H_0 is true. $Q(\lambda) = \sum_{i=1}^{l} \lambda_i \chi^2(1)$ is a weighed chi square random variable where λ_i for $i = 1, \ldots, l$ is the set of all non zero eigenvalues of

$$A\Gamma = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1J} \\ A_{21} & A_{22} & \cdots & A_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ A_{J1} & A_{J2} & \cdots & A_{JJ} \end{bmatrix} \begin{bmatrix} E(g_1g_1^{\mathrm{T}}) & E(g_1g_2^{\mathrm{T}}) & \cdots & E(g_1g_J^{\mathrm{T}}) \\ E(g_2g_1^{\mathrm{T}}) & E(g_2g_2^{\mathrm{T}}) & \cdots & E(g_2g_J^{\mathrm{T}}) \\ \vdots & \vdots & \ddots & \vdots \\ E(g_Jg_1^{\mathrm{T}}) & E(g_Jg_2^{\mathrm{T}}) & \cdots & E(g_Jg_J^{\mathrm{T}}) \end{bmatrix}$$

where

$$A_{jk} = \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_k}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_k g_k^{\mathrm{T}})\right\}^{-1}$$

and W is as defined in Theorem 4.1.

Proof. The details follow directly from Theorem 2.2. First for a given value of θ

$$\ell_{CE}(\theta) = \sum_{j=1}^{J} \left(\frac{n}{2} \left\{ \frac{1}{n} \sum_{i=1}^{n} g_j^{\mathrm{T}}(z_{i,j}, \theta) \right\} \left\{ \frac{1}{n} \sum_{i=1}^{n} g_j(z_{i,j}, \theta) g_j^{\mathrm{T}}(z_{i,j}, \theta) \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} g_j(z_{i,j}, \theta) \right\} \right) + o_p(1)$$

so

$$\ell_{CE}(\theta_0) = \sum_{j=1}^{J} \left(\frac{n}{2} Q_{1j}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} Q_{1j}(\theta_0, 0) \right) + o_p(1)$$

and by Taylor expansion

$$\begin{aligned}
\ell_{CE}(\hat{\theta}_{CE}) = \ell_{CE}(\theta_0) & \\
&+ \sum_{j=1}^{J} \left(-\frac{n}{2} Q_{1j}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial \theta}\right) W E\left(\frac{\partial g_j}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} Q_{1j}(\theta_0, 0) \right) \\
&+ \sum_{j \neq k}^{J} \left(-\frac{n}{2} Q_{1j}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial \theta}\right) W E\left(\frac{\partial g_k}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_k g_k^{\mathrm{T}})\right\}^{-1} Q_{1k}(\theta_0, 0) \right) \\
&+ o_p(1).
\end{aligned}$$

The test statistic is

$$\begin{aligned}
T &= 2\ell_{CE}(\theta_0) - 2\ell_{CE}(\hat{\theta}_{CE}) \\
&= \sum_{j=1}^{J} \left(\sqrt{n} Q_{1j}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial \theta}\right) W E\left(\frac{\partial g_j}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} \sqrt{n} Q_{1j}(\theta_0, 0) \right) \\
&\quad + \sum_{j \neq k}^{J} \left(\sqrt{n} Q_{1j}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial \theta}\right) W E\left(\frac{\partial g_k}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_k g_k^{\mathrm{T}})\right\}^{-1} \sqrt{n} Q_{1k}(\theta_0, 0) \right)
\end{aligned}$$

$$+ o_p(1).$$

$\sqrt{n}Q_{1j}(\theta_0, 0)$ for $j = 1, \ldots, J$ converge to multivariate normal with a mean of 0 and covariance Γ. The matrix A is positive definite and we can write T as

$$\left[\begin{array}{ccc} \sqrt{n}Q_{11}^{\mathrm{T}}(\theta_0, 0) & \cdots & \sqrt{n}Q_{1J}^{\mathrm{T}}(\theta_0, 0) \end{array} \right] \times A \times \left[\begin{array}{c} \sqrt{n}Q_{11}(\theta_0, 0) \\ \vdots \\ \sqrt{n}Q_{1J}(\theta_0, 0) \end{array} \right]$$

which completes the proof. □

Corollary 4.1 *Assume that $cov(g_j, g_k) = 0$ for all $j \neq k$. Then*

$$T \longrightarrow \chi^2(p)$$

as $n \longrightarrow \infty$ when H_0 is true. T is defined in Theorem 4.2.

Proof. Following the steps from Corollary 2.1 we can rewrite the test statistic so that the j, k-th element of A is

$$A_{jk} = \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g_j}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_k}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_k g_k^{\mathrm{T}})\right\}^{-1/2}.$$

Note

$$-\sqrt{n}Q_{1j}(\theta_0, 0) \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1/2} \longrightarrow N(0, I)$$

with

$$E(g_j g_k^{\mathrm{T}}) = cov(g_j, g_k) = 0$$

for all $j \neq k$ so

$$\Gamma = I.$$

We need only show that A is idempotent for T to be asymptotically chi square. Define $(A^2)_{j,k}$ as the j,k-th element of the square of the matrix A, then we have

$$\begin{aligned}(A^2)_{j,k} &= \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g_j}{\partial \theta}\right) \\ &\quad \times W^{-1}\left\{\sum_{j=1}^{J} E\left(\frac{\partial g_j}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial \theta}\right)\right\} W^{-1} \\ &\quad \times E\left(\frac{\partial g_k}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_k g_k^{\mathrm{T}})\right\}^{-1/2} \\ &= \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1/2} E\left(\frac{\partial g_j}{\partial \theta}\right) W^{-1} W W^{-1} E\left(\frac{\partial g_k}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_k g_k^{\mathrm{T}})\right\}^{-1/2} \\ &= A_{j,k}.\end{aligned}$$

Therefore A is idempotent which completes the proof. □

Theorem 4.3 *Let $\theta = [\phi^{\mathrm{T}}, \nu^{\mathrm{T}}]^{\mathrm{T}}$ be a p dimensional vector where ϕ is a $q \times 1$ vector and ν is a $(p-q) \times 1$ vector. The profile composite empirical likelihood ratio test statistic to test $H_0 : \phi = \phi_0$ is*

$$T_2 = 2\ell_{CE}(\phi_0, \hat{\nu}_{CE}(\phi_0)) - 2\ell_{CE}(\hat{\phi}_{CE}, \hat{\nu}_{CE})$$

where $\hat{\nu}_{CE}(\phi_0)$ minimizes $\ell_{CE}(\phi, \nu)$ with respect to ϕ_0. Under H_0

$$T_2 \longrightarrow Q(\lambda)$$

as $n \longrightarrow \infty$. $Q(\lambda) = \sum_{i=1}^{l} \lambda_i {\chi^2}(1)$ is a weighted chi square random variable where λ_i for $i = 1, \ldots, l$ is the set of all non zero eigenvalues of

$$A\Gamma = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1J} \\ A_{21} & A_{22} & \cdots & A_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ A_{J1} & A_{J2} & \cdots & A_{JJ} \end{bmatrix} \begin{bmatrix} E(g_1 g_1^{\mathrm{T}}) & E(g_1 g_2^{\mathrm{T}}) & \cdots & E(g_1 g_J^{\mathrm{T}}) \\ E(g_2 g_1^{\mathrm{T}}) & E(g_2 g_2^{\mathrm{T}}) & \cdots & E(g_2 g_J^{\mathrm{T}}) \\ \vdots & \vdots & \ddots & \vdots \\ E(g_J g_1^{\mathrm{T}}) & E(g_J g_2^{\mathrm{T}}) & \cdots & E(g_J g_J^{\mathrm{T}}) \end{bmatrix}$$

where

$$A_{jk} = \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} \left\{ E\left(\frac{\partial g_j}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_k}{\partial \theta}\right)^{\mathrm{T}} - E\left(\frac{\partial g_j}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_k}{\partial \nu}\right)^{\mathrm{T}} \right\} \left\{E(g_k g_k^{\mathrm{T}})\right\}^{-1},$$

$$W_2 = \sum_{j=1}^{J} E\left(\frac{\partial g_j}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial \nu}\right)$$

and W is as defined in Theorem 4.1

Proof. The proof follows from Theorems 3.1 and 4.2. In Theorem 4.2 we establish

$$\begin{aligned}
\ell_{CE}(\hat{\theta}_{CE}) = \ell_{CE}(\theta_0) & \\
&+ \sum_{j=1}^{J} \left(-\frac{n}{2} Q_{1j}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_j}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} Q_{1j}(\theta_0, 0) \right) \\
&+ \sum_{j \neq k}^{J} \left(-\frac{n}{2} Q_{1j}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial \theta}\right) W^{-1} E\left(\frac{\partial g_k}{\partial \theta}\right)^{\mathrm{T}} \left\{E(g_k g_k^{\mathrm{T}})\right\}^{-1} Q_{1k}(\theta_0, 0) \right) \\
&+ o_p(1).
\end{aligned}$$

where $\hat{\theta}_{CE} = [\hat{\phi}_{CE}^{\mathrm{T}}, \hat{\nu}_{CE}^{\mathrm{T}}]^{\mathrm{T}}$ and $\theta = [\phi^{\mathrm{T}}, \nu^{\mathrm{T}}]^{\mathrm{T}}$. Following Theorem 3.1 we can show

$$\begin{aligned}
\ell_{CE}(\phi_0, \hat{\nu}_{CE}(\phi_0)) = \ell_{CE}(\theta_0) & \\
&+ \sum_{j=1}^{J} \left(-\frac{n}{2} Q_{1j}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_j}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} Q_{1j}(\theta_0, 0) \right) \\
&+ \sum_{j \neq k}^{J} \left(-\frac{n}{2} Q_{1j}^{\mathrm{T}}(\theta_0, 0) \left\{E(g_j g_j^{\mathrm{T}})\right\}^{-1} E\left(\frac{\partial g_j}{\partial \nu}\right) W_2^{-1} E\left(\frac{\partial g_k}{\partial \nu}\right)^{\mathrm{T}} \left\{E(g_k g_k^{\mathrm{T}})\right\}^{-1} Q_{1k}(\theta_0, 0) \right) \\
&+ o_p(1),
\end{aligned}$$

where W_2 is as defined previously. The remainder of the proof follows from the steps shown in Theorem 2.2. □

Corollary 4.2 *Assume that $cov(g_j, g_k) = 0$ for all $j \neq k$. Then*

$$T_2 \longrightarrow \chi^2(q)$$

as $n \longrightarrow \infty$ when H_0 is true. T_2 is defined in Theorem 4.3.

We omit the details, but the proof follows directly from Corollary 3.1 by substituting T_2 from Corollary 4.1.

Estimation of the quadratic matrix and covariance matrix necessary to determine the weighted chi square is accomplished using the methods from Chapter 2. The elements of the covariance matrix are estimated using the consistent estimator

$$\widehat{\Gamma}_{jk}(\theta) = \frac{1}{n_{jk}} \sum g_j(z_{ij}, \theta) g_k^{\mathrm{T}}(z_{ik}, \theta)$$

where n_{jk} are the number of observed $z_{i,j}, z_{ik}$ pairs, and the elements for the quadratic matrix are estimated by

$$\widehat{W}(\theta) = \sum_{j=1}^{J} \left\{ \left(\frac{1}{n_j} \sum_{i=1}^{n_j} \frac{\partial g_j(z_{ij}, \theta)}{\partial \theta} \right)^{\mathrm{T}} \left(\frac{1}{n_j} \sum_{i=1}^{n_j} g_j^{\mathrm{T}}(z_{ij}, \theta) g_j(z_{ij}, \theta) \right)^{-1} \left(\frac{1}{n_j} \sum_{i=1}^{n_j} \frac{\partial g_j(z_{ij}, \theta)}{\partial \theta} \right) \right\}$$

$$\widehat{W}_2(\theta) = \sum_{j=1}^{J} \left\{ \left(\frac{1}{n_j} \sum_{i=1}^{n_j} \frac{\partial g_j(z_{ij}, \theta)}{\partial \nu} \right)^{\mathrm{T}} \left(\frac{1}{n_j} \sum_{i=1}^{n_j} g_j^{\mathrm{T}}(z_{ij}, \theta) g_j(z_{ij}, \theta) \right)^{-1} \left(\frac{1}{n_j} \sum_{i=1}^{n_j} \frac{\partial g_j(z_{ij}, \theta)}{\partial \nu} \right) \right\}$$

$$\begin{aligned}
\widehat{A}_{jk}(\theta) = & \left(\frac{1}{n_j} \sum_{i=1}^{n_j} g_j(z_{ij}, \theta) g_j^{\mathrm{T}}(z_{ij}, \theta) \right)^{-1} \\
& \times \left\{ \left(\frac{1}{n_j} \sum_{i=1}^{n_j} \frac{\partial g_j(z_{ij}, \theta)}{\partial \theta} \right) \widehat{W}(\theta)^{-1} \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \frac{\partial g_k(z_{ik}, \theta)}{\partial \theta} \right)^{\mathrm{T}} - \right. \\
& \qquad \left. \left(\frac{1}{n_j} \sum_{i=1}^{n_j} \frac{\partial g_j(z_{ij}, \theta)}{\partial \nu} \right) \widehat{W}_2(\theta)^{-1} \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \frac{\partial g_k(z_{ik}, \theta)}{\partial \nu} \right)^{\mathrm{T}} \right\} \\
& \times \left(\frac{1}{n_k} \sum_{i=1}^{n_k} g_k(z_{ik}, \theta) g_k^{\mathrm{T}}(z_{ik}, \theta) \right)^{-1}.
\end{aligned}$$

This is a multivariate extension of the form from Section 3.2. When there are no nuisance parameters $\partial g_j / \partial \nu = 0$ which results in $\widehat{W}_2(\theta) = 0$, so we get the multivariate analog of the form from Section 2.3.

4.3 Numerical Studies Using Canadian Climate Data

4.3.1 Introduction

The theoretical results in Chapters 2 and 3 are specific cases ($J = 2$) of the results shown in this chapter. The simulations in Chapters 2 and 3 verify the theoretical results and show that composite empirical likelihood performs comparably to empirical likelihood. Given this we will not examine the generalized cases presented in this chapter through simulation, but we will examine real data. We demonstrate a case where there is a single parameter of interest along with multiple nuisance parameters, and a case where there is a single parameter but the estimating equations are based on all the data instead of a single subset.

The data are provided by Dr. Rosalind Bueckert of the University of Saskatchewan Department of Plant Science. The data are measured throughout the primary agricultural region of Canada in order to examine how climate affects the annual yields of 15 crops. The region spans the provinces of Alberta, Manitoba and Saskatchewan with weather data collected from 41 locations. The data include daily minimum, maximum and average temperature along with precipitation, but we will focus on the temperature data during the summer months of June, July and August.

4.3.2 Climate Change

A current scientific question of interest concerns climate change. There are many viewpoints concerning the current state of the climate along with prediction of how the climate will change in the future. A simple way to measure this effect is to examine if a given summary measure of the temperature is changing over time. This can be done using a simple linear regression model and examining the confidence interval for the slope to determine if (and in what direction) the measure may be changing. We use a high dimensional construction of composite empirical likelihood to estimate the slope parameter and create a corresponding 95% confidence interval. Our variables of interest are the change in temperature over a 30 year span (1976 - 2005) for the summer months of June, July and August. The temperatures are recorded in degrees Celsius.

The three variables we will examine are the median of the daily average temperature, the 10% quantile of the daily minimum temperature and the 90% quantile for the daily maximum temperature by month, resulting in 30 observations for each month-variable combination at each location. We use the following model

$$E(Z_j) = \mu_j + \tau(year)$$

where μ_j is the mean for site j, and the variable $year$ is index so that 1976 is 1/30 and 2005 is 1. The parameter of interest is τ, which measures the average temperature change of all j sites for the last 30 years. The estimating equation for each location j is

$$g_j(z_{ij}, \theta) = \begin{bmatrix} z_{ij} - \mu_j - \tau(year) \\ year \left\{ z_{ij} - \mu_j - \tau(year) \right\} \end{bmatrix}.$$

Due to a high number of missing observations seven of the locations were removed. The locations used in the analyses are shown in Figure 4.1.

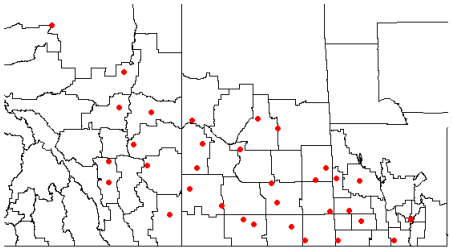


Figure 4.1: Locations of recorded Canadian weather data for climate change analysis.

The distribution of the test statistic follows from Theorem 4.2 We estimate the quadratic and covariance matrix using the empirical version shown in Section 4.2, and we compute the weighted chi square using the method from Section 2.3 (we will use the MCELE as our estimate of θ). We first examine each province separately, then extend the analysis to the entire region. For reference Alberta is the westernmost province and Manitoba is the easternmost province.

Table 4.1: 95% confidence intervals of τ for each variable from the province of Alberta for June, July and August.

	Variable		
Month	Median of Average	10% Quantile of Minimum	90% Quantile of Maximum
Jun	(-0.7724, 0.3240)	(-0.3493, 0.6135)	(-2.1385, -0.3710)
Jul	(-0.1017, 1.1881)	(-0.3623, 0.7193)	(-0.1089, 2.0086)
Aug	(-0.2134, 1.7476)	(-0.6552 , 0.9822)	(0.9775, 3.2463)

Table 4.2: 95% confidence intervals of τ for each variable from the province of Manitoba for June, July and August.

	Variable		
Month	Median of Average	10% Quantile of Minimum	90% Quantile of Maximum
Jun	(-1.5102, 0.6153)	(-0.3709, 1.9033)	(-1.9180, 0.7730)
Jul	(-0.7798, 0.6267)	(-0.9602, 0.1134)	(-2.0350, -0.0378)
Aug	(-0.6196, 1.6807)	(1.1154, 2.7239)	(-1.1345, 2.0673)

Table 4.3: 95% confidence intervals of τ for each variable from the province of Saskatchewan for June, July and August.

	Variable		
Month	Median of Average	10% Quantile of Minimum	90% Quantile of Maximum
Jun	(-2.1295, -0.3937)	(-0.5641, 1.2623)	(-2.5248, -0.4579)
Jul	(-1.0727, 0.4143)	(-0.6417, 0.1733)	(-1.2029, 0.8535)
Aug	(-0.6412, 1.7125)	(-0.1435, 1.4712)	(0.1092, 2.9988)

Table 4.1 shows that in Alberta over the last 30 years the 90% quantile of the maximum temperatures have changed, while the median and minimum has not significantly increased or decreased. The measure of the high and middle temperature decreased in June, stayed the same in July and increased in August. The lower endpoint of the confidence interval for July is very close to zero, suggesting that even if the maximum temperature has decreased, the amount has been very little.

The results from Table 4.2 for Manitoba indicate that the 10% quantile of the minimum recorded temperatures during the month of August has increased over the last 30 years while the 90% quantile of the maximum temperature has decreased in July. No other variables have appeared to have increased or decreased over the last 30 years.

Table 4.3 indicate that the only changes in the center of the agriculture region (Saskatchewan) involve the 90% quantile of the maximum. In June this measure has decreased, and in August it has increased.

Over these three areas we see evidence that in the central and western part of the agricultural region the higher temperatures in June have decreased some while high temperatures are increasing during the month of August. Manitoba has seen a decrease in high temperatures during the month of July, but instead of the increased maximums in August witnessed in the other two provinces, the increase is seen in the 10% quantile of the minimum. So it appears that there is an overall change in climate during the month of August that suggests a warming trend, but the actual increase in the east is seen by the low temperature increasing over the last 30 years.

Next we extend this approach to cover the entire region. The model remains the same, but now the composite empirical likelihood has $J = 38$, with one shared parameter τ and a μ_j for each j, so θ is a 39×1 vector. Again we present the 95% confidence intervals of τ for each month and variable combination.

Table 4.4: 95% confidence intervals of τ for each variable from all three Canadian provinces for June, July and August.

	Variable		
Month	Median of Average	10% Quantile of Minimum	90% Quantile of Maximum
Jun	(-1.4698, -0.1105)	(-0.3527, 0.9945)	(-2.1467, -0.3479)
Jul	(-2.6089, 1.1656)	(-0.5566, 0.2004)	(-1.1118, 0.7727)
Aug	(-0.4381, 1.6625)	(0.0276, 1.5415)	(0.3119, 2.8121)

The confidence intervals from Table 4.4 reveals that over the last 30 years that there has been a decrease in the 90% quantile of the maximum daily temperature in June, an increase in the 90% quantile of the maximum daily temperature in August and an increase in the 10% quantile of the minimum temperature in August. There has been no significant change in July of average or extremes temperatures over the last 30 years. These results reflect a similar pattern to what we saw with the individual provinces. It appears that over the entire region the high June temperatures have gone down, while there is a temperature increase during August. The increase in August does not show up in the median; instead the region does not have low temperatures as cold as it did 30 years ago, and the high temperatures are getting hotter.

4.3.3 Kernel Smoothing Parameter for Weighted Mean Prediction

An issue with any sort of spatial analysis is the limited number of observed locations. In many cases the data are unevenly spaced, further complicating inference and model fitting over the entire area. Spatial analysis addresses how correlated the data are over the entire area, which entails estimating the range parameter for a variogram, and then using this information to predict values at unobserved locations by a process called kriging. The two main limitations with this approach are an assumption of normally distributed errors and there cannot be more than one observation at a given location.

Our proposed alternative is to use the composite empirical likelihood to estimate the smoothing parameter of a kernel density in order to create a weighted average model to predict values at

unobserved locations. This method does not require a distributional assumption for the errors and uses more than one observation at a given location. The kernel density function is used to determines a set of weights based on how far an observation is from the current point of interest. These weights are used to compute a weighted mean, which becomes the estimate of the expected value for an unobserved location. Kernel densities utilize a bandwidth parameter h, which changes the ratio of weight to distance for the data points. In addition to the value of h determining the weights, this also gives a relative idea of how widespread a given effect is, with a larger h indicating that values are similar at larger distances.

The weighted average for an unobserved location z_0 is

$$\widehat{E(Z_0)} = \sum_{j=1}^{J} \frac{K_h(d_{j,0})}{\sum_{j=1}^{J} K_h(d_{j,0})} z_j$$

where $K_h(d)$ is the kernel function and $d_{j,0}$ is the distance between the j-th site and the unobserved site. Dividing by the sum of the weights is for standardization so that the weights sum to one.

Using a minimization of sums of square criterion, for a given site j we find the h that minimizes

$$\sum_{i=1}^{n} \left(z_{i,j} - \sum_{k \neq j}^{J} \frac{K_h(d_{j,k})}{\sum K_h(d_{j,k})} z_{i,k} \right)^2 .$$

If there is a unique minimum where the gradient is 0 with respect to h, we have the following

$$\sum_{i=1}^{n} \left\{ \left(z_{i,j} - \sum_{k \neq j}^{J} \frac{K_h(d_{j,k})}{\sum K_h(d_{j,k})} z_{i,k} \right) \sum_{k \neq j}^{J} \frac{K_h'(d_{j,k}) \sum K_h(d_{j,k}) - K_h(d_{j,k}) \sum K_h'(d_{j,k})}{\left(\sum K_h(d_{j,k}) \right)^2} z_{i,k} \right\} = 0$$

where $K_h'(d)$ is the first derivative of the kernel function. We use this solution to define the estimating equations at each location.

We use the following kernel density function

$$K_h(d) = \exp\left(\frac{-d}{h} \right)$$

where d denotes the distance between two points. This is not a standard choice of kernel density function, but we are using this function due to its similarity to the exponential variogram model.

The estimator of h is the MCELE derived from the composite empirical likelihood with the following estimating equation

$$g_j = \left(z_{i,j} - \sum_{k \neq j}^{J} \frac{\exp(-d_{j,k}/h)}{\sum_{jk \neq j}^{J} \exp(-d_{j,k}/h)} z_{i,k} \right) \times \left(\sum_{k \neq j}^{J} \frac{\frac{d_{j,k}\exp(-d_{j,k}/h)}{h^2} \sum_{k \neq j}^{J} \exp(-d_{j,k}/h) - \exp(-d_{j,k}/h) \sum_{k \neq j}^{J} \frac{d_{j,k}\exp(-d_{j,k}/h)}{h^2}}{\left(\sum_{k \neq j}^{J} \exp(-d_{j,k}/h) \right)^2} z_{i,k} \right).$$

The distance $d_{j,k}$ between two locations j and k is computed by the haversine formula

$$d_{j,k} = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{\psi_k - \psi_j}{2} \right) + \cos(\psi_j)\cos(\psi_k)\sin^2\left(\frac{\lambda_k - \lambda_j}{2} \right)} \right)$$

where ψ_j, ψ_k are the latitudes of locations j and k (in radians), λ_j, λ_k are the longitudes of locations j and k (in radians) and R is the radius of the sphere. Here we set $R = 6371$ which results in $d_{j,k}$ being the distance between locations j and k in kilometers.

The data consist of the daily mean, minimum and maximum temperature for June, July and August for the year 2005. Certain locations do not have observations for the three months of interest, so locations missing any observations are removed. The locations used to fit the bandwidth parameter are shown in Figure 4.2.

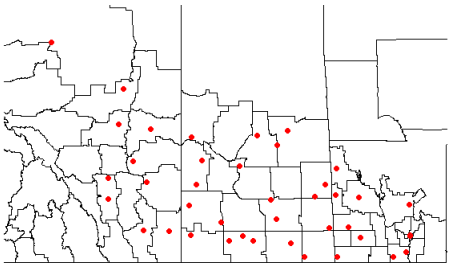


Figure 4.2: Locations of recorded Canadian weather data for bandwidth parameter estimation.

Due to the complexity of the composite empirical likelihood function in this example we cannot guarantee that the optimization algorithm finds the global minimum of ℓ_{CE}. To account for this

we fit the models using the following start values for h: 5, 50, 100, 200, 500, 1500, 5000, 10000 and 30000. We then select the solution with the lowest value of $\ell_{CE}(\theta_{CE})$. The maximum composite empirical likelihood estimators for h are shown in Table 4.5.

Table 4.5: Estimate of h for each variable from all three Canadian provinces for June, July and August using 2005 data.

	Variable		
Month	Daily Mean	Daily Minimum	Daily Maximum
Jun	49.17	56.42	1.04
Jul	1.01	73.64	1.03
Aug	52.72	76.02	55.51

We see from examination of Table 4.5 the bandwidth estimates for the June and July maximum along with the July mean are quite small. The remaining parameter estimates are fairly similar, with the minimum of July and August being slightly larger. Although the bandwidth parameter does not have any natural interpretation in terms of distance these results do give a general idea of how 'uniform' over the area these temperatures are. For example the July and August minimums would be similar over a larger area than the June minimum, whereas the June and July maximum temperatures would not necessarily be consistent over a large area. So during the first two months of the summer the high temperatures may not be similar over any given area; instead they resemble local phenomenon. From the standpoint of prediction this means for the June maximum and July mean and maximum temperature an accurate prediction for an unobserved site will be difficult since there is not a lot of correlation for those values, whereas the other variables show some spatial correlation with nearby neighbors. The slightly higher bandwidth parameters for the July and August show that those minimum temperatures may be a bit more uniform over a larger area; so to observe a change in the minimums we have to go a greater distance. Overall this pattern means that July and August minimums have a slightly larger area of effect, so extreme low (or high) minimums would be experienced over a larger area than we would see with the other variables.

4.4 Discussion

We establish theoretical results for the general J component case of composite empirical likelihood which parallel the theoretical results pertaining to empirical likelihood, composite likelihood, and composite empirical likelihood when $J = 2$. This generalization allows for composite empirical likelihood to be extended to complex systems, while the lack of distributional requirements for testing enhances the usefulness of this methodology.

The main challenge with composite empirical likelihood is estimation of the necessary quantities to compute the weighted chi square distribution. The distribution of T_{CE} is dependent on unknown parameters, so the potential that our conclusions are in error increases. Furthermore the eigenvalues for the weighted chi square require the moments of the random variables, so either the moments have to be estimated using the data or assumed. As a non-parametric method ideally we would not want to make assumptions about the moments, but if there is previous knowledge relating to the required moments this may be preferable to using the data to approximate these values.

The analysis performed using the Canadian climate data would be problematic using parametric methods. Determining the distribution of quantiles is challenging in itself, but also requires the data mechanism be known. Given this determining the distribution of quantiles (even asymptotically) is not practical, and as a result makes computing confidence intervals challenging. Our method to estimate an optimal bandwidth parameter not only demonstrates an example of a high dimensional conditional composite empirical likelihood but suggests a data driven methodology for selecting an optimal bandwidth.

Chapter 5

Summary

5.1 Introduction

The two biggest strengths of the composite empirical likelihood are the ability to arbitrarily set sample sizes in each piece and the flexibility of the estimating equations. By borrowing from two methods which are both designed to overcome the main limitation of Fisher likelihood, we are able to take the best feature of each method and design a likelihood that maintains non-parametric properties and increases computational tractability.

The composite empirical likelihood does not require knowledge or assumption about the probability function itself, but there are assumptions about the estimating equations which increase the possibility of model misspecification. We have not explored how improper specification of the estimating equations affects the validity of any inferential methods using composite empirical likelihood, but it stands to reason the performance would be comparable to what would be observed with empirical likelihood if the estimating equations were incorrect.

There are still many questions about how to efficiently adapt composite empirical likelihood to certain problems, and development of more complex statistical tests. We briefly cover a few of these questions in the remainder of this chapter.

5.2 Alternate Constructions of the Composite Empirical Likelihood

A common issue when working with methods that require numerical optimization (such as composite empirical likelihood) is that large sample sizes can significantly increase the amount of time needed to compute solutions. One method to reduce computation time would be to choose a random subset of the data to work with, but the resultant analysis may be incorrect simply because the random sample was not representative of the data (or population).

Resampling method, such as the jackknife use subsets of the data to determine the precision of the sample statistic. Perhaps the most known resampling method is the bootstrap (Efron, 1979, 1981) which draws from the sample with replacement to empirically build the distribution of the test statistic.

The composite empirical likelihood is already designed in such a way that easily lends itself to random resampling methodology. We can define an appropriate resample size, and taking this one step further set a component size. Using the J likelihood components, each created from a random data subset n_j, we can either create a single random draw composite empirical likelihood or we could repeat the process in a bootstrap fashion and empirically build the distribution of the MCELE along with the distribution of the estimators at defined confidence levels.

Given the choice of number of likelihood components and sample sizes of the composite empirical likelihood, we might wish to determine what is the optimal number of likelihood components for a given sample size or use an incomplete resample to further reduce the computational load. We may also wish to consider the case of using a bootstrap sample for each likelihood component, or take bootstrap samples from multiple likelihood components.

The following examples demonstrate these methods formally. Let n be the total sample size, and we will use J to denote the number of likelihood components.

Example 5.1 The simplest method is to take J samples without replacement where each sample size n_j for $j = 1, \ldots, J$ is much smaller than n. In this scenario we do not use all the data, however the asymptotic behavior of the composite empirical likelihood would follow directly from the results in Chapter 4. In this scenario the goal is to determine the optimal balance between choice of n_j and J.

Example 5.2 A bootstrap method is to create composite empirical likelihood using a small J, except the sample sizes n_b used for each component are fairly large and the samples to form the likelihood can be collected via resampling. Again we would want to balance the size of n_b against how many likelihood components (J) we create. This would be more representative of drawing from the population to create the composite empirical likelihoods, but the theoretical results from Chapter 4 may not hold due to the bootstrap samples.

Example 5.3 A likelihood bootstrap style method is to create a large number of likelihood components by exhaustively sampling from the data. Likelihood components using sample sizes n_b

are drawn with replacement, resulting in a bootstrap sample of empirical likelihoods. We then draw a small subset J_b of the bootstrapped empirical likelihoods (either with or without replacement) many times to create a distribution of composite empirical likelihoods. With this approach the empirical distribution of composite empirical likelihoods can be used to assess the accuracy of point estimates and confidence boundaries.

Determining how to balance the choice of sample and component size is limited by the fact that our theoretical results only show that the MCELE and the test statistics converge, but do not specify an order of convergence based on sample size and number of likelihood components. Empirical likelihood normally has a coverage error of n^{-1}, but also inherits the Bartlett correctability of parametric likelihood so the coverage error can be reduced to n^{-2} (DiCiccio et al., 1991). It is plausible that composite empirical likelihood also has similar convergence orders since each component in the composite empirical likelihood is a proper empirical likelihood, but we would also have to address how choice of J affects the order of convergence.

5.3 Estimating Equations and Optimization

The ease of applicability and distribution of the test statistic is dependent on the choice of estimating equations. Chapter 3 demonstrates an example where there is an optimal choice for the estimating equations, both in terms of accuracy of coverage and the distribution of the test statistic.

Despite the computational advantages gained from composite empirical likelihood, optimization can still present challenges. Many optimization methods derive the inverse of the Hessian matrix in order to optimize the constrained function. For an empirical likelihood component the Hessian is based on the estimating equations. Since each likelihood component $\ell_E^{(j)}(\theta)$ requires an optimization step for a fixed value of θ, there are problems if the estimating equations result in likelihood components having a singular Hessian. As an example could have assumed that the mean temperature for the Canadian climate data from Chapter 4 is

$$E(Z_j) = \mu + \mu_x(longitude_j) + \mu_y(latitude_j).$$

Although the design matrix using all locations would not be singular, the design matrix of each likelihood component would be

$$X_j = \begin{bmatrix} 1 & longitude_j & latitude_j \\ \vdots & \vdots & \vdots \\ 1 & longitude_j & latitude_j \end{bmatrix}$$

which is singular and therefore each likelihood component has an invertible Hessian. Since there is no constraint on how the components are broken up we could combine two or more locations into a single likelihood component to avoid this issue, but that decreases the functionality of the composite empirical likelihood. If we were interested in parameters that both involved time and location simultaneously the design of each component becomes even more complex.

There are alternate optimization algorithms, such as the evolutionary and swarm, that do not rely on inversions of matrices. The downside with these methods is that they tend to be slower and more computationally demanding than algorithms based on Newton-Raphson's method. In order to maximize the value of composite empirical likelihood it is necessary to balance the computational gains of computing in parallel versus loss of using a more numerically intensive optimization scheme.

5.4 Composite Multiple Hypothesis Testing

A common use of the likelihood ratio test is hypothesis testing for the validity of using a complex versus a simple model. In this context we test hypotheses of the form

$$H_0 : \theta_1 = \ldots = \theta_k = 0$$

$$H_A : \text{at least one } \theta_k \text{ is not zero}$$

or

$$H_0 : \theta_1 = \ldots = \theta_k$$

$$H_A : \text{at least one } \theta_k \text{ is different}$$

where a decision of do not reject would indicate there is no additional information or explanatory power gained from the more complex model.

The theorems and corollaries worked out in this dissertation do not address hypothesis tests of this nature. Owen (2001) gives a brief example of how to perform a composite multiple hypothesis test for an ANOVA setting focusing on the second hypothesis (all population means are the same versus at least one is different). If we assume all parameters are equal then for k groups

$$\arg\max_{\mu} -2\log R_E(\mu,\ldots,\mu) \longrightarrow \chi^2(k-1)$$

in distribution as $n \longrightarrow \infty$ when the null hypothesis is true (see Owen, 2001, Chapter 4). Empirical likelihood has the advantage over ANOVA since the variances do not have to be equal between groups.

Unlike empirical likelihood, composite empirical likelihood allows for correlation between the groups, hence composite multiple hypothesis testing may not conform to normal theory results. In the case of composite empirical likelihood we would use

$$T_{CE} = 2\ell_{CE}(\mu) - 2\ell_{CE}(\mu_1,\ldots,\mu_k) \tag{5.1}$$

which would likely have a weighted chi square distribution. The advantage of using composite empirical likelihood is that we would not have to assume independence between the groups. The triangular array empirical likelihood theory assumes independence between all observations, and consequently between groups, where composite empirical likelihood would only assume independence within groups.

Since we have shown many cases where empirical likelihood is a specific case of composite empirical likelihood we would expect that the asymptotic distribution of the test statistic shown in Equation 5.1 would be a chi square with $k-1$ degrees of freedom if all j groups are independent.

5.5 Conclusion

Research questions utilizing large amounts of data are quickly becoming commonplace. Despite the advances in computing power many methods are hindered by storage, computing speed or the limits of numerical precision. These issues are even more apparent with non-parametric methods. Given many modern research questions involve data with complex correlation structures, missing data, high measurement variability and nonstandard distributions add to the necessity of computationally tractable non-parametric methods.

We have expanded on the general framework of both empirical likelihood and composite likelihood with the development of a method that combines the distinguishing features of both these methods. Furthermore we have shown the asymptotic behavior of the resultant estimators using composite empirical likelihood follow naturally from empirical likelihood and the asymptotic distribution of the test statistic is a similar form seen in composite likelihood. With the additional assumption of independence between the likelihood components the results derived using a composite empirical likelihood are identical to those seen with Fisher likelihoods.

The distinguishing point of composite empirical likelihood compared to composite likelihood is that assumptions are not directly based on the probabilistic behavior defining the variables, but on properly defining the moment dependency through the estimating equations. As we have shown, the accuracy of the bivariate confidence region for the means is dependent on properly defining the conditional moment. This places an additional set of assumptions that the empirical likelihood does not require, but there may be situations where these moment assumptions are obvious, or preferable to no answer.

There are still many aspects of composite empirical likelihood to examine. Determining if the composite empirical likelihood is Bartlett correctable would be of both theoretical and practical interest, given that with very few exceptions (see Lazar and Mykland, 1999) the empirical likelihood and parametric likelihoods permit a Bartlett correction. The correction holding for the composite empirical likelihood would further confirm the idea that the composite empirical likelihood is a general form, with empirical likelihood and parametric likelihoods being specific cases. From an applied aspect the Bartlett correction would allow for a correction to the test statistic, resulting in more accurate inference.

Ultimately composite empirical likelihood presents a new generalization of likelihood methods which performs comparably to empirical likelihood. Although we did not examine performance versus parametric likelihood, based on results shown using empirical likelihood we predict that our method would also perform better than parametric likelihood if the distribution is improperly specified. An interesting question that our results bring up concerns the exact nature of a likelihood estimator. In many cases, the asymptotic distribution of the values that maximize the likelihood function using Fisher, composite, empirical and composite empirical methods are identical despite the perceived differences in each method. Examination of these equivalent cases may further our understanding of the relationship between parametric distributions, sufficient statistics and the overall general behavior of functions of data.

Appendix

A Stochastic Order Notation and Properties

Definition A.1 Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. We say X_n converges in probability to zero, written as $X_n = o_p(1)$ if $\forall\ \varepsilon > 0\ P(|X_n| > \varepsilon) \to 0$ as $n \to \infty$.

Definition A.2 Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. We say X_n is bounded in probability, written as $X_n = O_p(1)$ if $\forall\ \varepsilon > 0\ \exists\ M_\varepsilon$ such that $P(|X_n| > M_\varepsilon) < \varepsilon\ \forall\ n$.

Properties. Let $\{X_n\}$ and $\{Y_n\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ and let $\{a_n\}$ and $\{b_n\}$ be a sequence of strictly positive real numbers. Then

(1) $X_n = o_p(a_n)$ if and only if $a_n^{-1} X_n = o_p(1)$, and

(2) $X_n = O_p(a_n)$ if and only if $a_n^{-1} X_n = O_p(1)$.

Furthermore if $X_n = o_p(a_n)$ and $Y_n = o_p(b_n)$, then

(1) $X_n Y_n = o_p(a_n b_n)$, and

(2) $X_n + Y_n = o_p(\max(a_n, b_n))$, and

(3) $X_n^r = o_p(a_n^r)$ for $r > 0$.

If $X_n = O_p(a_n)$ and $Y_n = o_p(b_n)$, then

(1) $X_n Y_n = o_p(a_n b_n)$.

B Matrix Inverse

Let

$$S = \begin{bmatrix} S_{11} & 0 & S_{13} \\ 0 & S_{22} & S_{23} \\ S_{31} & S_{32} & 0 \end{bmatrix}$$

where S_{11} is $p_1 \times p_1$, S_{22} is $p_2 \times p_2$, S_{13} is $p_1 \times p_3$, S_{23} is $p_2 \times p_3$, S_{31} is $p_3 \times p_1$ and $S_{32} = p_3 \times p_2$. Also S_{11}^{-1} and S_{22}^{-1} exists. Then

$$S^{-1} = \begin{bmatrix} S_{11}^{-1}\left(I - S_{13}S_{33.1}^{-1}S_{31}S_{11}^{-1}\right) & -S_{11}^{-1}S_{13}S_{33.1}^{-1}S_{32}S_{22}^{-1} & S_{11}^{-1}S_{13}S_{33.1}^{-1} \\ -S_{22}^{-1}S_{23}S_{33.1}^{-1}S_{31}S_{11}^{-1} & S_{22}^{-1}\left(I - S_{23}S_{33.1}^{-1}S_{32}S_{22}^{-1}\right) & S_{22}^{-1}S_{23}S_{33.1}^{-1} \\ S_{33.1}^{-1}S_{31}S_{11}^{-1} & S_{33.1}^{-1}S_{32}S_{22}^{-1} & -S_{33.1}^{-1} \end{bmatrix}$$

where

$$S_{33.1} \equiv \left(S_{31}S_{11}^{-1}S_{13} + S_{32}S_{22}^{-1}S_{23}\right).$$

References

Arnold, B. C. and Strauss, D. (1991). Pseudolikelihood estimation: Some examples. *Sankhyā: The Indian Journal of Statistics*, 53(2):233–243.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*. Chapman & Hall/CRC, Boca Raton.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B*, 26(2):192–236.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 24(3):179–195.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of anaysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, 25(2):290–302.

Casella, G. and Berger, R. L. (2002). *Statistical Inference (2nd ed)*. Duxbury, Pacific Grove.

Chandler, R. E. and Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, 94(1):167–183.

Chen, J. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, 80(1):107–116.

Chen, J. and Sitter, R. R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica*, 9(2):385–406.

Chen, J., Variyath, A. M., and Abraham, B. (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, 17(2):426–443.

Conover, W. J. (1999). *Practical Nonparametric Statistics (3rd ed)*. John Wiley & Sons, Inc, New York.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3):729–737.

DiCiccio, T. J., Hall, P., and Romano, J. P. (1991). Empriical likelihood is Bartlett-correctable. *The Annals of Statistics*, 19(2):1053–1061.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Efron, B. (1981). Nonparametric standard errors and confidence intervals. *The Canadian Journal of Statistics*, 9(2):139–172.

Ferguson, T. S. (1996). *A Course in Large Sample Theory*. Chapman & Hall, London.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society, A*, 22:309–368.

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

Grendar, M. and Judge, G. G. (2010). Revised empirical likelihood. Technical Report 1106, CUDARE Working Paper Series, University of California at Berkeley.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction (2nd ed)*. Springer, New York.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1):73–101.

Kent, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika*, 69(1):19–27.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics*, 25(5):2084–2102.

Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models. *Statistica Sinica*, 4(1):199–218.

Larsen, R. J. and Marx, M. L. (1986). *An Introduction to Mathematical Statistics and its Applications*. Prentice-Hall, Englewood Cliffs.

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

Lazar, N. A. and Mykland, P. A. (1999). Empirical likelihood in the presence of nuisance parameters. *Biometrika*, 86(1):203–211.

Liang, K. Y. (1987). Extended Mantel-Haenszel estimating procedure for multivariate logistic regression models. *Biometrics*, 43(2):289–299.

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

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

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

Moschopoulos, P. G. and Canada, W. B. (1984). The distribution function of a linear combination of chi-squares. *Computers & Mathematics with Applications*, 10(4-5):383–386.

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

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120.

Owen, A. B. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, 19(4):1725–1747.

Owen, A. B. (1992). Empirical likelihood and generalized projection pursuit. Technical Report 393, Department of Statistics, Stanford University.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, Boca Raton.

Qin, J. (1994). Semi-empirical likelihood ratio confidence intervals for the difference of two sample means. *The Annals of the Institute of Statistical Mathematics*, 46(1):117–126.

Qin, J. (2000). Combining parametric and empirical likelihoods. *Biometrika*, 87(2):484–490.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22(1):300–325.

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

Qin, J. and Zhang, B. (2007). Empirical-likelihood-based inference in missing response problems and its application in observational studies. *Journal of the Royal Statistical Society Series B*, 69(1):101–122.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. Wiley, New York.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6):110–114.

Shi, J. and Lau, T. (2000). Empirical likelihood for partially linear models. *Journal of Multivariate Analysis*, 72(1):132–148.

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association*, 70(352):865–871.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29(3-4):350–362.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62.

Wu, C. and Rao, J. N. K. (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *The Canadian Journal of Statistics*, 34(3):359–375.