

BIAS CORRECTION FOR WELFARE MEASURES IN NON-MARKET VALUATION

by

RUI ZHANG

(Under the Direction of J. Scott Shonkwiler)

ABSTRACT

Welfare measures in non-market valuation can be biased by nonrandom sampling or nonlinear transformation of estimated parameters. This dissertation consists of two studies focusing on correcting each type of bias.

The first study compares the performances of the delta method, jackknife, and bootstrap in correcting nonlinear transformation bias through Monte Carlo simulation. The simulated models include Poisson, logit, probit, and misspecified probit. The results indicate that the delta method and jackknife can correct the bias of welfare measures for the listed models. With distributional misspecification, the delta method outperforms the jackknife. The bootstrap result has more expected bias, but also the smallest mean square error, especially when the sample size is large enough. Furthermore, the delta method was modified by correcting the bias of estimated parameters beforehand, and then adding the first order bias term or replacing parameter estimates with corrected ones. The simulation results show that adding first order bias term can lower the MSE of willingness to pay (WTP) estimates at the cost of increasing their expected bias. The simulation results show that the delta method is also effective to estimate median WTP.

The second study address the sampling bias decomposed as one part correlated with covariates and the other part correlated with the result variable. Non-response is common in

surveys used in non-market valuation studies which can bias the parameter estimates and mean WTP estimates. One approach to correct this bias is to reweight the sample so that the distribution of the characteristic variables of the sample can match that of the population. Kernel Mean Matching (KMM) is used to produce resampling weights in a non-parametric manner. KMM's performance is tested through Monte Carlo simulations under multiple scenarios and it is shown that KMM can effectively correct mean WTP estimates, especially when the sample size is small and sampling process depends on covariates. The results also reveal KMM's robustness to skewed bid design and model misspecification.

INDEX WORDS: bias, WTP, the delta method, jackknife, bootstrap, sampling bias, kernel mean matching, Monte Carlo

BIAS CORRECTION FOR WELFARE MEASURES IN NON-MARKET VALUATION

by

RUI ZHANG

M.S., Chinese Academy of Agricultural Sciences, 2012

B.S., Renmin University of China, 2009

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

2017

© 2017

Rui Zhang

All Rights Reserved

BIAS CORRECTION FOR WELFARE MEASURES IN NON-MARKET VALUATION

by

RUI ZHANG

Major Professor: J. Scott Shonkwiler

Committee: Craig Landry
Travis A. Smith

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2017

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	ix
CHAPTER	
1 INTRODUCTION	1
1.1 Background	1
1.2 Motivation and Major Objectives	2
1.3 Non-market Valuation Model with Welfare Measure Estimation	2
1.4 Organization	8
2 COMPARISON OF BIAS CORRECTION METHODS: MONTE CARLO SIMULATION	9
2.1 Introduction	9
2.2 Bias of Welfare Measures in Non-Market Valuation	11
2.3 Bias Correction Methods	13
2.4 Comparison of Bias Correction Methods	21
2.5 Poisson Model	25
2.6 Logit Model	31
2.7 Probit Model	35
2.8 Misspecified Probit Model	39
2.9 Small Sample Parameter Estimation Bias	43

2.10 Conclusion	56
References	58
3 CORRECTING SAMPLING BIAS WITH KERNEL MEAN MATCHING	65
3.1 Introduction.....	65
3.2 Sampling Bias	67
3.3 Kernel Mean Matching	70
3.4 Monte Carlo Simulation: Probit Model	73
3.5 Conclusion	95
References.....	96

LIST OF TABLES

	Page
Table 2.1: Monte Carlo analysis of WTP expected bias for Poisson model.....	27
Table 2.2: Monte Carlo analysis of WTP mean square error for Poisson model	28
Table 2.3: Monte Carlo analysis of WTP expected bias for logit model.....	33
Table 2.4: Monte Carlo analysis of WTP mean square error for logit model	34
Table 2.5: Monte Carlo analysis of WTP expected bias for probit model	37
Table 2.6: Monte Carlo analysis of WTP mean square error for probit model	38
Table 2.7: Monte Carlo analysis of WTP expected bias for misspecified probit model ...	40
Table 2.8: Monte Carlo analysis of WTP mean square error for misspecified probit model	41
Table 2.9: Monte Carlo analysis of mean WTP expected bias for logit model: combine the delta method and Cordeiro's method.....	50
Table 2.10: Monte Carlo analysis of mean WTP mean square error for logit model: combine the delta method and Cordeiro's method.....	51
Table 2.11: Monte Carlo analysis of median WTP expected bias for logit model: combine the delta method and Cordeiro's method.....	54
Table 2.12: Monte Carlo analysis of median WTP mean square error for logit model: combine the delta method and Cordeiro's method.....	55
TABLE 3.1 Simulation Results for Non-Response Bias (N=100, M=50, $a = .1$, $b = .1$, $\rho_1 =$ 0 , $\rho_2 = 10$, B=1000, $\sigma = .01$)	76

TABLE 3.2 Simulation Results for Non-Response Bias ($N=200$, $M=105$, $a = .1$, $b=.1$, $\rho_1 = 0$, $\rho_2 = 10$, $B=1000$, $\sigma = .01$)	77
TABLE 3.3 Simulation Results for Non-Response Bias ($N=400$, $M=224$, $a = .1$, $b=.1$, $\rho_1 = 0$, $\rho_2 = 10$, $B=1000$, $\sigma = .0005$)	78
TABLE 3.4 Simulation Results for Non-Response Bias($N=800$, $M=446$, $a = .1$, $b=.1$, $\rho_1 = 0$, $\rho_2 = 10$, $B=1000$, $\sigma = .0001$)	79
TABLE 3.5 Simulation Results for Sample Selection Bias ($N=100$, $M=50$, $a = .1$, $b=0$, $\rho_1 = .5$, $\rho_2 = 10$, $B=1000$, $\sigma = .01$)	81
TABLE 3.6 Simulation Results for Sample Selection Bias ($N=200$, $M=104$, $a = .1$, $b=0$, $\rho_1 = .5$, $\rho_2 = 10$, $B=1000$, $\sigma = .002$)	82
TABLE 3.7 Simulation Results for Sample Selection Bias ($N=400$, $M=222$, $a = .1$, $b=0$, $\rho_1 = .5$, $\rho_2 = 10$, $B=1000$, $\sigma = .002$)	83
TABLE 3.8 Simulation Results for Sample Selection Bias and Non-Response Bias ($N=100$, $M=50$, $a = .1$, $b=.1$, $\rho_1 = .5$, $\rho_2 = 10$, $B=1000$, $\sigma = .01$)	85
TABLE 3.9 Simulation Results for Sample Selection Bias and Non-Response Bias ($N=200$, $M=105$, $a = .1$, $b=.1$, $\rho_1 = .5$, $\rho_2 = 10$, $B=1000$, $\sigma = .001$)	86
TABLE 3.10 Simulation Results for Skewed Bid Design ($N=100$, $M=48$, $a = .1$, $b=0$, $\rho_1 = 0$, $\rho_2 = 10$, $B=1000$, $\sigma = .001$)	88
TABLE 3.11 Simulation Results for Skewed Bid Design ($N=200$, $M=104$, $a = .1$, $b=0$, $\rho_1 = 0$, $\rho_2 = 10$, $B=1000$, $\sigma = .001$)	89
TABLE 3.12 Simulation Results for Skewed Bid Design ($N=200$, $M=104$, $a = .1$, $b=.1$, $\rho_1 = 0$, $\rho_2 = 10$, $B=1000$, $\sigma = .001$)	90

TABLE 3.13 Simulation Results for Skewed Bid Design ($N=400$, $M=222$, $a = .1$, $b=0$, $\rho_1 = 0$, $\rho_2 = 10$, $B=1000$, $\sigma = .0001$)	91
TABLE 3.14 Simulation Results with Misspecification ($N=200$, $M=105$, $a = .1$, $b=.1$, $\rho_1 = 0$, $\rho_2 = 10$, $B=1000$, $\sigma = .005$)	93
TABLE 3.15 Simulation Results with Misspecification ($N=200$, $M=104$, $a = .1$, $b=0$, $\rho_1 = 0$, $\rho_2 = 10$, $B=1000$, $\sigma = .005$)	94

LIST OF FIGURES

	Page
FIGURE 2.1: Monte Carlo analysis of bias correction in maximum likelihood parameter estimates: compare expected bias under different sample sizes	47
FIGURE 2.2: Monte Carlo analysis of bias correction in maximum likelihood parameter estimates: compare mean square error under different sample sizes	48

CHAPTER 1

INTRODUCTON

1.1 Background

Welfare measures based upon observed behavior or stated preferences are rarely estimated in a straightforward fashion. Haab and McConnell (Valuing Environmental and Natural Resources, 2002) have summarized this as "The need for statistical inference and econometrics arises because individual actions, whether behavior that is observed in quasi-market settings or responses to hypothetical questions, almost never reveal precisely the economic value that a researcher wishes to measure."

In generalized linear models used in non-market valuation, the welfare measures are often derived as a nonlinear function of estimated parameters. Per Jensen's inequality, the nonlinear function of the expectations of the parameters is usually different from the expectation of the nonlinear function of the parameters. Consequently, the nonlinear transformation of estimated parameters can cause biased mean willingness to pay (WTP) estimates. Because the bias is on WTP per capita, it can proportionally cause drastic bias in the aggregate valuation results. Limited research has been done to address this bias or methods for its correction.

Non-response is common and widespread in contingent valuation research, which can bias the WTP results when the non-response occurrence is not random. It is necessary to collect data on non-respondents and correct the sampling bias caused by nonrandom non-response.

1.2 Motivation and Major Objectives

The two major parts of this dissertation deal with nonlinear transformation bias correction and sampling bias correction separately.

For nonlinear transformation bias, we compare the available statistical method for bias correction, including the delta method, jackknife and bootstrap through Monte Carlo simulation. “Misspecified models” are also studied in the simulation study to test the robustness of the bias correction methods. When applying the delta method, in addition to the basic forms that assume unbiasedness of parameter estimates, the parameter estimation bias correction is also studied to explore the potential of accuracy improvement of the delta method.

For sampling bias, we test the performance of a sample-reweight method called kernel mean matching. This method deals with the distributional differences between non-respondents and in-sample respondents, and apply this information in a nonparametric way to derive new parameters estimates and WTP estimates.

1.3 Non-market Valuation Model with Welfare Measure Estimation

In this section the review the models used in following chapters, which are also the most commonly used models in non-market valuation.

1.3.1 Random Utility Models

If we assume that an individual knows her preferences with certainty, but a researcher may fail to observe some component of it and treat it as random; then the individual’s indirect utility function can be written as:

$$v(y, s, q, \varepsilon) = u(y, s, q) + \varepsilon \quad (1)$$

where y is the income, s is a vector of socio-economic characteristics of individual, q is quantity of non-market item, ε is the stochastic component with zero expected value. The individual will accept an offer of an amount A for a change of q (from q^0 to q^1) only if $v(y - A, s, q^1, \varepsilon^1) \geq v(y, s, q^0, \varepsilon^0)$, where ε^1 and ε^0 are identically and independent distributed (iid) random variables with zero mean. Thus,

$$Prob(response \text{ is yes}) = Prob \{ v(y - A, s, q^1, \varepsilon^1) \geq v(y, s, q^0, \varepsilon^0) \} \quad (2)$$

And the welfare measure, in this case, the compensating variation (C), can be defined by

$$v(y - C, s, q^1, \varepsilon^1) = v(y, s, q^0, \varepsilon^0), \quad (3)$$

So $C = C(q^0, q^1, y, s, \varepsilon^0, \varepsilon^1)$ is the individual's compensating variation. Thus, the equivalent condition to (3) is

$$Prob(response \text{ is yes}) = Prob \{ C(q^0, q^1, y, s, \varepsilon^0, \varepsilon^1) \geq A \} \quad (4)$$

The researcher treats C as a random variable and assumes some cumulative distribution function (cdf) $G_c(\cdot)$. Thus, (4) can be written as

$$Prob(response \text{ is yes}) = 1 - G_c(A) \quad (5)$$

Haneman (1984) and Cameron (1988) formulate statistical models separately based on (2), (4) and (5). McConnell (1990) shows that those two approaches are dual to each other. Here we illustrate the models by Haneman's approach.

Haneman (1984) specifies a certain indirect utility function $v(y, s, q, \varepsilon)$ and a particular cdf for ε . Many models in existing literature using Haneman's approach can be nested in a Box-Cox utility function

$$v_q = \alpha_q + \beta_q \left[\frac{y^\lambda - 1}{\lambda} \right] + \varepsilon_q, \quad q = 0 \text{ or } 1 \quad (6)$$

when $\alpha_1 \geq \alpha_0$ and $\beta_1 \geq \beta_0 > 0$, (6) can be treated as a CES utility function in q and y . And the corresponding compensating variation is

$$C = y - \left[\frac{\lambda\alpha + \beta_0 y^\lambda + \beta_1 - \beta_0 - \lambda\eta}{\beta_1} \right]^{\frac{1}{\lambda}} \quad (7)$$

where $\alpha = \alpha^1 - \alpha^0$, and $\eta = \varepsilon^1 - \varepsilon^0$. By assuming $\beta_1 = \beta_0 \stackrel{\text{def}}{=} \beta$, we get

$$C = y - \left[y^\lambda - \frac{\alpha}{\beta} - \frac{\eta}{\beta} \right]^{\frac{1}{\lambda}} \quad (8)$$

Then by specifying value of λ , we can get linear model and logarithmic model, which are mostly used in CV due to their simplicity. λ is the parameter that determines income elasticity. When $\lambda = 1$, (6) becomes the linear model

$$v_q = \alpha_q + \beta_q y + \varepsilon_q, \quad q = 0 \text{ or } 1 \quad (9)$$

$$C = \frac{\alpha + \eta}{\beta} \quad (10)$$

When $\lambda = 0$, (6) becomes the logarithmic model

$$v_q = \alpha_q + \beta_q \ln y + \varepsilon_q, \quad q = 0 \text{ or } 1 \quad (11)$$

$$C = y \left[1 - e^{-\left(\frac{\alpha+\eta}{\beta}\right)} \right] \quad (12)$$

It is noticeable that C is a non-linear combination of α, β and η . The last step to complete the model is to specify the probability distribution of η . For a linear model as (9), when $G_\eta(\cdot)$ is standard normal cdf, (5) becomes a probit model

$$Prob(response \text{ is yes}) = \Phi(\alpha - \beta A) \quad (13)$$

When $G_\eta(\cdot)$ is standard logistic cdf, (5) becomes a logit model

$$Prob(response \text{ is yes}) = \frac{1}{1 + e^{-\alpha + \beta A}} \quad (14)$$

Other distributions are also used to describe distribution of η , including lognormal, log-logistic and Weibull. Kerr (1996) gave scaled and shifted variations for such distributions used in dichotomous choice models. An alternative to specifying a distribution is the non-parametric method, which will be covered in next section.

1.3.2 Theoretical Restrictions for the Random Utility Model

For a linear model, as shown in (10), C has a domain $(-\infty, +\infty)$, which does not comply with a theoretical economic restriction.

$$C(q^0, q^1, y, s, \varepsilon^0, \varepsilon^1) \leq y \quad (15)$$

If we see the change from q^0 to q^1 as an improvement, then

$$0 \leq C(q^0, q^1, y, s, \varepsilon^0, \varepsilon^1) \leq y \quad (16)$$

Haab and McConnell (1997, 1998) used three ways to solve the problem: truncating the distribution of the estimate of willingness to pay to be positive, specifying a nonnegative beta distribution in parametric estimation, or employing a distribution-free nonparametric approach. Cooper (2002) adopted a semi-nonparametric method to calculate willingness to pay between zero and respondents' income.

Actually, for Box-Cox models, only when $\lambda = 0$, (16) is satisfied. For other models, this problem can be solved by truncating or censoring the distribution of η .

For welfare measure $C(q^0, q^1, y, s, \varepsilon^0, \varepsilon^1)$, previous literature mostly focused on two summary statistics, the mean C^+

$$C^+ = E\{C(q^0, q^1, y, s, \varepsilon^0, \varepsilon^1)\} \quad (17)$$

And the median C^*

$$G(C^*) = 0.5 \quad (18)$$

In a Box-Cox model, if we assume $\beta_1 = \beta_0 = b$, from (7) we can get

$$C^+ = y - E \left[\left(y^\lambda - \frac{\alpha}{b} - \frac{\eta}{b} \right)^{\frac{1}{\lambda}} \right] \quad (19)$$

$$C^* = y - \left(y^\lambda - \frac{\alpha}{b} - \frac{\eta^*}{b} \right)^{\frac{1}{\lambda}} \quad (20)$$

Where η^* is the median of distribution of η , in a linear model

$$C^+ = \frac{\alpha + E(\eta)}{\beta} \quad (21)$$

$$C^* = \frac{\alpha + \eta^*}{\beta} \quad (22)$$

When the survey data come from questioning different individuals in a population, the individual welfare estimate will be multiplied by the number of people in the population to get an aggregate estimate. Thus, the mean, C^+ , is in accordance with Kaldor-Hicks compensation; while the median, C^* , is in accordance with majority voting.

An extension of the standard dichotomous model is to propose a second bid based on the response of the first bid. If the first response is "yes," the second bid is some amount greater than the first bid; while, if the first response is "no," the second bid is some amounts smaller.

Hanemann (1991) stated this double bound can improve the precision and efficiency of the estimation. McLeod (1999) demonstrates the initial bid may represent information with which respondents update their willingness to pay, thus any expected precision increase may be lost.

1.3.3 Double Bounded Model

In order to improve the information efficiency of WTP in the random utility models, a discrete choices survey with follow-up questions was proposed by Hanemann (1985). Let A denote the amount of the first bid. If the individual answered “no” to A , then the second bid will be some lower bound $A_d < A$; while if the response is “yes” to A , then the second bid will be some higher bound $A_u > A$. Thus the 4 possible responses would be: “yes, yes”, “yes, no”, “no, yes”, “no, no”. And the probability of response is given by

$$Prob(\text{yes, yes}) = 1 - G_c(A_u) \quad (23)$$

$$Prob(\text{no, no}) = G_c(A_d) \quad (24)$$

$$Prob(\text{yes, no}) = G_c(A_u) - G_c(A) \quad (25)$$

$$Prob(\text{no, yes}) = G_c(A) - G_c(A_d) \quad (26)$$

$G_c(A)$ can be specified according to the functional and distributional assumption. And the model can be estimated by MLE. Hanemann (1991) showed that comparing to the traditional single bounded model, this double bounded model provides more precise variance-covariance matrix of the coefficient estimates, and a generally lower median WTP estimate. The intuition for this improvement is that when the first bid is poorly designed, the second bid can always get closer to the optimal design.

1.3.4 Travel Cost Model

Another type of non-market valuation model we want to cover is travel cost models. The basic premise of the travel cost method is that the time and travel cost expenses that people incur to visit a site represent the “price” of access to the site. Thus, peoples’ willingness to pay to visit the site can be estimated based on the number of trips that they make at different travel costs. We

used Poisson and negative binomial specifications as representative for the travel cost model in later discussions.

1.4 Organization

This dissertation consists of three chapters. The first chapter introduces the background, motivation, and organization of the dissertation. The second chapter discusses nonlinear transformation bias and parameter estimation bias, and compares performance of bias correction methods including the delta method, jackknife and bootstrap. The third chapter deals with the sampling bias, its bias decomposition, and uses Monte Carlo to show how kernel mean matching can be used to correct the sampling bias under different bias compositions.

CHAPTER 2

COMPARISON OF BIAS CORRECTION METHODS: MONTE CARLO SIMULATION

2.1 Introduction

Several bias correction methods have been proposed for estimating welfare measures including the delta method, jackknife, bootstrap and distributional-free nonparametric method. In practice the delta method is the most commonly used bias correction method, but the reliability of these methods has not been validated and compared. The delta method can be difficult to apply in that the statistical differentials (for univariate), or variance-covariance matrix (for multivariate) are not readily available. And for many large and complicated models of the type encountered in non-market valuation, the bootstrap is not a feasible alternative. Thus it should be in the practitioners' interest to see if the jackknife can be a reliable alternative to the delta method.

Most of the statistical literature in this field compares the asymptotic properties of bias correction estimators for simple statistics like the sample mean or correlation coefficient (Efron, 1977, 1981, 1983; Parr, 1983; Chow, 1985; Withers, 2013, 2014), which cannot be directly applied to models used in non-market valuation, especially for small sample estimation. Only a few studies compared these methods for welfare measures in non-market valuation (Cooper, 1994; Langford, 1998; Bliemer, 2013). These studies used Monte Carlo simulation to show the small sample performance of bias correction methods, but they only used limited models with little variation of circumstances, which can hardly represent versatile cases in practice with different model specifications, sample sizes, and possible distributional misspecification.

Besides, in these studies the Monte Carlo simulation results have not been linked with the asymptotic properties for these bias correction methods, and no mathematic explanations for the simulation results was proposed, although there are extensive theoretical literatures about resampling methods.

This dissertation compares the performance of these bias correction methods for welfare measures in non-market valuation by asymptotic expansion as well as Monte Carlo simulation. We address this research question by achieving the objectives below:

- Why is there bias? Identify the possible driving factors of bias during and after estimation. Understand the interaction of these factors, and measure their impacts at different sample sizes by Monte Carlo simulations.
- Which is the optimal bias correction method for large sample estimation? Derive asymptotic expansions of these bias corrected estimators. The results show the performance of these bias correction methods when sample size is sufficiently large.
- Which is the optimal bias correction method for small sample estimation?
Compare the performance of these bias correction methods using Monte Carlo simulation with variations of sample size, model specification, and distributional assumption. The findings provide a reference for practitioners when trying to select the optimal bias correction method.
- Why does a bias correction method perform better under certain circumstances?
Study the inherent mechanisms of bias correction procedures, their driving forces and the correlations between them; provide theoretical explanations for Monte Carlo simulation results.

- Can the jackknife, or any other method work as a reliable and efficient alternative for the delta method? Explore, recombine and modify the existing bias correction procedures to discover possible alternatives.
- Can parameter estimate bias correction be combined with the delta method to improve the accuracy of the WTP estimates? If so, how to implement it?

2.2 Bias of Welfare Measures in Non-Market Valuation

Several types of bias can be incurred in the process to obtain the welfare measures in non-market valuation. Firstly, parameter estimation bias, namely the estimation results of underlying parameters are biased due to the insufficient sample size. This type of bias can be alleviated by applying the jackknife method. In non-market valuation practice this source of bias is entangled with other sources of bias, which complicates the analysis of the performance of bias correction methods. That is a reason why we tend to rely on simulation results instead of an analytical approach when comparing bias correction methods.

Secondly, non-linear transformation bias, namely the welfare measures obtained by unbiased parameter estimation can still be biased because the welfare measure is usually calculated as a non-linear function of the estimated parameters, and it is well-known that the expected value of a non-linear function of unbiased parameter estimates is itself generally biased. Welfare measures in non-market valuation are usually derived from nonlinear function of parameters. If θ is a vector of k parameters, under maximum likelihood estimation

$$\hat{\theta} \sim N(\theta, V(\theta)) \quad (27)$$

where $\sim$ means approximately distributed as, $V(\theta)$ is the variance-covariance matrix. Assume welfare measure $W = h(\theta)$, for which the usually estimator is $\hat{W} = h(\hat{\theta})$. Here $\hat{\theta}$ usually takes

the form of sample mean $\bar{\theta}$. When h is a nonlinear function, the function of parameter estimates $h(\hat{\theta})$ will be a biased estimate for WTP. If h is a convex function, then $E(h(\hat{\theta})) > E(h(\bar{\theta}))$; if h is a concave function, then $E(h(\hat{\theta})) < E(h(\bar{\theta}))$. With sample size increasing, the expected bias will decrease, but will always have the same sign.

Thirdly, distributional misspecification bias, namely the data generating process assumed by the statistical model can be different from the economic model specified. When applying random utility models, we derive the model specification by assuming a distribution of the random part of utility or corresponding probability distribution. For instance, the logit model with logistic probability distribution is obtained by assuming that each random part of utility is independently, identically distributed extreme value, namely Gumbel or type I extreme value distribution; while a probit model is used with the normal distribution specified.

Common probability distributions are the normal, logistic, log-logistic, Weibull and exponential (Kerr, 2000). With real world datasets the difficulty is that the underlying true distribution cannot be revealed even we compare results with different possible distributional specifications. However, simulated data can be generated by setting up a certain distribution, and by comparing results under “true” and alternative models for the influence of distributional misspecification. The effects of skewness or kurtosis from the “true” distribution can be driving the bias in different ways.

We will illustrate these sources of bias with frequently used models in non-market valuation practice as examples, including Poisson and negative binomial for travel cost models and binary probit and logit for random utility models. We will also consider more complicated models including double bound models, random effect models, and scaled or drifted distributional

specifications as suggested by Kerr (2000). We will show in simulations how large and significant each type of bias can be under different sample sizes.

These different sources of bias may become the major driving factors for different models at different sample sizes. Basically the parameter estimation bias is prominent when sample size is small, and it can also complement the nonlinear transformation bias at first order. As shown in (33), when the sample size is large enough to neglect the parameter estimation bias, the nonlinear transformation bias becomes the major source of bias. The asymptotic result can be seen as the case when sample size approaches infinity, and the consistent parameter estimation has zero bias. Although in practice generally we care more about small sample results than asymptotic results, the asymptotic result can help explain the dynamics of these driving factors.

2.3 Bias Correction Methods

In CV models the welfare measures are often obtained by nonlinear transformation of estimated parameters, which are known to be biased due to nonlinearity. Also, as summarized by Chen (2012), the maximum likelihood parameter estimates can be biased in finite samples. A problem of interest is to reduce the bias as well as MSE of such estimators. The most common method to do so is to apply Taylor's series expansion to analytically derive bias at the $O(n^{-1})$ corrected estimator, namely the Delta method. Oehlert (1992) gave a review of the delta method and concluded that the Taylor series approximation for expected value of the function is consistent if the underlying sequences of random variables have enough bound moments.

Alternatively, resampling procedures like the jackknife and bootstrap can also remove the $O(n^{-1})$ bias, which entails a negligible increase in variance as shown by Quenouille (1956), or even decrease the variance as shown by Durbin (1959). Here we mainly focus on the jackknife,

as for many large and complicated models of the type encountered in non-market valuation, the bootstrap is not a feasible alternative; besides, Efron (1979) showed that the jackknife is a linear approximation for the bootstrap. The “delete-one” jackknife was extended by Jaeckel (1972) to the infinitesimal jackknife, and by Gray (1975) and Sharot (1976) to the generalized jackknife. Gray (1975) and Efron (1979) pointed out that the bias and variance expressions suggested by the infinitesimal jackknife transform equivalently to the Delta method, and the Delta method is a limiting case of the generalized jackknife. Kim (1998) derived conditions under which bias corrected estimates for smooth function and non-smooth function statistics with the jackknife and bootstrap method lead to increase in MSE in terms of second-order asymptotic superiority, which can be extended to regression models. Withers (2013) assessed through a simulation study the performance of the Delta method bias corrected estimators for the odds ratio in a Binomial (n, θ) , including the maximum likelihood estimator, using the delta estimators of order up to three. Bias and MSE of the four estimators are computed in 10,000 simulations. The steps are repeated for sample size $1 \sim 100$, and $\theta = 0.3, 0.5, 0.7, 0.9$. Results showed that the delta method can reduce both bias and MSE in multinomial odds ratio estimation, and the results are consistent under different values of θ .

In section 2.3 we list several relevant bias correction methods, and in section 2.4 we will compare the performance of some of these methods by Monte Carlo simulation.

2.3.1 Analytical Bias Correction Method for Maximum Likelihood Parameter Estimation

The maximum likelihood parameter estimation bias can be corrected by computing the bias of parameter estimates at first or higher orders, and then subtracting the bias from the original estimate. Here we give several examples of correcting the bias of the ML estimator $\hat{\beta}$.

For a logit model, assume $Pr(y = 1) = \pi$, with which a set of covariates $\mathbf{X}$ is associated.

And the log odds ratio is a linear function of $\mathbf{X}$, specified as $\mathbf{X}'\boldsymbol{\beta}$. To estimate $\hat{\boldsymbol{\beta}}$, we need to solve the equation

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_r} \stackrel{\text{def}}{=} g(\beta_r) = \sum (y_i - \hat{\pi}_i) \mathbf{X}_{ir} = 0 \quad (28)$$

where $l(\boldsymbol{\beta})$ is the log-likelihood function, and $g(\beta_r)$ is the score function. In matrix notation

$\boldsymbol{\pi} = (1 + e^{-\mathbf{X}'\boldsymbol{\beta}})^{-1}$, and $g(\boldsymbol{\beta}) = \mathbf{X}'(\mathbf{y} - \hat{\boldsymbol{\pi}})$. The fisher information is $\mathbf{F}(\boldsymbol{\beta}) = \mathbf{X}'\mathbf{W}\mathbf{X}$, where

$\mathbf{W} = \text{diag}(\hat{\boldsymbol{\pi}}(1 - \hat{\boldsymbol{\pi}}))$. Then $\hat{\boldsymbol{\beta}}$ is obtained by applying the iterative Newton-Raphson method:

$$\boldsymbol{\beta}^{t+1} = \boldsymbol{\beta}^t + \mathbf{F}(\boldsymbol{\beta}^t)^{-1} g(\boldsymbol{\beta}) \quad (29)$$

where $\boldsymbol{\beta}^t$ is the estimate at t-th stage. This step is repeated until convergence criterion is reached.

This method is called iterative maximum likelihood estimation.

Firth (1993) suggested that the bias of MLE can be reduced by modifying the score function

(41) as

$$g_f(\beta_r) = \sum ((y_i - \hat{\pi}_i) + h_i(0.5 - \hat{\pi}_i)) \mathbf{X}_{ir} \quad (30)$$

where h_i is the i-th diagonal element of the matrix $\mathbf{H} = \mathbf{W}^{\frac{1}{2}} \mathbf{X} (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}^{\frac{1}{2}}$. Then use $g_f(\beta_r)$

for the iterative Newton-Raphson method.

Cordeiro (1991) proposed a bias correction estimator for the generalize linear model

$$\boldsymbol{\beta} = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\boldsymbol{\xi} \quad (31)$$

For the logit model $\boldsymbol{\xi} = \text{diag}[\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'](\hat{\boldsymbol{\pi}} - 0.5)$.

2.3.2 The Delta Method

The Delta method can correct the nonlinear transformation bias by approximating the nonlinear function with a Taylor series. When the function h has continuous second derivatives, by taking a second order Taylor series approximation of $h(\hat{\boldsymbol{\theta}})$, we get

$$h(\hat{\boldsymbol{\theta}}) \cong h(\boldsymbol{\theta}) + \sum_{i=1}^k E[(\hat{\theta}_i - \theta_i)] \frac{\partial h(\boldsymbol{\theta})}{\partial \theta_i} + \frac{1}{2} \sum_{i,j=1}^k V_{ij}(\hat{\boldsymbol{\theta}}) \frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \quad (32)$$

Taking expectations of both sides, we get a bias corrected estimator

$$\hat{W}_D \cong E[h(\hat{\boldsymbol{\theta}})] - \sum_{i=1}^k E[(\hat{\theta}_i - \theta_i)] \frac{\partial h(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_i} - \frac{1}{2} \sum_{i,j=1}^k V_{ij}(\hat{\boldsymbol{\theta}}) \frac{\partial^2 h(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_i \partial \hat{\theta}_j} \quad (33)$$

This first-order approximation is usually referred to as the delta method. Although equation (33) can be evaluated at every data point, typically interest is focused on the sample mean $h(\bar{\boldsymbol{\theta}})$. The delta method can eliminate the $1/n$ bias leading term from the nonlinear function (Miller, 1964).

If $\hat{\theta}_i$ is treated as an unbiased estimator of θ_i , then the first order term in (33) can be eliminated, then we get the most commonly used version of delta method estimator.

$$\hat{W}_D \cong E[h(\hat{\boldsymbol{\theta}})] - \frac{1}{2} \sum_{i,j=1}^k V_{ij}(\hat{\boldsymbol{\theta}}) \frac{\partial^2 h(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_i \partial \hat{\theta}_j} \quad (34)$$

The method mentioned in 2.3.1 can be incorporated into the delta method to account for the first order term in (33). Theoretically this procedure produces an estimator with less expected bias.

For small samples, the first order bias is prominent, which weakens the performance of the delta method if this term is omitted. For certain models, the bias of $\hat{\theta}_i$ has been derived, so that a more precise result can be attained. We will talk about this small sample bias in more detail at 2.9.

2.3.3 Jackknife

The basic idea behind the jackknife bias corrected estimators lies in recomputing the statistic of interest using all but one of the observations. Suppose $\hat{\theta}$ is estimated from a dataset

$X_1, X_2, X_3, \dots, X_n$. And welfare measure is $\hat{W} = h(\hat{\theta})$, then denote

$$\hat{W}_{-i} = h(\hat{\theta}_{-i}) \quad (35)$$

Here $\hat{\theta}_{-i}$ is estimated with all but i_{th} observation X_i for $i = 1, 2, \dots, n$. Denote the i_{th} pseudo-value by $\hat{W}_i = n\hat{W} - (n-1)\hat{W}_{-i}$. Then the jackknife bias corrected estimator is

$$\hat{W}_J = \frac{1}{n} \sum_{i=1}^n \hat{W}_i = n h(\hat{\theta}) - \frac{n-1}{n} \sum_{i=1}^n h(\hat{\theta}_{-i}) \quad (36)$$

The Jackknife can also be defined as a form of weighted observations. If we attach a weight to each observation, then leaving one observation out is the same as giving an observation a weight of zero. Instead of doing that for the ordinary jackknife, we give the observation only a slightly less weight than the others, then we get the infinitesimal jackknife. We assign weights $w_1, w_2, w_3, \dots, w_n$ to observations $X_1, X_2, X_3, \dots, X_n$ respectively. Then $\hat{W}$ can be written as a function of $2 \times n$ variables $T(X_1, X_2, X_3, \dots, X_n; w_1, w_2, w_3, \dots, w_n)$, since all observations have equal weight of $1/n$, we get

$$\hat{W} = T(X_1, X_2, X_3, \dots, X_n; \frac{1}{n}, \dots, \frac{1}{n}) \quad (37)$$

For $\hat{W}_{-i}$, the weight of X_i becomes 0, and all other observations have same weight, so we have

$$\hat{W}_{-i} = T(X_1, X_2, X_3, \dots, X_n; \frac{1}{n-1}, \dots, \frac{1}{n-1}, 0, \frac{1}{n-1}, \dots, \frac{1}{n-1}) \quad (38)$$

To simplify some of the algebra, we extend the definition of T as follows. If G is any probability distribution for which T is defined, and c is a positive constant, then we let $T(cG) = T(G)$. Thus,

if we are considering discrete distributions with possible values X_i , $i=1, \dots, n$. we can assign to each X_i the weight w_i without requiring that $\sum_{i=1}^n w_i = 1$.

If we reduce w_i by ϵ and let $\epsilon \rightarrow 0$. We have

$$\widehat{W}_{-i}(\epsilon) = T(X_1, X_2, X_3, \dots, X_n; \frac{1}{n}, \dots, \frac{1}{n}, \frac{1}{n} - \epsilon, \frac{1}{n}, \dots, \frac{1}{n}) \quad (39)$$

Let $\widehat{D}_i = \frac{\partial T}{\partial w_i}$, $\widehat{D}_{ii} = \frac{\partial^2 T}{\partial w_i^2}$, since $T(cG) = T(G)$, it can be proved that $\frac{1}{n} \sum \widehat{D}_i = 0$. We can form the

Taylor series expansion

$$\begin{aligned} \widehat{W}_{-i}(\epsilon) - \widehat{W} &= T\left(\frac{1}{n}, \dots, \frac{1}{n} - \epsilon, \dots, \frac{1}{n}\right) - T\left(\frac{1}{n}, \dots, \frac{1}{n}\right) \\ &= -\epsilon \widehat{D}_i + \frac{\epsilon^2}{2} \widehat{D}_{ii} - \dots \end{aligned} \quad (40)$$

Suppose $\widehat{W}_{(\cdot)}(\epsilon) = \frac{1}{n} \sum \widehat{W}_{-i}(\epsilon)$, we can define a bias estimate $\widehat{B}(\epsilon)$ by

$$n^2 \epsilon^2 \widehat{B}(\epsilon) = n(1 - \epsilon)(\widehat{W}_{(\cdot)}(\epsilon) - \widehat{W}) \quad (271)$$

Letting $\epsilon \rightarrow 0$, we have

$$\widehat{B}(0) = \frac{1}{2n^2} \sum \widehat{D}_{ii} \quad (42)$$

Finally, we get the infinitesimal jackknife estimate

$$\widehat{W}_{IJ} = \widehat{W} - \widehat{B}(0) = \widehat{W} - \frac{1}{2n^2} \sum \widehat{D}_{ii} \quad (43)$$

which is exactly the result of applying delta method to $\widehat{W}$ (Gray, 1975).

The usage of the jackknife can be extended to many aspects of estimation. Kézdi (2002) proposed a jackknife minimum distance estimator designed to reduce the finite-sample bias of the optimal minimum distance estimator. Lee (2012) showed that the jackknife maximum likelihood estimator is consistent, and the jackknife estimate of the log likelihood is asymptotically unbiased.

2.3.4 Bootstrap

An alternative resampling method, the bootstrap procedure can also be used for bias correction in welfare measures with nonlinear transformation. The procedures are as follows:

Step 1. Construct the empirical probability distribution $\hat{F}$ of each dataset with n observations by putting probability mass n^{-1} on each observation X_i .

Step 2. Draw a random sample of size n from $\hat{F}$ as $X_1^*, X_2^*, X_3^*, \dots, X_n^* \sim \hat{F}$, and get $\hat{\theta}^* = f(\bar{X}^*)$, where $\bar{X}^* = \frac{1}{n} \sum_{i=1}^n X_i^*$.

Step 3. Repeat step 2 of a large number T times, get $\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, \dots, \hat{\theta}_T^*$. With $\hat{W} = h(\hat{\theta})$, get $\hat{W}_1^*, \hat{W}_2^*, \hat{W}_3^*, \dots, \hat{W}_T^*$. Then we get the estimate of bias

$$\widehat{bias}(\hat{W}) = \frac{1}{T} \sum_{t=1}^T (\hat{W}_t^* - \hat{W}) \quad (44)$$

So the bootstrap bias corrected estimator is

$$\hat{W}_B = 2\hat{W} - \frac{1}{T} \sum_{t=1}^T \hat{W}_t^* \quad (45)$$

The Bootstrap can take different forms by different schemes of random drawing, here for simplicity we only consider the most commonly used method.

For many large and complicated models of the type encountered in non-market valuation, the bootstrap is not a feasible alternative due to the need to re-estimate the model T times. Thus we only use bootstrap simulation results as a reference for simple model cases, but do not make further extensions of it.

2.3.5 Distribution-Free Nonparametric Method

In parametric estimation the distribution of random utility can be incorrectly specified, which could cause inefficient and/or biased estimates. Kerr (2000) applied multiple distributions with

shifted and scaled versions to two existing dichotomous choice data sets, and found that competing models generate widely divergent estimates of mean consumer surplus. Kriström (1990) first adopted nonparametric estimation of welfare measures in discrete response models by Distribution Free Maximum Likelihood (DFML). Creel and Loomis (1997) and Chen and Randall (1997) proposed a semi-nonparametric estimator for welfare measures for binary discrete response models without imposing distributional or function form assumptions, which is a replacement for rigid models when they do not pass the specification test. Watanabe (2009) also proposed a distribution free consistent linear projection estimation of mean WTP. Giraud (2001) compared a number of parametric, semi-nonparametric and nonparametric estimation techniques using empirical data, and showed there are significant differences among estimation techniques with a hypothesis test based on a bootstrap method.

When the true distribution is hard to identify, choose the “moderate” one so as to not cause severe bias. Creel (1998) stated empirical studies have found that estimates of mean WTP derived from simple parametric models often differ little from those derived from nonparametric methods. This indicates that simple models can potentially yield WTP estimators that are relatively unbiased.

The typical discrete choice survey gives a sequence of proportions $\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2, \hat{\pi}_3, \dots, \hat{\pi}_k)$, and an increasing sequence of bids $\hat{A} = (A_1, A_2, A_3, \dots, A_k)$, where $\hat{\pi}_i$ is the proportion of responses to the lowest bid A_i ; and $\pi_i = k_i/n_i$, where k_i and n_i are respectively the number of “yes” responses and number of total responses for bid A_i . If $\hat{\pi}$ forms a monotone increasing sequence, then this sequence provides a Distribution Free Maximum Likelihood estimator of the probability of acceptance at each bid point as $\bar{\pi}$ (Ayer, 1955). If the sequence is not monotonic, it can be made monotonic by applying a pool-adjacent-violator algorithm: if $\hat{\pi}_i < \hat{\pi}_{i+1}$ for some i ,

then $\bar{\pi}_i = \bar{\pi}_{i+1} = (k_i + k_{i+1})/(n_i + n_{i+1})$. And the procedure is repeated until the sequence is monotonic in i . By applying this algorithm, we get a consistent estimation of the true probability of acceptance. By certain interpolation of the points of probabilities of acceptance at each bid amount, we obtain the empirical survival function, and the mean WTP can be approximated by the area under the curve.

2.4 Comparison of Bias Correction Methods

2.4.1 Literature Review

Efron (1981) used bootstrap, jackknife and infinitesimal jackknife and twelve other nonparametric methods in a Monte Carlo experiment for calculating the correlation coefficient. The point estimates, variances, and MSEs are compared. The simulation results showed that jackknife has less bias than the delta method; the bootstrap generally performs the best among all methods at the cost of less computational efficiency.

Parr (1983) showed that the Delta method, jackknife and bootstrap coincide at least to the first order term, and the absence of skewness leads to second-order coincidence of the jackknife and delta methods.

Shein-Chunq Chow (1985) studied the Delta method, jackknife and bootstrap estimators for nonlinear functions of local parameters and derived asymptotic expansions of the mean squared error up to the third order. There is no consistent relationship among these estimators, the performance depends on the underlying distribution and particular form of the nonlinear function. A least MSE estimator was proposed and compared with these three estimators, in simulations for several commonly used nonlinear functions, $\log x$, $1/x$, and $\sqrt{x}$ of sample mean estimates, and distributions including gamma, Poisson, and lognormal, with sample size $n=10$

and $n=20$. A total of one thousand runs were made for each combination. This study provides a direct basis for our study, but the sample size is too small for any non-market valuation practice, and only a nonlinear function of a single parameter estimate was studied. So we will extend to nonlinear functions of more than one estimated parameter, and increase the simulation sample size range, as well as functional forms, according to the importance and usage frequency in non-market valuation practice.

The comparison of these method in the field of non-market valuation was firstly studied by Cooper (1994). A Monte Carlo simulation technique is used to assess the reliability of the Krinsky and Robb, jackknife, bootstrap, and the Cameron confidence interval routines as applied to the logit regression model with simulated sample size $N=100$, $N=500$ and $N=1000$. The simulated data sets are created using the logistic distributions, thus the robustness of these methods under misspecification of skewed distribution can be tested. Bias, standard deviation, and frequency of inclusion of true mean in the estimated confidence interval were used as comparison criterions. All methods produced similar results with larger samples, but for $N=1000$ the logistic distribution bootstrap performed the best. An analytical approach like the Cameron method performs the best when $N=100$, but tends to underestimate the standard deviation. Under falsified Weibull distribution, all methods perform worse than when the logistic is the true underlying distribution, but the jackknife performs the best when $N=500$ and $N=1000$.

Langford (1998) compared bootstrap and Delta method bias corrected estimations of median WTP in the Howe Logit model, and showed that with bid level effect included (random effect model), the bootstrap result is more skewed and with more kurtosis than the Delta method results, and the variances of two results are also significantly different.

Withers (2014) compared the computational efficiency of bias correction estimators using the delta method, jackknife and bootstrap. It is shown that the p th order delta estimator requires $\sim n$ calculations, the p th order “drop-one-observation” jackknife estimator requires $\sim n$ calculations if $p = 1$ and $\sim np-1$ calculations if $p \geq 2$, the p th order “drop-m-observation” jackknife estimator requires $\sim nms+1$ calculations or $\sim nms$ calculations if $s = p - 1 > 0$; and bootstrap is second order and requires n^2 calculations.

2.4.2 Monte Carlo Simulation

Monte Carlo Simulation analysis is started by assuming a distribution, which can be the true underlying distribution of the model to be estimated, or a different distribution. The reason for the distributional divergence is to assess the bias under distributional misspecification. Then a large number S simulated datasets are created based on the assumed distribution. With each simulated dataset, $\widehat{W}_D$, $\widehat{W}_J$ and $\widehat{W}_B$ are calculated. With these S estimates based on the simulated distribution of $\widehat{W}$, expected bias, variance, mean square error, and confidence intervals can be calculated and compared for $\widehat{W}_D$, $\widehat{W}_J$ and $\widehat{W}_B$.

Simulations need to run under a wide range of variations to represent the possible alternative cases. We believe the key variations include: model specification/functional form; sample size; distributional assumption.

Firstly, different bias correction methods may have optimal performance with different model specifications. For instance, the welfare measures may be nonlinear functions of one parameter, multiple parameters, or estimated parameters with observed values. We will start with commonly used models like logit, probit, and Poisson; and extend to more advanced ones like double bound probit and the Box-Cox model.

Secondly, different bias correction methods may perform best under different sample size. For instance, as sample size in each repetition increases, the variance of estimates in each repetition gets smaller, and the nonlinear transformation becomes closer to linear when it happens within a narrower interval, so that the nonlinear transformation bias becomes smaller when sample size increases. On the other hand, the delta method requires sample size to be above certain thresholds to produce robust result. However, for the jackknife the performance has no such correlation to sample size. Simulation will be done under sample sizes of 50,100,200 and 400.

Thirdly, under distributional misspecification the welfare measure will be further biased, for which different bias correction methods may show different levels of robustness. The misspecification may be about skewness as in Weibull /logistic, or kurtosis as in normal/logistic. Besides, several nonparametric/semi-nonparametric methods can be alternative solutions for distributional misspecification, although they are in general less efficient than MLE.

2.4.3 Asymptotic Expansion

It is well known that all three methods discussed above can remove the $1/n$ bias term (Miller 1964, Efron 1982). Chow (1985) made an asymptotic expansion of bias correction estimates using delta method, jackknife and bootstrap up to $1/n^2$. Let $X_1, X_2, X_3, \dots, X_n$ be i.i.d. random variables with mean μ and variance σ^2 , and let f be a nonlinear function with continuous k -th derivative. Suppose that $\theta = f(\mu)$ is the target statistic, and $\hat{\theta}_D, \hat{\theta}_J$ and $\hat{\theta}_B$ be the bias corrected estimates using delta method, jackknife and bootstrap. Then we have

$$bias(\hat{\theta}_D) = -n^{-2}(2a + b) + O(n^{-2}) \quad (46a)$$

$$bias(\hat{\theta}_J) = -n^{-2}(a + b) + O(n^{-2}) \quad (46b)$$

$$bias(\hat{\theta}_B) = -n^{-2}(3a + 2b - c) + O(n^{-2}) \quad (46c)$$

where $a = \frac{1}{6}\mu_3 f^{(3)}(\mu)$, $b = \frac{1}{8}\sigma^4 f^{(4)}(\mu)$, $c = \frac{1}{2}\sigma^2 f^{(2)}(\mu)$. μ_3 is the third central moment and $f^{(k)}(\mu)$ is the k_{th} derivative at μ . Based on these expansions, we can see that if $\mu_3 = 0$, (i.e., the distribution is symmetric about μ) then $\hat{\theta}_D$ and $\hat{\theta}_J$ both reduced bias equally up to the n^{-2} term.

This result can be extended to the asymptotic expansion of the welfare measures in non-market valuation as nonlinear transformations of multiple ML estimated parameters, which by the asymptotic properties of MLE have asymptotic normal distribution. Thus for these models the delta method estimate and jackknife estimate should be asymptotically equivalent up to the n^{-2} term.

The asymptotic result shows the performance of bias correction methods when sample size approaches infinity, and the consistent parameter estimator has zero bias. Asymptotic results can be used as an upper bound baseline for the small sample case, although in practice the sample size is barely large enough to directly apply them.

2.5 Poisson Model

The travel cost model of recreational demand can take the form of a Poisson specification

$$\lambda = e^{\beta_0 + \beta_1 x} \quad (28)$$

with λ representing the number of trips to a recreation site and the variable x representing travel costs. If we assume $\beta_0 = 2$, $\beta_1 = 0.25$, and x being a positive valued uniform variable with mean 4 and minimum 1.5. Then $-1/\beta_1$ is the per-trip consumer surplus, the true value of which under the population model should be 4.

2.5.1 Monte Carlo Simulation

The model is simulated 1,000 times for each estimation. Then the 1000 outcomes are treated as a new sample with 1,000 observations, with which the mean, mean of bias, standard deviation, and MSE are calculated. For standard deviation, we use the one calculated over 1,000 simulations, instead of the mean of the 1,000 standard deviation calculated in each simulation; because we are interested in the precision and reliability of the bias estimates, instead of the precise estimates of standard deviation.

Using (34) the general formula of the delta method, we derive the delta method estimator of the specified Poisson model as

$$\widehat{W}_D \cong -\frac{1}{\beta_1} + \frac{V(\beta_1)}{\beta_1^3} \quad (48)$$

For the bootstrap in each simulation we use a sample with the same size as the original one, so that it can remain at a similar computational efficiency level as other method. The estimates without any bias correction procedure are summarized in “original” row.

Here the point estimate results (with true value 4) is listed in Table 2.1, and standard deviation and MSE results are listed in Table 2.2.

Table 2.1: Monte Carlo analysis of WTP expected bias for Poisson model

Sample size	50	70	100	200	400	1000
Original	4.224	4.185	4.137	4.080	4.048	4.013
Delta	3.840	3.964	3.994	4.016	4.017	4.001
Jackknife	3.830	3.967	3.999	4.019	4.019	4.002
Bootstrap	4.270	4.185	4.083	4.051	4.015	4.015

Table 2.2: Monte Carlo analysis of WTP mean square error for Poisson model

Sample size	50	70	100	200	400	1000
STD.DEV						
Original	1.355	0.938	0.781	0.512	0.354	0.215
Delta	0.858	0.762	0.692	0.488	0.346	0.214
Jackknife	0.870	0.764	0.693	0.487	0.346	0.214
Bootstrap	1.316	1.080	0.722	0.513	0.337	0.212
MSE						
Original	1.8860	0.9142	0.6284	0.2688	0.1272	0.04651
Delta	0.7613	0.5820	0.4795	0.2381	0.1198	0.04558
Jackknife	0.7859	0.5842	0.4800	0.2380	0.1199	0.04560
Bootstrap	1.8048	1.1999	0.5283	0.2654	0.1139	0.04508

The “original” bias declines when sample size increases. The nonlinear transformation always caused the estimates to be biased upwards because of the convexity of the function $-\frac{1}{\beta_1}$ with β_1 being negative.

The delta method and jackknife both decrease bias as well as standard deviation with sample size level 50~1000, and they have very close results at all sample size levels. A converging trend can be observed for the bias to shift from downwards to upwards when sample size increases over 100. The jackknife has smaller bias roughly with sample size from 70~100; while the delta method has smaller bias roughly with sample size above 200. The reason for the superiority of delta method at larger sample size may be that for a small sample size the major source of bias is the bias of $\widehat{\beta}_1$, and when sample size is large enough, the bias of $\widehat{\beta}_1$ become negligible compared to the second order expansion of $\widehat{W}_D$ as calculated in (48). When the standard deviations from these two methods are compared, consistent results cannot be reached at different sample sizes; similarly with the MSE. Thus, by the standard of MSE, the delta method and the jackknife can be seen similar to each other for the Poisson model when sample size is below 1000.

The reason for the delta method and jackknife to have similar performance may be that the delta method is equivalent to infinitesimal jackknife, which is a smoothed version of the ordinary leave-one-out jackknife. The gap between the delta method and the jackknife can be seen more clearly when sample size is large enough. Despite this gap, when the precision requirement is not so high, or when the delta method is not convenient to apply due to a complicated second derivative, the jackknife can be an eligible alternative for the delta method for the Poisson model.

It is noticeable that the original estimation result is biased upwards under all sample sizes; while both the delta and jackknife result are biased downward when sample size is small, then

both convert to be biased upward at sample size 200, and as sample size continues to increase they remain biased upwards but in smaller absolute value. Again, the different combinations of bias sources and their interactions under different sample size can help explain this, but this analysis is beyond the scope of this study.

The bias of the bootstrap result is larger than the original result at sample size 50,70, and 1000. And when sample size is sufficiently large, the bootstrap estimate has the smallest MSE, but still a larger bias. One advantage of the bootstrap is that even for small sample estimation, we can always improve the precision of the bootstrap by increasing random draws in each simulation. It is noticeable that the sign of bias is always upward and keeps decreasing as the sample size increases.

2.5.2 Asymptotic Expansion of Bias Corrected Estimates

Per (46a) (46b) (46c), for Poisson model, the welfare measure estimates after bias correction using delta method, jackknife and bootstrap are

$$bias(\widehat{W}_D) = -n^{-2} \left(\frac{2\mu_3}{\hat{\beta}^4} - \frac{3\sigma^4}{\hat{\beta}^4} \right) + O(n^{-2}) \quad (49a)$$

$$bias(\widehat{W}_J) = -n^{-2} \left(\frac{\mu_3}{\hat{\beta}^4} - \frac{3\sigma^4}{\hat{\beta}^4} \right) + O(n^{-2}) \quad (49b)$$

$$bias(\widehat{W}_B) = -n^{-2} \left(\frac{3\mu_3}{\hat{\beta}^4} - \frac{6\sigma^4}{\hat{\beta}^4} + \frac{\sigma^2}{\hat{\beta}^3} \right) + O(n^{-2}) \quad (49c)$$

where μ_3 is the third central moment of $\hat{\beta}$, and σ^2 is the variance of estimate $\hat{\beta}$. Based on these expansions, suppose $\mu_3 = 0$, then $\widehat{W}_D$ and $\widehat{W}_J$ reduced bias equally up to the n^{-2} term when the sample size is large enough and $\hat{\beta}$ can be regarded as unbiased. If $\mu_3 < 0$, then $|bias(\widehat{W}_D)| > |bias(\widehat{W}_J)|$; if $\mu_3 > 0$, there is no conclusion about the size of $bias(\widehat{W}_D)$ and $bias(\widehat{W}_J)$. Thus,

for the Poisson model of per-trip consumer surplus, the jackknife estimate has a smaller asymptotic bias than the delta method estimate when $\hat{\beta}$ is biased upwards, and is equivalent to the delta method estimate when $\hat{\beta}$ is unbiased.

The asymptotic expansion can provide some intuition for the performance differences of the simulation result, especially for a simple model like Poisson and its single parameter nonlinear transformation. For more complex models with WTP estimates derived as nonlinear functions of multiple estimated parameters, the asymptotic results become difficult to be obtained or analyzed.

2.6 Logit Model

A model is assumed with a binary choice variable y representing whether to accept the bid, with x representing conditioning variables, and A representing the bid amount offered. By assuming latent unobserved WTP as both logistic and Weibull distributions, we can get the true data generating process specified as

$$Prob(y = 1) = \frac{e^{\beta_0 + \beta_1 x + \beta_2 A}}{1 + e^{\beta_0 + \beta_1 x + \beta_2 A}} \quad (50)$$

If we assume logistic distribution as in (50), the expected WTP is

$$E(WTP|x) = \frac{\beta_0 + \beta_1 \bar{x}}{-\beta_2} \quad (51)$$

For Monte Carlo, we generate simulated data using (50), and fit the data using the logit model, so the estimates for the Weibull dataset can be assumed as a distributional misspecification case.

We assume $\beta_0 = 1.4136$, $\beta_1 = 0.00372$, and $\beta_2 = -0.008561$. Amenity variable x is generated with random draws from a normal distribution with mean 95.78 and standard deviation 119.226.

A is generated by a repeating sequence (2.5, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 150,

200, 250, 300, 400, 500, 700). The dependent variable y is generated by estimating $P_i = \widehat{Prob}(y_i = 1)$ for each observation with (50), and using a Bernoulli distribution with parameter P_i to generate a “1 or 0” response. Following these procedures, for logistic, the true $E(WTP)$ is \$206.74. Because x is randomly generated, for every simulation we plug the $\bar{x}$ calculated from the generated x into (51) to get the distribution of true WTP for each sample size.

For simulated datasets, we choose sample sizes of 40, 80, 100, 200, 400, 1000 to ensure equal numbers of each bid amount are included for each sample size. For each sample size the WTP estimate is derived with direct nonlinear parameter transformation, delta method, and jackknifing. For each analysis 1,000 simulations are conducted to obtain 1,000 point estimates of WTP forming up the simulated distribution, of which the mean, expected bias, standard deviation, and MSE are reported.

For each sample size, there is a separate true mean because although the $\beta_i, i = 1, 2, 3$ do not change with sample size, x changes for each random draw at different sample size. Thus, for each sample size the expected $\bar{W}$ is calculated as the true mean, and then the bias for each method at each sample size is calculated by $\widehat{W}_i - \bar{W}, i = D, J, B$.

From (34), the delta method estimator is

$$\widehat{W}_D \cong \frac{\widehat{\beta}_0 + \widehat{\beta}_1 \bar{x}}{-\widehat{\beta}_2} - \frac{Cov(\widehat{\beta}_0, \widehat{\beta}_2) + Cov(\widehat{\beta}_1, \widehat{\beta}_2) \bar{x}}{\widehat{\beta}_2^2} + \frac{(\widehat{\beta}_0 + \widehat{\beta}_1 \bar{x}) * V(\widehat{\beta}_2)}{\widehat{\beta}_2^3} \quad (52)$$

The simulation results are presented in Table 2.3 and Table 2.4.

Table 2.3: Monte Carlo analysis of WTP expected bias for logit model

Sample size	40	80	100	200	400	1000
MEAN						
True	206.736	206.932	206.694	206.916	206.721	206.766
Original	213.014	207.411	208.778	208.326	206.723	207.160
Delta	199.452	202.043	204.464	206.215	205.687	206.744
Jackknife	204.879	206.682	208.212	208.133	206.643	207.129
Bootstrap	207.993	207.413	207.609	207.308	207.314	206.489
BIAS						
Original	6.278	0.479	2.084	1.411	0.002	0.394
Delta	-7.284	-4.889	-2.230	-0.701	-1.035	-0.022
Jackknife	-1.857	-0.249	1.518	1.217	-0.079	0.363
Bootstrap	1.258	0.481	0.916	0.392	0.592	-0.277

Note: true values are different for each sample size due to randomly drew variable x.

Table 2.4: Monte Carlo analysis of WTP mean square error for logit model

Sample size	40	80	100	200	400	1000
STD.DEV						
Original	66.723	39.652	36.732	24.388	16.674	10.536
Delta	57.756	37.852	35.364	23.971	16.537	10.502
Jackknife	101.144	38.684	35.948	24.178	16.612	10.521
Bootstrap	58.159	38.548	35.118	23.715	16.733	11.075
MSE						
Original	4491.311	1572.489	1353.598	596.766	278.026	111.162
Delta	3388.809	1456.645	1255.567	575.081	274.536	110.296
Jackknife	10233.516	1496.499	1294.593	586.045	275.955	110.829
Bootstrap	3384.098	1486.141	1234.144	562.536	280.346	122.722

The logit result looks indecisive. For the expected bias, each method performs best under two sample sizes. The original and the delta has the smallest MSE when sample size gets larger, while the bootstrap has the smallest MSE when sample size is smaller. The delta method generally performs better when the sample size gets large.

The original estimation result is biased upwards, while the delta result is biased downwards under all sample sizes. The intuition behind this is that the delta method can “overcorrect” by omitting a higher order term that has an opposite sign as the second order term included in the formula. One evidence backing this is that at sample size 400 when the original result has only slightly upward bias which is also the smallest (0.002), the delta method has the largest bias in the opposite direction (-1.035), although it still has the smallest MSE. The sign of bias for resampling techniques like bootstrap and jackknife can be indeterminate under different sample sizes. Thus, in practical scenarios when knowing the direction of the bias is important, the delta method has extra advantage.

2.7 Probit Model

Like the logit model, the probit model has a binary choice variable y representing whether to accept the bid, x representing conditioning variables, and A representing the bid amount offered. By assuming latent unobserved WTP as Gaussian distributed, we can get the true data generating process specified as

$$Prob(y = 1) = \Phi(\beta_0 + \beta_1\bar{x} + \beta_2A) \quad (53)$$

$\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Here $\bar{x}$ is the sample mean of x . The expected WTP formula is same as (51) for logit.

For Monte Carlo, we generate simulated data using (53), and fit the data using the probit model. Identical to the logit model settings, we assume $\beta_0 = 1.4136$, $\beta_1 = 0.00372$, and $\beta_2 = -0.008561$. Amenity variable x is generated with random draws from a normal distribution with mean 95.78 and standard deviation 119.226. A is generated by a repeating sequence (2.5, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 150, 200, 250, 300, 400, 500, 700). The dependent variable y is generated by estimating $P_i = \widehat{Prob}(y_i = 1)$ for each observation with (53), and using a Bernoulli distribution with parameter P_i to generate a “1 or 0” response. Following these procedures, the true $E(WTP)$ are \$206.74. Because x is randomly generated, for every simulation we plug the $\bar{x}$ calculated from the generated x into (51) to get the distribution of true WTP for each sample size. The sample sizes and number of simulations are also identical as for the logit model. And the delta method estimator is also calculated as (52).

Since almost all specifications of simulation are identical for logit and probit except the link function, we can attribute the difference in the simulation results to the effect of the change of link function of generalized linear model.

The simulation results are presented in Table 2.5 and Table 2.6.

Table 2.5: Monte Carlo analysis comparing several methods for mean & bias of welfare measure for Proit model

Sample size	40	80	100	200	400	1000
MEAN						
TRUE	206.610	206.932	206.694	206.916	206.721	206.766
Original	204.383	206.636	205.680	207.080	206.304	206.852
Delta	199.768	203.188	202.947	205.714	205.627	206.581
Jackknife	-	207.292	206.317	206.317	206.488	206.927
BIAS						
Original	-2.227	-0.296	-1.014	0.164	-0.417	0.087
Delta	-6.842	-3.743	-3.746	-1.201	-1.095	-0.185
Jackknife	-	0.361	-0.376	-0.598	-0.233	0.161

Table 2.6: Monte Carlo analysis of WTP mean square error for Probit model

Sample size	40	80	100	200	400	1000
STD.DEV						
Original	32.680	28.544	25.966	17.288	12.379	7.833
Delta	31.807	27.961	25.510	17.142	12.330	7.820
Jackknife	-	28.475	25.866	17.291	12.383	7.832
MSE						
Original	1072.943	814.825	675.240	298.894	153.425	61.367
Delta	1058.473	795.836	675.240	295.279	153.216	61.190
Jackknife	-	810.949	669.210	299.245	153.386	61.364

For the probit model, the original estimation has least expected bias for sample sizes 80, 200, and 1000, and the jackknife result has the least expected bias for sample size 100 and 400; so basically the jackknife and delta method are not effective in correcting the expected bias. However, when we look at the MSE results, the delta method still has the best performance, and the jackknife performs the best at sample size 100. The probit also has less MSE reduction through bias correction methods than logit.

The reason for the performance change between logit and probit is related to the bias in estimated parameters because the nonlinear transformations are identical. When sample size is small, the parameter estimates of probit are less biased than those of logit, and this gap is magnified through the nonlinear transformation. In section 2.9 we will analyze the bias correction of parameter estimates and associated WTP bias correction.

2.8 Misspecified Probit

To test the robustness of bias correction methods to distributional misspecification, we randomly generated the data using logistic assumptions as (50), but estimate the data using a probit model as (53). The simulation and bias correction procedures are essentially the same as what were done with the probit in the last section, and we only used the delta method and jackknife for bias correction. The expected bias and MSE results are listed in Table 2.7 and Table 2.8.

Table 2.7: Monte Carlo analysis comparing several methods for mean & bias of welfare measure for misspecified Probit model

Sample size	40	80	100	200	400	1000
MEAN						
True	206.733	206.931	206.696	206.916	206.718	206.766
Original	213.882	209.031	210.868	210.781	209.668	210.251
Delta	200.735	204.085	206.919	208.861	208.732	209.876
Jackknife	207.356 (210.231)	209.606	211.472	211.197	209.938	210.374
BIAS						
Original	7.149	2.100	4.172	3.865	2.950	3.485
Delta	-5.998	-2.846	0.224	1.945	2.014	3.110
Jackknife	0.620 (3.496)	2.674	4.779	4.282	3.217	3.609

Note: the value in the parenthesis under jackknife is the value calculated after eliminating one extreme estimate.

Table 2.8: Monte Carlo analysis of WTP mean square error for misspecified Probit model

Sample size	40	80	100	200	400	1000
STD.DEV						
Original	66.728	39.513	36.844	24.480	16.803	10.542
Delta	58.260	37.815	35.565	24.096	16.680	10.511
Jackknife	106.997	38.719	36.277	24.327	16.768	10.538
	(56.52)					
MSE						
Original	4503.703	1565.662	1374.910	614.211	291.052	123.279
Delta	3430.171	1438.069	1264.893	584.424	282.263	120.159
Jackknife	11448.739	1506.292	1338.851	610.140	291.522	124.077
	(3206.719)					

Note: the value in the parenthesis under jackknife is the value calculated after eliminating one extreme estimate.

As is shown in the simulations the misspecification results in inconsistent estimation. Even at the sample size of 1000, there is still systematic bias after any bias correction. However, in general, the delta method is more robust to this misspecification. And for jackknife, in one simulation the estimate is so severely biased that it has a tremendous effect on the average value. In a future study, we will use different misspecification with multiple scales of deviation of skewness and kurtosis to compare the performances of bias correction methods.

The implication from the misspecified result is very important to understand the effectiveness of applying these bias correction methods in practice, and the key is the interaction between different bias sources. The delta method can effectively correct bias caused by nonlinear transformation, and it can also correct the part of other type of biases interacted with nonlinear transformation. Thus, no matter what are the bias sources, as long as they are magnified or interacted with nonlinear transformation, the interaction bias influence can be corrected by the delta method.

The order of the occurrences of bias also matters for bias correction. If our goal is to obtain more precise WTP estimates, then the nonlinear transformation bias is crucial because it is the last step to get the WTP. By applying the delta method, we correct the nonlinear transformation bias and its interaction with other bias sources caused in previous steps. However, as shown in section 2.9, correcting the parameter estimation bias before applying the delta method does not produce better results. The reason is that the bias correction of any bias source also influences the interaction part, if we deal with different but interacted bias sources at the same time with separate methods, the interaction part would be treated twice, which causes less effective bias correction. Unless an integral bias correction method is developed which can fix the bias

interaction problem, the practical strategy is to just correct the most prominent or “latest” bias source (e.g. for WTP, the nonlinear transformation bias).

2.9 Small Sample Parameter Estimation Bias

In common delta method research and application, the parameter is usually assumed to be unbiased so that the first order bias term is eliminated. However, when sample size is small the estimation bias of parameter estimates becomes nontrivial and can have combined effect with the nonlinear transformation bias. Here we recover the first order bias in the delta method and use (33) to calculate bias in welfare measures. To show this result in matrix-vector notation, let $\mathbf{b}$ be a $k \times 1$ vector of estimated parameter biases, let $\mathbf{d}$ be a $k \times 1$ vector whose j th element is the mean of $\partial h(\hat{\boldsymbol{\theta}})/\partial \hat{\theta}_j$ over the sample, and let $\mathbf{D}$ be the $k \times k$ matrix whose j,l th element is the mean of $\partial^2 h(\hat{\boldsymbol{\theta}})/\partial \hat{\theta}_j \partial \hat{\theta}_l$ over the sample, then (33) can be written as

$$\overline{h(\boldsymbol{\theta})} \cong \overline{h(\hat{\boldsymbol{\theta}})} - \mathbf{b}'\mathbf{d} - 1/2 \text{trace}(\mathbf{V}(\hat{\boldsymbol{\theta}})\mathbf{D}) \quad (54)$$

Because $\mathbf{d}$ can be derived and calculated easily for a given model specification, we only need to estimate $\mathbf{b}$ to implement (54). Here we only discuss first order bias in maximum likelihood estimation because ML is most commonly used to estimation in the CV literature. When distributional assumptions are correct, ML estimates are more efficient than nonparametric estimates. The bias of ML estimate can be analytically derived asymptotically or through simulation in finite samples.

2.9.1 Bias in Small Sample Maximum Likelihood Parameter Estimates

It has been noted in earlier studies (Cox & Snell (1968, §3), Anderson & Richardson (1979) and Schaefer (1983)) that bias is large for ML estimator in models with small and moderate sample sized or with small total Fisher information.

Cox & Snell (1968, §3) derived the expression for first order bias of ML estimates for some basic models. Sowden (1971, 1972) gave expressions for bias of maximum likelihood estimators correct to terms of order n^{-1} for logit model, and showed ML performs better than minimum chi-squared and modified minimum chi-squared in the quantal response model. Anderson & Richardson (1979) and Schaefer (1983) corrected first order bias for logistic regressions.

Amemiya (1981) showed that maximum likelihood estimates of the logit model are consistent and asymptotically normal with desirable properties. Amemiya (1980, 1984) derived analytic expressions for the $O(n^{-2})$ MSE of the MLE estimator for logit model.

Cordeiro (1991) derived general formulae for first-order biases of maximum likelihood estimates of the linear parameters, linear predictors, the dispersion parameter and fitted values in generalized linear models. Cordeiro (1997) also derived general formulae for second-order biases of maximum likelihood estimates for some multivariate nonlinear regression models. Firth (1993) showed that first-order bias of maximum likelihood estimates can be removed by a suitable modification of the score function, and in some other generalized linear models, the Jeffreys prior penalty function can be imposed for bias correction.

MacKinnon (1998) derived a “bias function”, which relates the bias of some estimators to the parameter values, through simulation estimation and analytical approximation; and concluded that if the bias function slopes downward, the bias-correcting estimators will have larger variances, even larger MSE than the initial estimator. If the bias function slopes upward, the

bias-correcting estimators will have smaller variances than the initial estimators. If the bias function is non-linear, the bias-correcting estimators will be biased at $O(n^{-2})$. Chen (2012) gave first-order bias and second-order mean squared error function for finite sample maximum likelihood estimator in the binary logit model with random covariates. For various data distributions, the bias of the estimator is signed the same as the covariate's coefficient, and both the absolute bias and the mean squared errors increase symmetrically with the absolute value of that parameter.

2.9.2 Bias Correction for Small Sample Maximum Likelihood Parameter Estimates in Generalized Linear Models

As listed in 2.3.1, there are multiple approaches to correct the parameter estimation bias in maximum likelihood. Maiti (2008) compared the performance of Maximum Likelihood (MLE), Conditional Maximum Likelihood (CMLE), Firth (1993)'s adjusting score function method, and Cordeiro (1991)'s closed form bias correction method in correcting small sample estimation bias correction for logistic regression in extensive simulations, and recommended Cordeiro (1991)'s method because it is the closest to the actual value and it has a closed form expression. In the following logistic regression example, we will use Cordeiro's method to correct small sample estimation bias.

2.9.3 Simulation for Small Sample Parameter Estimation Bias Correction in Logistic Model

When applying the delta method in previous simulation studies, we assume the unbiasedness of the parameter estimates. However, if the sample size is small then the parameter estimation bias becomes nontrivial. For the same logistic data generating process as in 2.7, we can use

Cordeiro's method to correct the estimation bias. We denote the corrected parameter estimates using Cordeiro's method as $\overline{\beta_0}, \overline{\beta_1}, \overline{\beta_2}$ corresponding to the original parameter estimates $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}$. The true values of the coefficients are $\beta_0 = 1.4136$, $\beta_1 = 0.00372$, and $\beta_2 = -0.008561$. The parameter bias correction results are listed in FIGURE 2.1 and FIGURE 2.2. FIGURE 2.1 shows the expected bias calculated as the mean of bias over the 1000 simulations, and FIGURE 2.2 shows the MSE calculated over the 1000 simulations.

2.9.3.1 Simulation Results: Parameter Estimates Bias Correction

FIGURE 2.1 and FIGURE 2.2 show that the Cordeiro's method can effectively reduce the parameter estimation bias when sample size is small. For constant coefficient β_0 and bid variable coefficient β_2 , the expected bias can be reduced significantly as the sample size grows all the way to 800, so are the MSE.

For amenity variable coefficient β_1 , when sample size is 400 and 800, the expected bias becomes very small and the expected biases increase after applying the Cordeiro's method, although the MSEs are slightly reduced.

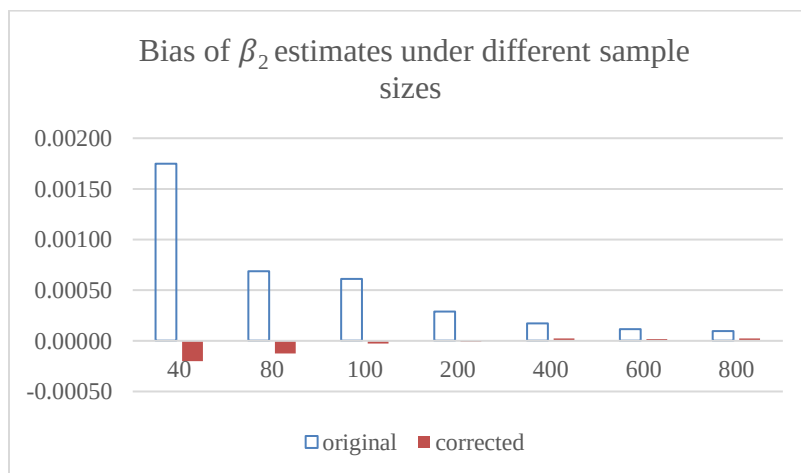
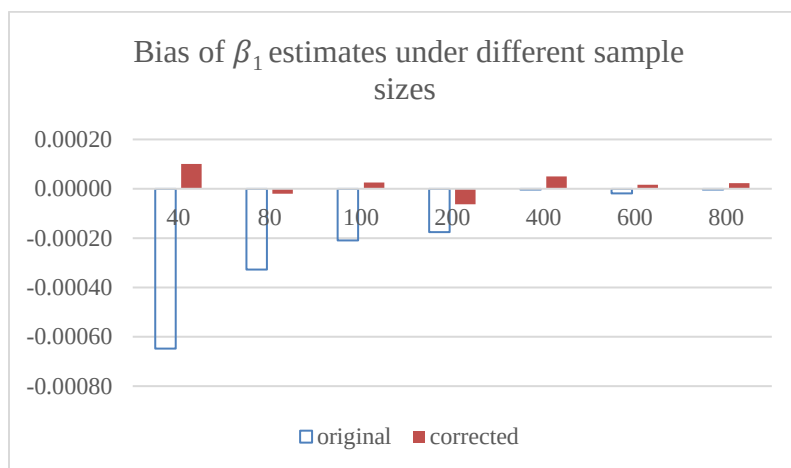
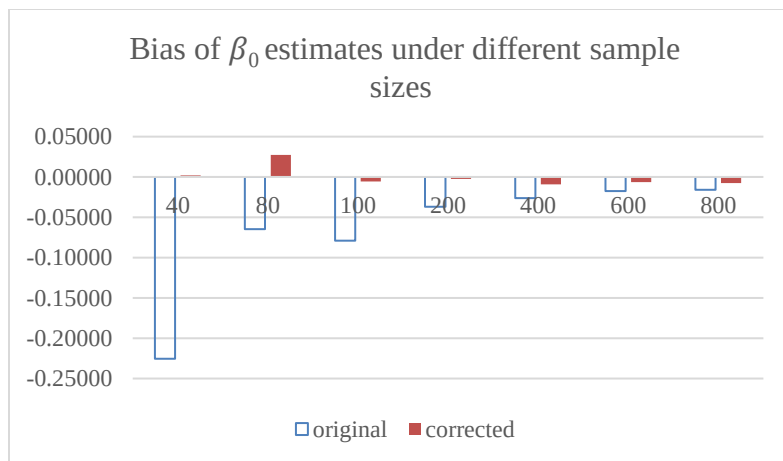


FIGURE 2.1: Monte Carlo analysis of bias correction in maximum likelihood parameter estimates: compare expected bias under different sample sizes

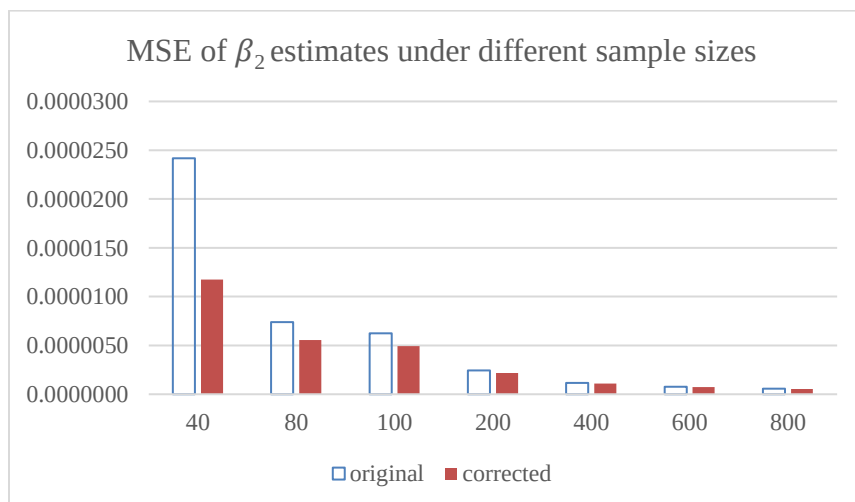
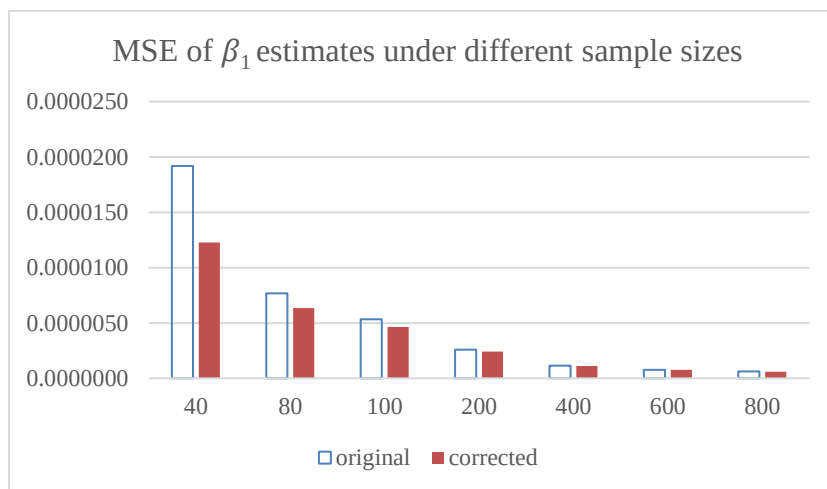
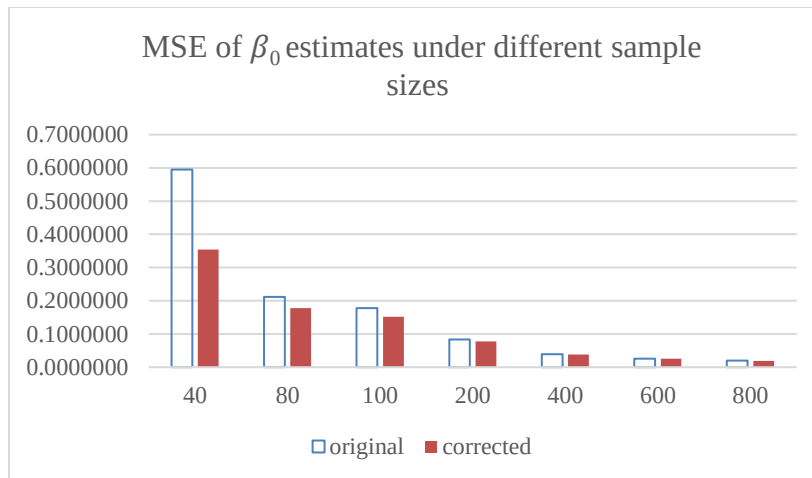


FIGURE 2.2: Monte Carlo analysis of bias correction in maximum likelihood parameter estimates: compare mean square error under different sample sizes

2.9.3.2 Simulation Results: Mean WTP Estimates Bias Correction

We then incorporate the corrected parameter estimates into the delta method procedures in different ways. We can directly replace the coefficients' mean estimates in the delta method formula or in the WTP formula, we can also add the first order term of the Taylor series expansion into the formula.

(a) Plug $\bar{\beta}_1$ directly into the original WTP formula (51)

$$\hat{W}_C \cong \frac{\bar{\beta}_0 + \bar{\beta}_1 \bar{x}}{-\bar{\beta}_2} \quad (55)$$

(b) Plug $\bar{\beta}_1$ into formula (52), variance and covariance terms do not change

$$\hat{W}_{CD} \cong \frac{\bar{\beta}_0 + \bar{\beta}_1 \bar{x}}{-\bar{\beta}_2} - \frac{Cov(\hat{\beta}_0, \hat{\beta}_2) + Cov(\hat{\beta}_1, \hat{\beta}_2) \bar{x}}{\bar{\beta}_2^2} + \frac{(\bar{\beta}_0 + \bar{\beta}_1 \bar{x}) * V(\hat{\beta}_2)}{\bar{\beta}_2^3} \quad (56)$$

(c) Use original parameter estimates, but add the first order term $\mathbf{b}'\mathbf{d}$ in formula (54) into the delta method formula (52)

$$\begin{aligned} \hat{W}_{D1} \cong & \frac{\hat{\beta}_0 + \hat{\beta}_1 \bar{x}}{-\hat{\beta}_2} + \frac{b_0}{\hat{\beta}_2} + \frac{b_1 \bar{x}}{\hat{\beta}_2} - b_2 \frac{\hat{\beta}_0 + \hat{\beta}_1 \bar{x}}{\hat{\beta}_2^2} - \frac{Cov(\hat{\beta}_0, \hat{\beta}_2) + Cov(\hat{\beta}_1, \hat{\beta}_2) \bar{x}}{\hat{\beta}_2^2} \\ & + \frac{(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}) * V(\hat{\beta}_2)}{\hat{\beta}_2^3} \end{aligned} \quad (57)$$

(d) Plug $\bar{\beta}_1$ into the Taylor series expansion formula (54), this is different from (56) in that it has the first order bias term in the Taylor series expansion.

$$\begin{aligned} \hat{W}_{CD1} \cong & \frac{\bar{\beta}_0 + \bar{\beta}_1 \bar{x}}{-\bar{\beta}_2} + \frac{b_0}{\bar{\beta}_2} + \frac{b_1 \bar{x}}{\bar{\beta}_2} - b_2 \frac{\bar{\beta}_0 + \bar{\beta}_1 \bar{x}}{\bar{\beta}_2^2} - \frac{Cov(\hat{\beta}_0, \hat{\beta}_2) + Cov(\hat{\beta}_1, \hat{\beta}_2) \bar{x}}{\bar{\beta}_2^2} \\ & + \frac{(\bar{\beta}_0 + \bar{\beta}_1 \bar{x}) * V(\hat{\beta}_2)}{\bar{\beta}_2^3} \end{aligned} \quad (58)$$

The expected bias and MSE results are listed at Table 2.9 and Table 2.10.

Table 2.9: Monte Carlo analysis of mean WTP expected bias for logit model: combine the delta method and Cordeiro's method

Sample size	40	80	100	200	400	600	800
MEAN							
True	206.74	206.93	206.69	206.92	206.72	206.75	206.81
Original	213.01	207.41	208.78	208.33	206.71	206.95	206.91
Delta	199.45	202.04	204.46	206.21	205.67	206.26	206.39
C	224.11	212.54	212.85	210.31	207.68	207.60	207.40
CD	201.71	205.66	207.61	207.99	206.60	206.89	206.87
D1	178.87	192.67	196.91	202.55	203.84	205.05	205.49
CD1	176.76	195.46	199.54	204.21	204.74	205.66	205.95
BIAS							
Original	6.28	0.48	2.08	1.41	-0.02	0.20	0.10
Delta	-7.28	-4.89	-2.23	-0.70	-1.05	-0.49	-0.42
C	17.37	5.60	6.16	3.40	0.96	0.85	0.59
CD	-5.03	-1.28	0.92	1.07	-0.13	0.13	0.06
D1	-27.87	-14.26	-9.78	-4.36	-2.88	-1.71	-1.33
CD1	-29.97	-11.48	-7.16	-2.71	-1.98	-1.09	-0.86

Note: true values are different for each sample size due to randomly drawn variable x.

Table 2.10: Monte Carlo analysis of mean WTP mean square error for logit model: combine the delta method and Cordeiro's method

Sample size	40	80	100	200	400	600	800
STD.DEV.							
Original	65.75	38.64	36.31	24.34	16.59	13.53	11.58
Delta	56.78	36.84	34.95	23.93	16.46	13.45	11.53
C	68.74	39.59	37.00	24.57	16.67	13.57	11.61
CD	63.65	37.39	35.40	24.12	16.53	13.49	11.56
D1	53.08	35.53	33.89	23.62	16.34	13.40	11.49
CD1	60.19	36.00	34.30	23.80	16.41	13.44	11.52
MSE							
Original	4453	1573	1350	596	278.2	192.49	138.22
Delta	3337	1434	1252	576	273.6	190.41	137.09
C	4862	1650	1402	607	280.8	193.65	138.85
CD	4164	1476	1284	585	276.0	191.51	137.70
D1	2931	1339	1180	561	270.0	188.94	136.18
CD1	3737	1374	1208	570	272.3	190.03	136.78

The rank of lowest expected bias: CD, original, delta, C, CD1, D1.

The rank of least MSE: D1, CD1, delta, CD, original, C.

The “original” results and “C” results are biased upwards, both of which do not use the delta method formula. “C” results are derived by plugging corrected parameter estimates directly into the WTP nonlinear function, which have even more expected bias and MSE than the original results. This indicates that if we only correct bias in estimated parameters, but do not correct nonlinear transformation bias in WTP, we could end up getting even worse WTP estimates.

The “CD” results are derived by plugging corrected parameter estimates into the delta method formula, which generally have the least expected bias. The “delta” results rank 2nd lowest in expected bias. Both use the original delta method formula and do not add the first order bias term. This indicates that the delta method formula is more effective in lowering the expected bias than the modified formula adding the first order bias term.

The “D1” results are derived by adding the first order bias term derived by using the original parameter estimates. The “CD1” results are derived by replacing the estimated parameters in the “D1” formula with the correct parameters. “D1” results have the least MSE, followed by “CD1”; while “D1” results have the highest expected bias, also followed by “CD1”. This indicates that the “complete” delta method formula with the first order term added can increase the efficiency of WTP estimation at the cost of adding some expected bias.

2.9.3.3 Simulation Results: Median WTP Estimates Bias Correction

In this section, we extend the simulation to median WTP. Because median is calculated in a nonlinear fashion, the resampling method like jackknife is expected to be less effective comparing with when used in the mean WTP situation.

The data generating process is the same as the mean WTP simulation in 2.9.3.2, but here we estimate the median WTP instead. Based on Hanemann (2001) formula (38) (39), we only need to replace the $\bar{x}$ in (55) - (58) with median of x to generate median WTP estimates. Here are the simulation results. We firstly estimated the median in each simulation and averaged over 1000 simulations to get the median WTP expected bias, then calculated the variance over 1000 simulations. The jackknife results are also included as a reference.

The “CD” result still has the lowest expected bias as in the mean WTP case, but the rank of performance by expected bias become more indecisive. The “D1” result still has the lowest standard deviation, but has more expected bias. Surprisingly, the “delta” result has the smallest MSE among all methods. In generally for the median WTP results all bias correction method become less effective because the MSE decrease from the “original” result becomes smaller. However, the delta method can still consistently reduce the bias.

Table 2.11: Monte Carlo analysis of median WTP expected bias for logit model: combine the delta method and Cordeiro's method

Sample size	40	80	100	200	400	600	800
MEAN							
True	206.65	206.91	206.70	206.94	206.74	206.77	206.81
Original	212.79	207.23	208.76	208.38	206.71	206.96	206.89
Jackknife	204.73	206.53	208.21	208.19	206.63	206.91	206.85
Delta	199.20	201.88	204.45	206.27	205.67	206.27	206.37
C	223.96	212.37	212.85	210.37	207.68	207.61	207.37
CD	201.50	205.50	207.61	208.05	206.60	206.90	206.84
D1	178.67	192.52	196.91	202.61	203.84	205.06	205.46
CD1	176.65	195.32	199.55	204.27	204.74	205.67	205.93
BIAS							
Original	6.14	0.33	2.06	1.44	-0.03	0.19	0.08
Jackknife	-1.92	-0.37	1.51	1.25	-0.11	0.14	0.05
Delta	-7.45	-5.03	-2.25	-0.67	-1.07	-0.50	-0.44
C	17.30	5.46	6.14	3.43	0.95	0.84	0.57
CD	-5.15	-1.40	0.90	1.11	-0.14	0.13	0.04
D1	-27.98	-14.39	-9.79	-4.33	-2.90	-1.71	-1.35
CD1	-30.01	-11.59	-7.16	-2.67	-2.00	-1.10	-0.88

Note: true values are different for each sample size due to randomly drawn variable x.

Table 2.12: Monte Carlo analysis of median WTP mean square error for logit model: combine the delta method and Cordeiro's method

Sample size	40	80	100	200	400	600	800
STD.DEV.							
Original	67.35	39.86	36.77	24.81	16.74	13.88	11.78
Jackknife	101.69	38.90	36.01	24.60	16.68	13.85	11.75
Delta	58.15	38.05	35.41	24.39	16.60	13.81	11.73
C	70.45	40.82	37.48	25.04	16.82	13.92	11.80
CD	64.94	38.62	35.88	24.58	16.67	13.85	11.75
D1	54.67	36.82	34.41	24.10	16.49	13.75	11.69
CD1	61.72	37.32	34.84	24.28	16.56	13.79	11.72
MSE							
Original	4574	1589	1357	618	280	192.72	138.71
Jackknife	10345	1514	1299	607	278	191.87	138.17
Delta	3437	1473	1259	595	277	190.85	137.75
C	5262	1697	1442	639	284	194.57	139.66
CD	4244	1493	1288	606	278	191.73	138.18
D1	3771	1563	1280	599	280	192.08	138.47
CD1	4710	1527	1265	597	278	191.46	138.04

2.10 Conclusion

The delta method, jackknife and bootstrap can effectively correct the bias of welfare measures caused by nonlinear transformation in small samples (less than 400). When sample size is too small, the bias of parameters is transitory, which can disturb and weaken the bias correction method. When sample size is large enough (over 1000), the results converge to the asymptotic situation, which almost eliminates both the parameter estimation bias and nonlinear transformation bias, but not the systematic misspecification bias.

The delta method generally performs best in the sense that it usually has the smallest expected bias and MSE among these three, although the delta method and jackknife have similar performance for the Poisson model. Also, the delta method would generally have expected bias with opposite direction as the original estimation does, which is useful when the direction of bias is important information. With distributional misspecification, the delta method can still effectively correct the bias, but the jackknife method cannot. When the model is correctly specified, jackknife can work as a substitute for the delta method given that it is very easy to apply and computationally inexpensive. Bootstrap performs the best when the sample is sufficiently large (more than 1,000), which is consistent with asymptotic results.

Cordeiro's method can reduce the bias of parameter estimates, and adding the first order bias term derived by Cordeiro's method into the delta method formula can improve the efficiency of WTP estimates. The expected bias can be lowered at the cost of higher variance and MSE. The probable reason is that although the absolute value of the expected bias of parameter estimates can be corrected through Cordeiro's method, the direction of bias in estimated parameter can be altered, which may increase the bias in WTP when these corrected parameter results are plugged into the nonlinear function of the delta method WTP formula or the original WTP formula.

Besides, the parameter estimation bias is interacted in complicated ways with the nonlinear transformation bias as can be seen in the Taylor series expansion, the change of sign of bias can have different effects for odd terms and even terms in the Taylor series expansion. If the WTP can be derived as a linear function of estimated parameters as shown by Cameron (1988), then the parameter estimation bias may perform effectively in bias correction of WTP estimates.

Finally, the interaction between different bias sources, and the order of the occurrences of different types of biases are important for choosing the bias correction methods. If our goal is to obtain more precise WTP estimates, then the nonlinear transformation bias is crucial because it is the last step to get the WTP. By applying the delta method, we correct the nonlinear transformation bias and its interaction with other bias sources caused in previous steps. If we only want better parameter estimates, then Cordeiro's method can meet our need. However, as shown in section 2.9, correcting the parameter estimation bias before applying the delta method does not produce more precise WTP estimates. Unless an integral bias correction method is developed which can fix the bias interaction problem, the practical strategy is to just correct the most prominent or "latest" bias source. For WTP estimates derived from nonlinear transformation of estimated parameters, the most effective way is to correct the nonlinear transformation bias using the delta method or jackknife.

REFERENCES

- Alberini, A. (1995). Efficiency vs Bias of Willingness-to-Pay Estimates: Bivariate and Interval-Data Models. *Journal of Environmental Economics and Management*, 29(2), 169-180.
- Alberini, A. (1995). Optimal Designs for Discrete Choice Contingent Valuation Surveys: Single-Bound, Double-Bound, and Bivariate Models. *Journal of Environmental Economics and Management*, 28(3), 287-306.
- Álvarez Díaz, M., González Gómez, M., Saavedra González, Á., & De Uña Álvarez, J. (2010). On dichotomous choice contingent valuation data analysis: Semiparametric methods and Genetic Programming. *Journal of Forest Economics*, 16(2), 145-156.
- Amemiya, T. (1980). The n^{-2} Order Mean Squared Errors of the Maximum Likelihood and the Minimum Logit Chi-Square Estimator. *The Annals of Statistics*, 8(3), 488-505.
- Amemiya, T. (1981). Qualitative response models: a survey. *Journal of Economic Literature*, 19, 1483-1536.
- Araña, J. E., & León, C. J. (2005). Flexible mixture distribution modeling of dichotomous choice contingent valuation with heterogeneity. *Journal of Environmental Economics and Management*, 50(1), 170-188.
- Belluzzo, W. (2004). Semiparametric Approaches to Welfare Evaluations in Binary Response Models. *Journal of Business & Economic Statistics*, 22(3), 322-330.
- Bengochea-Morancho, A., Fuertes-Eugenio, A. M., & Del Saz-Salazar, S. (2005). A comparison of empirical models used to infer the willingness to pay in contingent valuation. *Empirical Economics*, 30(1), 235.

- Bishop, R. C., & Heberlein, T. A. (1979). Measuring Values of Extramarket Goods: Are Indirect Measures Biased? *American Journal of Agricultural Economics*, 61(5), 926-930.
- Bliemer, M. C. J., & Rose, J. M. (2013). Confidence intervals of willingness-to-pay for random coefficient logit models. *Transportation Research Part B: Methodological*, 58, 199-214.
- Boman, M., Bostedt, G., xf, ran, Kristr, xf, & m, B. (1999). Obtaining Welfare Bounds in Discrete-Response Valuation Studies: A Non-Parametric Approach. *Land Economics*, 75(2), 284-294.
- Calia, P., & Strazzera, E. (2000). Bias and efficiency of single versus double bound models for continent valuation studies : a Monte Carlo analysis. *Applied economics*, 32(10), 1329-1336.
- Cameron, T. A. (1988). A new paradigm for valuing non-market goods using referendum data: Maximum likelihood estimation by censored logistic regression. *Journal of Environmental Economics and Management*, 15(3), 355-379.
- Cameron, T. A., & James, M. D. (1987). Efficient Estimation Methods for 'Closed-ended' Contingent Valuation Surveys. *Review of Economics and Statistics*, 69(2), 269-276.
- Chen, H. Z., & Randall, A. (1997). Semi-nonparametric estimation of binary response models with an application to natural resource valuation. *Journal of Econometrics*, 76(1-2), 323-340.
- Chen, Q., & Giles, D. (2012). Finite-sample properties of the maximum likelihood estimator for the binary logit model with random covariates. *Statistical Papers*, 53(2), 409.
- Chow, S.C. (1985). Resampling procedures for the estimation of nonlinear functions of parameters. (Doctoral dissertation)

- Cooper, J. C. (1994). A Comparison of Approaches to Calculating Confidence Intervals for Benefit Measures from Dichotomous Choice Contingent Valuation Surveys. *Land Economics*, 70(1), 111-122.
- Cooper, J. C. (2000). Nonparametric and Semi-nonparametric Recreational Demand Analysis. *American Journal of Agricultural Economics*, 82(2), 451-462.
- Cooper, J. C. (2002). Flexible Functional Form Estimation of Willingness to Pay Using Dichotomous Choice Data. *Journal of Environmental Economics and Management*, 43(2), 267-279.
- Cordeiro, G. M., & McCullagh, P. (1991). Bias Correction in Generalized Linear Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3), 629-643.
- Creel, M. (1998). A Note on Consistent Estimation of Mean WTP Using a Misspecified Logit Contingent Valuation Model. *Journal of Environmental Economics and Management*, 35(3), 277-284.
- Creel, M., & Loomis, J. (1997). Semi-nonparametric Distribution-Free Dichotomous Choice Contingent Valuation. *Journal of Environmental Economics and Management*, 32(3), 341-358.
- Crooker, J. R., & Herriges, J. A. (2004). Parametric and Semi-Nonparametric Estimation of Willingness-to-Pay in the Dichotomous Choice Contingent Valuation Framework. *Environmental and Resource Economics*, 27(4), 451-480.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. 1-26.
- Efron, B. (1981). Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap and Other Methods. *Biometrika*, 68(3), 589-599.

- Efron, B., & Gong, G. (1983). A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *The American Statistician*, 37(1), 36-48.
- Efron, B., & Stein, C. (1981). The Jackknife Estimate of Variance. 586-596.
- Firth, D. (1993). Bias correction of Maximum Likelihood Estimates. *Biometrika*, 80(1), 27-38.
- Giraud, K. L., Loomis, J. B., & Cooper, J. C. (2001). A Comparison of Willingness to Pay Estimation Techniques from Referendum Questions. *Environmental and Resource Economics*, 20(4), 331-346.
- Gray, H. L., Schucany, W. R., & Watkins, T. A. (1975). On the Generalized Jackknife and its Relation to Statistical Differentials. *Biometrika*, 62(3), 637-642.
- Hanemann, M., & Kanninen, B. (2001). 11 The Statistical Analysis of Discrete-Response CV Data147. *Valuing environmental preferences: theory and practice of the contingent valuation method in the US, EU, and developing countries*, 302.
- Haab, T. C., & McConnell, K. E. (1998). Referendum Models and Economic Values: Theoretical, Intuitive, and Practical Bounds on Willingness to Pay. *Land Economics*, 74(2), 216-229.
- Haab, T.C and McConnell, K.E. (2002) Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation. Edward Elgar Publishing, Northampton.
- Hanemann, M., Loomis, J., & Kanninen, B. (1991). Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics*, 73(4), 1255-1263.
- Hanemann, W. M. (1989). Welfare Evaluations in Contingent Valuation Experiments with Discrete Response Data: Reply. *American Journal of Agricultural Economics*, 71(4), 1057-1061.

- Hofler, R. A., & List, J. A. (2004). Valuation on the Frontier: Calibrating Actual and Hypothetical Statements of Value. *American Journal of Agricultural Economics*, 86(1), 213-221.
- Huang, J.-C., Nychka, D. W., & Smith, V. K. (2008). Semi-parametric discrete choice measures of willingness to pay. *Economics Letters*, 101(1), 91-94.
- *Jaeckel, L. A. The Infinitesimal Jackknife. Bell Laboratories
- Kanninen, B. J. (1995). Bias in Discrete Response Contingent Valuation. *Journal of Environmental Economics and Management*, 28(1), 114-125.
- Kerr, G. N. (2000). Dichotomous choice contingent valuation probability distributions. *Australian Journal of Agricultural and Resource Economics*, 44(2).
- Kézdi, G., Hahn, J., & Solon, G. (2002). Jackknife minimum distance estimation. *Economics Letters*, 76(1), 35-45.
- Kim, Y., & Singh, K. (1998). Sharpening estimators using resampling. *Journal of Statistical Planning and Inference*, 66(1), 121-146.
- Kling, C. L., & Thomson, C. J. (1996). The Implications of Model Specification for Welfare Estimation in Nested Logit Models, 103.
- Kosmidis, I. (2007). Bias correction in Exponential Family Nonlinear Models.
- Kriström, B. (1990). A Non-Parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies, 135.
- Langford, I. H., Bateman, I. J., Jones, A. P., Langford, H. D., & Georgiou, S. (1998). Improved Estimation of Willingness to Pay in Dichotomous Choice Contingent Valuation Studies. *Land Economics*, 74(1), 65-75.

- Lee, H., Jogesh Babu, G., & Rao, C. R. (2012). A jackknife type approach to statistical model selection. *Journal of Statistical Planning and Inference*, 142(1), 301-311.
- MacKinnon, J. G., & Smith, A. A., Jr. (1998). Approximate Bias Correction in Econometrics. *Journal of Econometrics*, 85(2), 205-230.
- MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3), 305-325.
- Maiti, T., & Pradhan, V. (2008). A comparative study of the bias corrected estimates in logistic regression. *Statistical Methods in Medical Research*, 17(6), 621.
- McLeod, D. M., & Bergland, O. (1999). Willingness-to-Pay Estimates Using the Double-Bounded Dichotomous-Choice Contingent Valuation Format: A Test for Validity and Precision in a Bayesian Framework. *Land Economics*, 75(1), 115-125.
- Miller, R. G. (1964). A Trustworthy Jackknife. *The Annals of Mathematical Statistics*, 35(4), 1594-1605.
- Oehlert, G. W. (1992). A Note on the Delta Method. *The American Statistician*, 46(1), 27-29.
- Parr, W. C. (1983). A Note on the Jackknife, the Bootstrap and the Delta Method Estimators of Bias and Variance. *Biometrika*, 70(3), 719-722.
- Parr, W. C., & Schucany, W. R. (1980). The Jackknife: A Bibliography. *International Statistical Review / Revue Internationale de Statistique*, 48(1), 73-78.
- Sharot, T. (1976). The Generalized Jackknife: Finite Samples and Subsample Sizes. *Journal of the American Statistical Association*, 71(354), 451-454.
- Sharot, T. (1976). Sharpening the Jackknife. *Biometrika*, 63(2), 315-321.

- SOWDEN, R. R. (1972). On the first-order bias of parameter estimates in a quantal response model under alternative estimation procedures. *Biometrika*, 59(3), 573-579.
- Vasconcellos, G. M. C. K. L. P. (1997). Bias correction for a class of multivariate nonlinear regression models. *Statistics and Probability Letters*, 35(2), 155.
- Venkatachalam, L. (2004). The contingent valuation method: a review. *Environmental Impact Assessment Review*, 24(1), 89-124.
- Watanabe, M., & Asano, K. (2009). Distribution Free Consistent Estimation of Mean WTP in Dichotomous Choice Contingent Valuation. *Environmental & Resource Economics*, 44(1), 1-10.
- Weber, N. C. (1986). The jackknife and heteroskedasticity: Consistent variance estimation for regression models. *Economics Letters*, 20(2), 161-163.
- Withers, C. S., & Nadarajah, S. (2013). Delta and jackknife estimators with low bias for functions of binomial and multinomial parameters. *Journal of Multivariate Analysis*, 118, 138-147.
- Withers, C. S., & Nadarajah, S. (2014). BIAS CORRECTION: THE DELTA METHOD VERSUS THE JACKKNIFE AND THE BOOTSTRAP. *Pakistan Journal of Statistics*, 30(1), 143-151.

CHAPTER 3

CORRECTING SAMPLING BIAS WITH KERNEL MEAN MATCHING

3.1 Introduction

Nonrandom sampling can bias the contingent valuation estimates in two ways. Firstly, when the sample selection process depends on the covariate, despite the parameter estimates are consistent, the WTP estimates are biased due to the divergence between the covariate distributions of the sample and the population; this is usually called non-response bias. Secondly, when the sample selection process depends on the dependent variable or the individual WTP, the parameter estimates become inconsistent, and the WTP estimates are biased and inconsistent (Heckman, 1976); this is usually called sample selection bias.

The existing literatures indicate sampling selection bias is common and large in non-market valuation studies and can be corrected in multiple ways. Edwards et. al. (1987) and Loomis (1987) discuss bias caused by nonrandom nonresponses when generalizing the sample WTP estimate to the population, and list correction methods like weighted average or weighted regressions based on the correlated variable. Bockstael et al. (1990) compare Tobit model, Heckman Model and the Cragg Model for correcting sample selection bias in an empirical recreation demand study. Dalecki et al. (1993) and Whitehead et al. (1993) find no sample selection bias in contingent valuation through phone/mail survey, but do find nonresponse bias for aggregate benefit estimates. Eklöf and Karlsson (1999) investigate the properties of tests for sample selection bias and the losses made by applying estimators assuming no sample selection. Morrison (2000) compares various approaches dealing with aggregation bias of *WTP* estimates

in stated preference studies caused by the divergence between sample and population characteristics and shows that the selection of estimation method can significantly affect the precision of the result. Yoo and Yang (2001) detect both sample selection bias and nonresponse bias in a double-bounded dichotomous choice contingent valuation survey and use a bivariate sample selection model.

This paper introduces the sample reweighting process technique called kernel mean matching (KMM) to non-market valuation. KMM can reweight the training points in a nonparametric way such that the means of the training and testing points in a reproducing kernel Hilbert space (RKHS) are close (Huang, 2006). More readily, KMM can match the joint distribution of multiple covariates, while the existing weighted average or weighted regression techniques can only do this for a single variable. We use Monte Carlo to simulate different sampling bias decomposition scenarios including:

- Only non-response bias;
- Only sample selection bias;
- Sample selection bias + non-response bias;
- Skewed bid design;
- Misspecification with/without non-response bias.

then we test KMM's bias correction performance under these scenarios. We compare KMM parameters and *WTP* estimation results with the results of the original estimation and of substituting the sample mean with the population means. The Monte Carlo results show that

- KMM can correct non-response bias in *WTP* estimates when the sample size is small so that such bias is prominent.
- KMM can marginally correct sample selection bias in *WTP* estimates.

- KMM can correct non-response bias in *WTP* when both non-response bias and sample selection bias exist.
- KMM has robust bias correction performance under skewed bid design.
- KMM can correct the non-response bias in *WTP* estimates that is magnified under misspecification of the model.
- KMM cannot correct the parameter estimates, unless the data generating process has a nonlinear form but a linear approximation model is used for estimation.

3.2 Sampling Bias

3.2.1 An Illustrative Example of Sampling Bias

We illustrate in a data generating process following Heckman (1976, 1979)'s two-step selection model how nonrandom nonresponses in contingent valuation studies can cause bias.

$$y_{i1} = \mathbf{X}_{i1}\boldsymbol{\beta}_1 + \varepsilon_{i1} \quad (59a)$$

$$y_{i2} = \mathbf{X}_{i2}\boldsymbol{\beta}_2 + \varepsilon_{i2} \quad (59b)$$

(59a) is the sample selection equation; (59b) is the *WTP* equation. y_{ij} is transformed into observable binary outcome Y_{ij} as

$$Y_{ij} = \begin{cases} 1, & y_{ij} > 0 \\ 0, & y_{ij} < 0 \end{cases} \text{ for } j = 1, 2. \quad (60)$$

When $Y_{i1} = 1$, the observation i is included in the sample and becomes available for the second equation. Given an initial generated population of size N , N_2 observations are selected into the sample. When $Y_{i2} = 1$, the respondent i would have a “yes” response to certain bid amount due to the utility increase after the compensation.

In contingent valuation practice, researchers typically estimate (59b) under certain distribution assumptions and model specifications. We assume $(\varepsilon_{i1}, \varepsilon_{i2})$ is drawn from a standard bivariate normal distribution with Pearson's correlation parameter ρ , so (59b) can be estimated using a probit model. Here we use a simple form of (59b) for illustration.

$$y_{i2} = \alpha + \gamma * bid_i + \delta * x_i + \varepsilon_{i2} \quad (61)$$

where bid_i is the amount of money individual i is asked to contribute for the welfare change caused by the project we want to value, x_i is the quantity of public goods that can be consumed by individual i . We denote $\bar{x}$ as the population mean of x_i . Then mean WTP is

$$E(WTP) = -\frac{\alpha + \delta * \bar{x}}{\gamma} \quad (62)$$

3.2.2 Decomposition of Sampling Bias

Under this data generating process biases can be caused in different ways. Firstly, when $\rho \neq 0$, nonresponses become dependent on individual WTP and can lead to inconsistent parameter estimates of α , γ and δ in (61) as shown in Heckman (1979)'s two-stage model. This type of bias is commonly called “sample selection bias” in contingent valuation literature (Whitehead et al., 1993; Eklöf and Karlsson, 1999).

If x_{i1} and x_{i2} have common or correlated characteristic variables, nonresponses can cause divergence of distribution of such variables between that of the sample and the population. Although the parameter estimates in (61) will not be affected by common/correlated regressors, nonresponses can still bias the results when we use the sample means of characteristic variables as a substitute of the population means to calculate aggregate WTP estimates. This type of bias caused by nonrandom sampling depending on covariates is typically called nonresponse bias (Whitehead et al., 1993).

3.2.3 Bid Design

Kaninen (1993) suggested that bid design can influence WTP estimation. The more the support of the bid deviates from the domain of the population WTP, the greater is the difference between the estimated mean WTP and the population mean WTP. This indicates that the support of the bid influences the estimated mean WTP.

Problematic bid design can also bias estimation in a similar way as non-responses do. For example, if the bid numbers are centered around the upper tail of the true WTP distribution, the respondents with low WTP would almost always say “no” to the bid, so barely any information will be revealed about preferences at the lower end. It would have an effect equal to allocating low WTP observations less weight in the estimation. There is extensive literature about “optimal bid design” and bias caused by bad bid design (Cooper and Loomis, 1992; Cooper, 1993; Kanninen, 1995; Alberini, 1995).

3.2.4 Misspecification

Habb (1999) uses four non-response model specifications in a simulation study to show existing econometric models designed to account for non-responses are extremely sensitive to misspecification bias. If the functional form of variables in $\mathbf{x}_{i2}$ is misspecified, and such variables are included in or correlated with $\mathbf{x}_{i1}$, the non-responses can further bias the parameter estimates and WTP estimates. For example, if marginal utility of x is diminishing as in (63),

$$y_{i2} = \alpha + \gamma * bid_i + \delta * x_i^{1/2} + \varepsilon_{i2} \quad (63)$$

but for estimation we use a linear form of “income” variable as in (61), then samples with a different distribution of x can produce different parameter estimates, among which the ones with x distributions closest to that of the population would be the best linear approximation. Many

researchers have studied model misspecification in contingent valuation and propose nonparametric/semi-nonparametric methods for bias correction (Kriström, 1990; Creel, 1997; Habb, 1999; Cooper, 2000; Huang, 2008; Criado and Veronesi, 2013).

3.3 Kernel Mean Matching

3.3.1 Sampling Weight

Many sampling bias correction techniques used in econometrics and machine learning consist of reweighting each observation or its function in the training set to more closely reflect the unbiased population distribution. DuMouchel and Duncan (1983) introduced sampling weights to least square estimation. Winship and Radbill (1994) noted that when sampling weights are a function of independent variables, unweighted OLS estimates are consistent and more efficient than Weighted Least Square (WLS); when sampling weights are a function of dependent variables (and thus of the error term), unweighted OLS estimates become inconsistent, and WLS can be preferred.

Shimodaira (2000) developed a pseudo-maximum likelihood estimation by weighting the observed samples in maximizing the log-likelihood function, and showed that under model misspecification (a linear approximation of the polynomial parametric model as illustrated in 3.4) and covariate shift, the optimal choice of the weight function is asymptotically the ratio of the density function of the covariate. Covariate shift assumes that the conditional probability distribution of the output variable given the input variable remains fixed in both the training and test set, i.e. the shift happens only for the marginal probability distributions of the covariates. To correct sampling bias caused by covariate shift the key is estimating the importance weight (density ratio) for the training sample.

Cortes et al. (2008) interpreted the sample reweighting bias correction using Bayes' theorem.

Let X denote the input space and Y the label set. Y can be a discrete choice variable in a classification problem or a subset of $\mathbb{R}$ in a regression estimation problem. Let D denote the population distribution from which we obtain a sample $z = \{(x_1, y_1), \dots (x_n, y_n)\}$ that follows a sample distribution D' . Let s denote the sample selection: when $s = 1$ the data point is sampled; when $s = 0$ it is not. The probability of drawing z per D : $Pr[z]$ is related to the probability of drawing z per D' : $Pr'[z]$ through optimal sampling weight w_i .

$$Pr[z_i] = \frac{Pr[z_i | s = 1] Pr[s = 1]}{Pr[s = 1 | z_i]} = \frac{Pr[s = 1]}{Pr[s = 1 | z_i]} Pr'[z_i] = w_i Pr'[z_i] \quad (64a)$$

$$w_i = \frac{Pr[z_i]}{Pr'[z_i]} \quad (64b)$$

Under covariate shift, the sampling probability is independent of Y , (64a) becomes

$$Pr[x_i] = \frac{Pr[s = 1]}{Pr[s = 1 | x_i]} Pr'[x_i] = w_i Pr'[x_i] \quad (65a)$$

$$w_i = \frac{Pr[x_i]}{Pr'[x_i]} \quad (65b)$$

where w_i is a reweighting factor for the selected sample. We reweight every observation (x_i, y_i) such that observations that are under-represented in Pr' obtain a higher weight, whereas over-represented cases are down-weighted. w_i can be used to correct sampling bias in multiple methods such as weighted least squares (DuMouchel and Duncan, 1983), weighted maximum likelihood (Shimodaira, 2000), and support vector classification (Huang, 2006).

3.3.2 An Empirical Kernel Meaning Matching Optimization Algorithm

Multiple methods have been proposed to estimate the optimal sampling weight from finite samples, including kernel mean matching (KMM) (Huang et al., 2007), logistic regression

(Bickel et al., 2009), and Kullback-Leibler importance estimation (Sugiyama et al., 2008). We follow an empirical KMM proposed by Huang et al. (2007) in the simulation study. Suppose we have two samples $X = (x_1, x_2, \dots, x_m)$ and $X' = (x'_1, x'_2, \dots, x'_n)$, drawn i.i.d. from Pr and Pr' respectively, Kernel Mean Matching can optimize the weights w of observations of X to correct the bias caused by the discrepancy between Pr and Pr' in a nonparametric manner. The empirical KMM algorithm can be summarized as a quadratic problem

$$\underset{w}{\text{minimize}} \quad \frac{1}{2} w^T K w - \kappa^T w \quad \text{subject to } w_i \in [0, W] \text{ and } \left| \sum_{i=1}^m w_i - m \right| \leq m\epsilon \quad (66a)$$

where $K = \{K_{ij} := k(x_i, x_j)\}$, $\kappa = \{\kappa_i := \frac{m}{n} \sum_{j=1}^n k(x_i, x'_j)\}$. $k(x_i, x_j)$ is the kernel, which is a similarity score function for two values (x_i and x_j) or two vectors. The intuition is that when κ_i is large, the corresponding observation in the training sample is more “similar” to the population and should be given more weight. The upper bound of weights W determines the allowed deviation between $w_i Pr'$ and Pr . $\epsilon = (\sqrt{n} - 1)/\sqrt{n}$ determines the convergence condition and ensures that $w(x)Pr(x)$ is close to a probability distribution. It is obvious that we need a larger sample size to get reasonable convergence if Pr' and Pr are very different.

Suppose we have a mapping $\varphi: \mathbb{R}^p \rightarrow \mathbb{R}^q$ that brings a vector in $\mathbb{R}^p$ to $\mathbb{R}^q$. Then the dot product of $\mathbf{x}$ and $\mathbf{y}$ in space $\mathbb{R}^q$ is $\varphi(\mathbf{x})^T \varphi(\mathbf{y})$. A kernel is a function that corresponds to this dot product $k(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x})^T \varphi(\mathbf{y})$. Kernel can be used to compute the dot product or any other “similarity score” of two vectors in some high dimensional feature space without knowing what this space is and what the mapping function φ is. (66a) is derived from its mapping function form as

$$\left\| \frac{1}{m} \sum_{i=1}^m w_i \varphi(x_i) - \frac{1}{n} \sum_{i=1}^n \varphi(x'_i) \right\|^2 = \frac{1}{m^2} w^T K w - \frac{2}{m^2} \kappa^T w + \text{constant} \quad (66b)$$

In this research, we use Gaussian kernel for the “similarity score”.

$$k(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}} \quad (67)$$

The Gaussian kernel represents this similarity as a decaying function of the distance between the vectors (i.e. the squared-norm of their distance). That is, if the two vectors are close together then, $\|x_i - x_j\|$ will be small. Then $k(x_i, x_j)$ will be larger. Thus, closer vectors have a larger Gaussian kernel value than farther vectors. This function is of the form of a bell-shaped curve. The parameter σ sets the width of the bell-shaped curve.

3.4 Monte Carlo Simulation: Probit Model

3.4.1 Data Generating Process

The data generating process follows the structure of the illustrative example in 3.2.1.

$$y_{i1} = a + b * x_{1i} + \varepsilon_{i1} \quad (68a)$$

$$y_{i2} = 1.4136 - 0.008561bid_i + 0.00372x_{2i} + \varepsilon_{i2} \quad (68b)$$

$$Y_{ij} = \begin{cases} 1, & y_{ij} > 0 \\ 0, & y_{ij} < 0 \end{cases} \text{ for } j = 1, 2. \quad (68c)$$

where (68a) is sample selection equation and equation (68b) is the utility difference equation.

We assume $(\varepsilon_{i1}, \varepsilon_{i2}) \sim N \left[\begin{bmatrix} 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \right]$, and $(x_{1i}, x_{2i}) \sim N \left[\begin{bmatrix} 1 & 95.78 \end{bmatrix}, \begin{bmatrix} 1 & \rho_2 \\ \rho_2 & 119.226 \end{bmatrix} \right]$,

the true mean of WTP under this specification would be 206.74; due to the variation of x_2 sampling the true mean for certain simulation scenarios may be slightly different.

3.4.1.1 Sampling Independence

We specify different values of a , b , ρ_1 , ρ_2 , and population size N to simulate different sampling bias scenarios. ρ_1 determines the sample selection process depending on y_2 , or the sample selection bias; when $\rho = 0$ there is no sample selection bias. ρ_2 and b determines whether sampling depends on x , or the non-response bias; when $b = 0$ or $\rho_2 = 0$ there is no non-response bias. a has a “drift” effect on the expected sample selection/response rate. Different combinations of a , b , ρ indicate different sampling bias decomposition. The population sizes used are $N = \{100, 200, 400, 800\}$ for estimation of (68b). The coefficient a is arbitrarily assigned to ensure the selected sample is neither too small for estimation nor too large to be indifferent with the population.

3.4.1.2 Bid Design

In the default setting the variable bid consists of repetitions of sequence $\{140, 175, 205, 245, 275\}$, which follow the “middle only” bid design that used approximately 30th, 40th, 50th, 60th, and 70th percentile of the true WTP distribution.

3.4.1.3 Calibration

When applying KMM, the upper bound $w_i \in [0, W]$ and the kernel parameter σ in (67) need to be calibrated to ensure the robustness of the reweight. A rule of thumb is that robust reweighting should only change the weight slightly so that in each simulation the weight is not “overfitting” and the results will not be too dispersed. The size of σ adjusts the trade-off between bias and variance of KMM estimators.

3.4.2 Simulation Results

Simulation results are obtained and compared under different scenarios. For each scenario, 1000 data sets are generated for simulation, averages and variances of parameter estimates and WTP estimates over 1000 simulation results are calculated to measure the accuracy of mean estimates. A probit model is used to estimate parameters and WTP for the population (“population” rows in the Tables) as well as the selected sample. The population estimation result works as a baseline reference; for a selected sample, two sets of results are obtained and compared: one uses KMM (“KMM” rows in the Tables), the other does not (“sample” rows in the Tables). For WTP estimates, in addition to KMM, a bias correction estimate (“sample_s” rows in the Tables) is obtained using the population mean of x_2 to replace the sample mean of x_2 . This average substitution method is commonly used to correct non-response bias.

3.4.2.1 Non-Response Bias

The Monte Carlo experiment shows that KMM can improve the accuracy of WTP estimates, especially when the non-response bias is large. The parameter estimates will be consistent when there is only non-response bias. KMM does not correct the parameter estimates itself and even increases the bias a little bit. However, when the sample size is small, the parameter estimates will be significantly biased; when they are used to calculate WTP estimates, the parameter of x_2 needs to be multiplied with the biased sample mean of x_2 , which can magnify the bias to an unacceptable scale. This is the situation when KMM become a powerful tool for bias correction. Replacing the sample mean of x_2 with population mean has the similar bias correction effect as KMM, but is less efficient as shown in Table 3.1, 3.2.

TABLE 3.1 Simulation Results for Non-Response Bias¹(N=100, M=50, $a = .1$, $b = .1$, $\rho_1 = 0$, $\rho_2 = 10$, B=1000, $\sigma = .01$)

	WTP	variance
population	206.49	337
sample	204.08	6403
sample_s	204.86	5672
KMM	207.40	2868

	beta1	beta2	beta3	v1	v2	v3
population	1.57	-0.0088	0.0027	1.82	7.92E-06	1.57E-04
sample	1.70	-0.0093	0.0023	4.24	1.97E-05	3.72E-04
KMM	1.71	-0.0093	0.0022	4.28	2.00E-05	3.76E-04

¹ N is the population size, M is the selected sample size, beta1, beta2, beta3 are the average of coefficient estimates for *constant*, *bid*, and x_2 ; v1, v2, v3 are the variances of coefficient estimates for *constant*, *bid*, and x_2 . B is upper bound of weight, σ is the Gaussian kernel standard deviation parameter.

TABLE 3.2 Simulation Results for Non-Response Bias

(N=200, M=105, $a = .1$, $b = .1$, $\rho_1 = 0$, $\rho_2 = 10$, B=1000, $\sigma = .01$)

	WTP	variance
population	206.67	134
sample	241.72	1244579
sample_s	224.92	354040
KMM	206.79	2080

	beta1	beta2	beta3	v1	v2	v3
population	1.46	-0.008731	0.0036	0.8262129	3.78E-06	7.03E-05
sample	1.48	-0.008861	0.0036	1.725996	8.20E-06	1.44E-04
KMM	1.48	-0.00888	0.0037	1.7753478	8.60E-06	1.47E-04

TABLE 3.3 Simulation Results for Non-Response Bias

(N=400, M=224, $a = .1$, $b = .1$, $\rho_1 = 0$, $\rho_2 = 10$, B=1000, $\sigma = .0005$)

	WTP	variance
population	206.63	51.973
sample	206.77	114.86
sample_s	206.47	115.4
KMM	206.75	114.7

	beta1	beta2	beta3	v1	v2	v3
population	1.4244	-0.0087	0.0038	0.41	1.92E-06	3.52E-05
sample	1.4553	-0.0087	0.0035	0.7427	3.53E-06	6.21E-05
KMM	1.4556	-0.0087	0.0035	0.7392	3.53E-06	6.18E-05

TABLE 3.4 Simulation Results for Non-Response Bias

(N=800, M=446, $a = .1$, $b = .1$, $\rho_1 = 0$, $\rho_2 = 10$, B=1000, $\sigma = .0001$)

	WTP	variance
population	206.59	29.259
sample	206.71	53.662
sample_s	206.37	54.029
KMM	206.7	53.624

	beta1	beta2	beta3	v1	v2	v3
population	1.4093	-0.0086	0.0038	0.2089	1.00E-06	1.82E-05
sample	1.3832	-0.0086	0.0041	0.415	1.88E-06	3.54E-05
KMM	1.3833	-0.0086	0.0041	0.4149	1.89E-06	3.54E-05

3.4.2.2 Sample Selection Bias

KMM can lower the variance of WTP when sample size is small, but cannot fix the inconsistency of WTP caused by nonrandom sampling correlated with y_2 (or ε_2). Even when the population size is as large as 400, the WTP estimates and parameter estimates are still severely biased, no matter whether KMM is applied or not. Sample selection bias that is correlated with y_2 (or ε_2) needs to be dealt with by other methods like bivariate probit (Eklöf and Karlsson, 1999). We only report the results up to sample size 400 because there is little difference between different methods at sample size 400.

TABLE 3.5 Simulation Results for Sample Selection Bias

(N=100, M=50, $a = .1$, $b=0$, $\rho_1 = .5$, $\rho_2 = 10$, B=1000, $\sigma = .01$)

	WTP	variance
population	206.41	353
sample	260.66	36402
sample_s	261.10	38597
KMM	259.70	15536

	beta1	beta2	beta3	v1	v2	v3
population	1.54	-0.00887	0.0030	1.77	8.09E-06	1.57E-04
sample	2.22	-0.01018	0.0033	4.26	2.06E-05	3.65E-04
KMM	2.24	-0.01023	0.0032	4.31	2.12E-05	3.69E-04

TABLE 3.6 Simulation Results for Sample Selection Bias

(N=200, M=104, $a = .1$, $b=0$, $\rho_1 = .5$, $\rho_2 = 10$, B=1000, $\sigma = .002$)

	WTP	variance
population	206.40	133
sample	252.73	580
sample_s	252.78	584
KMM	252.57	552

	beta1	beta2	beta3	v1	v2	v3
population	1.45	-0.00872	0.0037	0.78	3.85E-06	6.97E-05
sample	1.98	-0.00972	0.0046	1.82	8.51E-06	1.63E-04
KMM	1.98	-0.00974	0.0046	1.84	8.57E-06	1.64E-04

TABLE 3.7 Simulation Results for Sample Selection Bias

(N=400, M=222, $a = .1$, $b=0$, $\rho_1 = .5$, $\rho_2 = 10$, B=1000, $\sigma = .002$)

	WTP	variance
population	206.79	59
sample	250.26	177
sample_s	250.28	178
KMM	250.25	177

	beta1	beta2	beta3	v1	v2	v3
population	1.42	-0.00865	0.0038	0.412	1.90E-06	3.58E-05
sample	1.98	-0.00952	0.0041	0.801	3.74E-06	6.68E-05
KMM	1.98	-0.00951	0.0040	0.798	3.78E-06	6.66E-05

3.4.2.3 Sample Selection Bias Combined with Non-Response Bias

The results are similar with when there is only non-response bias except for the systematic sample selection bias upwards. KMM can only improve the accuracy of WTP estimate, but not the parameter estimates. When sample size is 200 there is little difference between different methods.

TABLE 3.8 Simulation Results for Sample Selection Bias and Non-Response Bias

(N=100, M=50, $a = .1$, $b = .1$, $\rho_1 = .5$, $\rho_2 = 10$, B=1000, $\sigma = .01$)

	WTP	variance
population	206.41	353
sample	272.22	370100
sample_s	273.53	408774
KMM	243.66	267737

	beta1	beta2	beta3	v1	v2	v3
population	1.54	-0.00887	0.0030073	1.767876	8.09E-06	1.57E-04
sample	2.57	-0.01022	-0.000295	4.255567	2.08E-05	3.52E-04
KMM	2.59	-0.01027	-0.000418	4.344559	2.13E-05	3.60E-04

TABLE 3.9 Simulation Results for Sample Selection Bias and Non-Response Bias

(N=200, M=105, $a = .1$, $b = .1$, $\rho_1 = .5$, $\rho_2 = 10$, B=1000, $\sigma = .001$)

	WTP	variance
population	206.40	133
sample	252.67	574
sample_s	252.61	580
KMM	252.58	561

	beta1	beta2	beta3	v1	v2	v3
population	1.45	-0.00872	0.0037045	0.776177	3.85E-06	6.97E-05
sample	2.29	-0.00975	0.0013827	1.778746	8.81E-06	1.55E-04
KMM	2.29	-0.00976	0.0013677	1.784776	8.82E-06	1.55E-04

3.4.2.4 Skewed Bid Design

KMM can correct the bias caused by skewed bid design, in spite of the fact that the reweighting variable on which KMM is used is not correlated with the bid variable in contingent valuation because the bid variable is randomly generated by drawing from several predetermined values.

Intuitively, the skewed bid design is a special case of biased sampling. A properly designed bid vector covers the most informative section of the WTP distribution, so it can reveal sufficient information to estimate the WTP; while a skewed bid design only covers the lower or upper part of the WTP distribution, so it has limited discriminative power for respondents with their WTP spotted outside the bid vector range (they would all simply say “yes” to the give it or take it question if their WTP are above the bid vector range). It is equivalent to giving the observations with underlying WTP outside the bid vector range less sampling weight.

In the simulations, a repetition of bid vector {140, 205} is used which fails to cover the upper half of the WTP distribution. We test with random sampling ($b = 0$, $\rho_1 = 0$) for sample sizes {100, 200, 400} and non-response bias ($b = .01$, $\rho_1 = 0$) for sample size 200. For small samples, even the population estimates have high tendency to be biased (the true mean is approximately 206); and the selected sample estimates can be severely biased as shown in Table 3.10, Table 3.11, and Table 3.12. For these cases KMM performs very well in bias correction.

Table 3.11 and Table 3.12 show that KMM would be effective when non-response bias and skewed bid design coexist.

TABLE 3.10 Simulation Results for Skewed Bid Design

(N=100, M=48, $a = .1$, $b=0$, $\rho_1 = 0$, $\rho_2 = 10$, B=1000, $\sigma = .001$)

	WTP	variance
population	220.66	33486
sample	181.90	3868003
sample_s	182.01	3926720
KMM	196.33	2023790

	beta1	beta2	beta3	v1	v2	v3
population	1.54	-0.009039	0.0033	1.9653	1.66E-05	1.57E-04
sample	1.54	-0.009198	0.0038	4.3529	3.62E-05	3.50E-04
KMM	1.54	-0.009203	0.0038	4.3604	3.63E-05	3.51E-04

TABLE 3.11 Simulation Results for Skewed Bid Design

(N=200, M=104, $a = .1$, $b=0$, $\rho_1 = 0$, $\rho_2 = 10$, B=1000, $\sigma = .001$)

	WTP	variance
population	230.18	339953
sample	4088.65	14892005300
sample_s	4113.38	15084564020
KMM	217.56	12519

	beta1	beta2	beta3	v1	v2	v3
population	1.485	-0.008863	0.003592	0.9816	8.66E-06	7.56E-05
sample	1.494	-0.008973	0.003714	1.9229	1.54E-05	1.50E-04
KMM	1.497	-0.008986	0.003717	1.9305	1.55E-05	1.51E-04

TABLE 3.12 Simulation Results for Skewed Bid Design

(N=200, M=104, $a = .1$, $b = .1$, $\rho_1 = 0$, $\rho_2 = 10$, B=1000, $\sigma = .001$)

	WTP	variance
population	230.18	339953
sample	189.20	259443
sample_s	188.52	269422
KMM	207.86	71310

	beta1	beta2	beta3	v1	v2	v3
population	1.485	-0.008863	0.003592	0.9816	8.66E-06	7.56E-05
sample	1.505	-0.008943	0.003539	1.8979	1.54E-05	1.47E-04
KMM	1.506	-0.008939	0.003528	1.9318	1.57E-05	1.49E-04

TABLE 3.13 Simulation Results for Skewed Bid Design

(N=400, M=222, $a = .1$, $b=0$, $\rho_1 = 0$, $\rho_2 = 10$, B=1000, $\sigma = .0001$)

	WTP	variance
population	208.82	175
sample	210.16	755
sample_s	210.16	755
KMM	210.00	728

	beta1	beta2	beta3	v1	v2	v3
population	1.42	-0.00867	0.0039	0.4653	4.13E-06	3.73E-05
sample	1.45	-0.00884	0.0038	0.8456	7.71E-06	6.91E-05
KMM	1.45	-0.00885	0.0038	0.8447	7.70E-06	6.92E-05

3.4.2.5 Misspecification

To simulate misspecifications, we use

$$y_{i2} = 1.4136 - 0.008561bid_i + 0.00372\sqrt{x_{2i}} + \varepsilon_{i2} \quad (69)$$

To generate the data, and use the linear form (63) for estimation. The goal is to identify the best fit linear approximation model. The true WTP mean would be 169.5 per equation (69). Bid vector is adjusted to {70, 110, 140, 175, 205} to eliminate the effect of skewed bid design.

Two sets of simulation results are obtained: one with non-response bias, the other without. By comparing WTP estimates in Table 3.13 and Table 3.14, we see that non-response bias can be magnified severely when misspecification is present. More specifically, the bias of the beta3 parameter estimate increases significantly when there are non-responses correlated with x_2 . KMM does not correct the beta3 estimate, but it can correct the WTP estimate, especially when non-response bias is also present.

TABLE 3.14 Simulation Results with Misspecification

(N=200, M=105, $a = .1$, $b = .1$, $\rho_1 = 0$, $\rho_2 = 10$, B=1000, $\sigma = .005$)

	WTP	variance
population	170.77	192
sample	173.94	4138
sample_s	173.88	4029
KMM	172.39	597

	beta1	beta2	beta3	v1	v2	v3
population	1.48	-0.008770	1.01E-05	0.7985	4.14E-06	7.53E-05
sample	1.50	-0.008947	8.99E-05	1.6071	8.69E-06	1.49E-04
KMM	1.50	-0.008959	9.53E-05	1.6447	8.98E-06	1.51E-04

TABLE 3.15 Simulation Results with Misspecification

(N=200, M=104, $a = .1$, $b=0$, $\rho_1 = 0$, $\rho_2 = 10$, B=1000, $\sigma = .005$)

	WTP	variance
population	170.77	192
sample	172.52	527
sample_s	172.57	541
KMM	172.39	515

	beta1	beta2	beta3	v1	v2	v3
population	1.48	-0.008770	1.01E-05	0.7985	4.14E-06	7.53E-05
sample	1.48	-0.008909	2.84E-04	1.6398	8.50E-06	1.56E-04
KMM	1.48	-0.008926	3.07E-04	1.6718	8.78E-06	1.58E-04

3.5 Conclusion

Nonrandom sampling can bias parameter estimates as well as WTP estimates. The Monte Carlo study shows that KMM can correct non-response bias in WTP estimates. However, KMM is less effective in correcting sample selection bias and cannot always correct bias in parameter estimates. KMM will perform the best when sampling bias is prominent. There are several sampling bias tests that can be applied before deciding whether bias correction is needed (Vella, 1992).

There are Matlab packages of KMM application for several different kernels, which are available on request. However, further exploration is needed in calibrating the upper bound $w_i \in [0, W]$ and the kernel parameter σ to ensure the KMM can perform effectively.

In real world applications, variations of bid design, model misspecification, and data instability can magnify or counteract the sampling bias. What we simulate in skewed bid design and misspecification are two simplified cases for such variations. One risk associated with such variations is to neglect a more important bias source. As in 3.6.4.4, the sampling bias becomes trivial compared with the bias caused by skewed bid design. Another risk is the complication of interactions between different bias sources. As in 3.6.4.5, the covariate shift has a multiplier effect on the bias of linear approximation. The use of bias correction techniques needs to be cautiously validated and benchmarking methods should be applied for affirmation.

REFERENCES

- Cortes, C., Mohri, M., Riley, M., & Rostamizadeh, A. (2008). *Sample selection bias correction theory*. Paper presented at the International Conference on Algorithmic Learning Theory.
- Creel, M., & Loomis, J. (1997). Semi-nonparametric Distribution-Free Dichotomous Choice Contingent Valuation. *Journal of Environmental Economics and Management*, 32(3), 341-358. doi:<http://dx.doi.org/10.1006/jeem.1997.0972>
- Criado, C. O., & Veronesi, M. (2013). Parametric vs nonparametric dichotomous choice contingent valuation models: testing the kernel estimator and its revealed performance. *work*.
- Dalecki, M. G., Whitehead, J. C., & Blomquist, G. C. (1993). Sample non-response bias and aggregate benefits in contingent valuation: an examination of early, late and non-respondents. *Journal of Environmental Management*, 38(2), 133-143.
- Dubin, J. A., & Rivers, D. (1989). Selection bias in linear regression, logit and probit models. *Sociological Methods & Research*, 18(2-3), 360-390.
- DuMouchel, W. H., & Duncan, G. J. (1983). Using sample survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association*, 78(383), 535-543.
- Edwards, S. F., & Anderson, G. D. (1987). Overlooked biases in contingent valuation surveys: Some considerations. *Land Economics*, 63(2), 168-178.
- Eklöf, J., & Karlsson, S. (1999). *Testing and correcting for sample selection bias in discrete choice contingent valuation studies* (No. 171). Stockholm School of Economics.

- Garcia, S., Harou, P., Montagné, C., & Stenger, A. (2009). Models for sample selection bias in contingent valuation: Application to forest biodiversity. *Journal of Forest Economics*, 15(1–2), 59-78. doi:<http://dx.doi.org/10.1016/j.jfe.2008.03.008>
- Gretton, A., Smola, A., Huang, J., Schmittfull, M., Borgwardt, K., & Schölkopf, B. (2009). Covariate shift by kernel mean matching. *Dataset shift in machine learning*, 3(4), 5.
- Haab, T. C. (1999). Nonparticipation or Misspecification? The Impacts of Nonparticipation on Dichotomous Choice Contingent Valuation. *Environmental and Resource Economics*, 14(4), 443-461.
- Hanemann, M., & Kanninen, B. (2001). 11 The Statistical Analysis of Discrete-Response CV Data147. *Valuing environmental preferences: theory and practice of the contingent valuation method in the US, EU, and developing countries*, 302.
- Heckman, J. (1990). Varieties of Selection Bias. *The American Economic Review*, 80(2), 313-318.
- Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models *Annals of Economic and Social Measurement*, Volume 5, number 4 (pp. 475-492): NBER.
- Heckman, J. J. (1977). Sample selection bias as a specification error (with an application to the estimation of labor supply functions): National Bureau of Economic Research Cambridge, Mass., USA.
- Huang, J.-C., Nychka, D. W., & Smith, V. K. (2008). Semi-parametric discrete choice measures of willingness to pay. *Economics Letters*, 101(1), 91-94.
doi:<http://dx.doi.org/10.1016/j.econlet.2008.06.010>

- Huang, J., Gretton, A., Borgwardt, K. M., Schölkopf, B., & Smola, A. J. (2006). *Correcting sample selection bias by unlabeled data*. Paper presented at the Advances in neural information processing systems.
- Kanninen, B. J. (1995). Bias in Discrete Response Contingent Valuation. *Journal of Environmental Economics and Management*, 28(1), 114-125.
doi:<http://dx.doi.org/10.1006/jeem.1995.1008>
- Kanninen, B. J., Kriström, B., & Han, B. (1993). Sensitivity of Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Valuation Models: Comment. *Land Economics*, 69(2), 199-202.
- Kriström, B. (1990). A Non-Parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies, 135.
- Loomis, J. B. (1987). Expanding Contingent Value Sample Estimates to Aggregate Benefit Estimates: Current Practices and Proposed Solutions. *Land Economics*, 63(4), 396-402.
- Morrison, M. (2000). Aggregation Biases in Stated Preference Studies. *Australian Economic Papers*, 39(2), 215.
- Shimodaira, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2), 227-244.
- Steinwart, I. (2002). Support Vector Machines are Universally Consistent. *Journal of Complexity*, 18(3), 768-791. doi:<http://dx.doi.org/10.1006/jcom.2002.0642>
- Stolzenberg, R. M., & Relles, D. A. (1997). Tools for Intuition about Sample Selection Bias and Its Correction. *American Sociological Review*, 62(3), 494-507.

- Sugiyama, M., Krauledat, M., & MÃžller, K.-R. (2007). Covariate shift adaptation by importance weighted cross validation. *Journal of Machine Learning Research*, 8(May), 985-1005.
- Vella, F. (1992). Simple tests for sample selection bias in censored and discrete choice models. *Journal of Applied Econometrics*, 7(4), 413-421.
- Watanabe, M. (2010). Nonparametric Estimation of Mean Willingness to Pay from Discrete Response Valuation Data, 1114.
- Watanabe, M., & Asano, K. (2009). Distribution-Free Consistent Estimation of Mean WTP in Dichotomous Choice Contingent Valuation. *Environmental & Resource Economics*, 44(1), 1-10.
- Whitehead, J. C., Groothuis, P. A., & Blomquist, G. C. (1993). Testing for non-response and sample selection bias in contingent valuation: Analysis of a combination phone/mail survey. *Economics Letters*, 41(2), 215-220.
- Winship, C., & Mare, R. D. (1992). Models for sample selection bias. *Annual review of sociology*, 18(1), 327-350.
- Winship, C., & Radbill, L. (1994). Sampling weights and regression analysis. *Sociological Methods & Research*, 23(2), 230-257.
- Yoo, S.-H., & Yang, H.-J. (2001). Application of Sample Selection Model to Double-Bounded Dichotomous Choice Contingent Valuation Studies. *Environmental and Resource Economics*, 20(2), 147-163.
- Yu, Y., & Szepesvári, C. (2012). Analysis of kernel mean matching under covariate shift. *arXiv preprint arXiv:1206.4650*.

Zadrozny, B. (2004). *Learning and evaluating classifiers under sample selection bias*. Paper presented at the Proceedings of the twenty-first international conference on Machine learning.