CORRECTING SAMPLING BIAS IN NON-MARKET VALUATION WITH KERNEL MEAN MATCHING

Rui Zhang

Department of Agricultural and Applied Economics

University of Georgia

rui.zhang25@uga.edu

Selected Paper prepared for presentation at the 2017 Agricultural & Applied Economics Association Annual Meeting, Chicago, Illinois, July 30 - August 1

Abstract

Non-response is common in surveys used in non-market valuation studies and can bias the parameter estimates and mean willingness to pay (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. We use a machine learning algorism Kernel Mean Matching (KMM) to produce resampling weights in a non-parametric manner. We test KMM's performance through Monte Carlo simulations under multiple scenarios and show that KMM can effectively correct mean WTP estimates, especially when the sample size is small and sampling process depends on covariates. We also confirm KMM's robustness to skewed bid design and model misspecification.

Key Words: contingent valuation, Kernel Mean Matching, non-response, bias correction, willingness to pay

1. Introduction

Nonrandom sampling can bias the contingent valuation estimates in two ways. Firstly, when the sample selection process depends on the covariate, the *WTP* estimates are biased due to the divergence between the covariate distributions of the sample and the population, even the parameter estimates are consistent; 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 method selection can significantly affect the precision of the result. Yoo and Yang (2001) detect both sample selection bias and nonresponse bias in the double-bounded dichotomous choice contingent valuation survey and use a bivariate sample selection model.

This paper introduces the sample reweighting process technique kernel mean matching (KMM) to the 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* estimate 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 have 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.

2. Sampling Bias

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.

$$y_{i1} = \boldsymbol{X_{i1}\beta_1} + \varepsilon_{i1} \tag{1a}$$

$$y_{i2} = \boldsymbol{X_{i2}\beta_2} + \varepsilon_{i2} \tag{1b}$$

(1a) is the sample selection equation; (1b) is the *WTP* equation. y_{ij} is transformed into observable binary outcome Y_{ij} as

$$Y_{ij} = \begin{cases} 1, y_{ij} > 0 \\ 0, y_{ij} < 0 \end{cases} \; for\ j = 1,2. \tag{1c}$$

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 (1b) 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 (1b) can be estimated using a probit model. Here we use a simple form of (1b) for illustration.

$$y_{i2} = \alpha + \gamma * bid_i + \delta * x_i + \varepsilon_{i2} \tag{2a}$$

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 valuate, 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

$$\mathrm{E(WTP)} = -\frac{\alpha + \delta * \bar{x}}{\gamma} \tag{2b}$$

2.2 Decomposition of Sampling bias

2.2.1 Sample Selection 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 in (2a) 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).

2.2.2 Non-Response Bias

If $\boldsymbol{x}_{i1}$ and $\boldsymbol{x}_{i2}$ have common or correlated characteristic variables, nonresponses can cause divergence of distribution of such variables between of the sample and the population. Although the parameter estimates in (1b) will not be affected by common/correlated repressors, 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).

2.2.3 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 in 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 the preference the observant at the lower end. It would have the equal effect 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).

2.2.4 Misspecification

Habb (1999) use 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 function form of variables in $\boldsymbol{x}_{i2}$ is misspecified, and such variables are included in or correlated with $\boldsymbol{x_{i1}}$, the non-responses can further bias the parameter and *WTP* estimates. For example, if marginal utility of *x* is diminishing as in (2c)

$$y_{i2} = \alpha + \gamma * bid_i + \delta * x_i^{1/2} + \varepsilon_{i2} \tag{2c}$$

however, we use a linear form of "income" variable as in (2a) for estimation, 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 study 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. Kernel Mean Matching

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 independent 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 2.1) 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.

Let *X* denote the input space and *Y* the label set, which may be (Aadland, Caplan, & Phillips, 2007) in classification or a subset of *R* in regression estimation problems. Let D denote the population distribution from which we obtain a sample $\mathrm{z} = \{(x_1, y_1), \dots (x_n, y_n)\}$ that follows a sample distribution

D'. Let *s* denote the sample selection: when $\mathrm{s} = 1$ the data point is sampled; when $\mathrm{s} = 0$ 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 .

$$Pr[z_i] = \frac{Pr[z_i|s=1]\, Pr[s=1]}{Pr[s=1|z_i]} = \frac{Pr[s=1]}{Pr[s=1|z_i]} Pr'^{[z_i]} = w_i Pr'[z_i] \qquad (3a)$$

$$w_i = \frac{Pr[z_i]}{Pr'[z_i]} \qquad (3b)$$

Under covariate shift, the sampling probability is independent of *Y*, (3a) becomes

$$Pr[x_i] = \frac{Pr[s=1]}{Pr[s=1|x_i]} Pr'[z_i] = w_i Pr'[x_i] \qquad (3c)$$

$$w_i = \frac{Pr[x_i]}{Pr'[x_i]} \qquad (3d)$$

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).

3.2 An Empirical Kernel Meaning Matching Optimization Algorism

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), 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 $\mathrm{X} = (x_1, x_2, \dots x_m)$ and $\mathrm{X}' = (x_1', x_2', \dots x_n')$, 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 algorism can be summarized as a quadratic problem

$$\underset{w}{minimize}\ \frac{1}{2}w^T K w - \kappa^T w\ subject\ to\ w_i \in [0, W]\ and\ \left|\sum_{i=1}^{m} w_i - m\right| \leq m\epsilon \tag{4a}$$

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 an 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 ensure that $w(x)Pr(x)$ is close to a probability distribution. It is obvious that we need larger sample size to get reasonable convergence if Pr' and Pr are very different.

Suppose we have a mapping $\varphi: \mathbb{R}^p \to \mathbb{R}^q$ that brings a vector in $\mathbb{R}^p$ to $\mathbb{R}^q$. Then the dot product of $\boldsymbol{x}$ and $\boldsymbol{y}$ in space $\mathbb{R}^q$ is $\varphi(\boldsymbol{x})^{\boldsymbol{T}}\varphi(\boldsymbol{y})$. A kernel is a function that corresponds to this dot product $k(\boldsymbol{x}, \boldsymbol{y}) = \varphi(\boldsymbol{x})^{\boldsymbol{T}}\varphi(\boldsymbol{y})$. 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. (4a) is derived from its mapping function form as

$$\left\|\frac{1}{m}\sum_{i=1}^{m} w_i \varphi(x_i) - \frac{1}{n}\sum_{i=1}^{n} \varphi(x_i')\right\|^2 = \frac{1}{m^2} w^T K w - \frac{2}{m^2}\kappa^T w + constant \tag{4b}$$

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

$$k(x_i, x_j) = e^{-\frac{\|x_i - x_j\|}{2\sigma^2}} \tag{4c}$$

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. The larger the value of σ, the narrower will be the bell.

4. Monte Carlo Simulation: Probit Model

4.1 Data Generating Process

The data generating process follows the structure of the illustrative example in 2.1.

$$y_{i1} = a + b * x_{1i} + \varepsilon_{i1} \tag{5a}$$

$$y_{i2} = 1.4136 - 0.008561bid_i + 0.00372x_{2i} + \varepsilon_{i2} \tag{5b}$$

$$Y_{ij} = \begin{cases} 1, y_{ij} > 0 \\ 0, y_{ij} < 0 \end{cases} \quad for\ j = 1,2. \tag{5c}$$

where (5a) is sample selection equation and equation (5b) is the utility difference equation. We assume $(\varepsilon_{i1}, \varepsilon_{i2})$ is)$\sim N\left[[0, \quad 0]\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}\right]$, and $(x_{1i}, x_{2i})\sim N\left[[1, \quad 95.78]\begin{bmatrix} 1 & \rho_2 \\ \rho_2 & 119.226 \end{bmatrix}\right]$, 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.

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. $\rho_2$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 combination of *a, b,* ρ indicate different sampling bias decomposition. The population sizes used are $N = \{100, 200, 400, 800\}$ for estimation of (5b). *a* is arbitrarily assigned to ensure the selected sample is neither too small for estimation nor too large to be indifferent with the population.

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.

4.1.3 KMM Parameter Calibration

When applying KMM, the upper bound $w_i \in [0, W]$ and the kernel parameter σ in (4c) need to be calibrated to ensure the robustness of the reweight. A rule of thumb is that robust reweight should only change the weight slightly so that in each simulation the weight is "overfitting" and the results will not be too dispersed. The size of σ adjusts the trade-off between bias and variance of KMM estimators.

4.2 Simulation Results

Simulation results are obtained and compared under different scenarios. For each scenario, 1000 data sets are generated for simulation, average 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 parameter 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 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 additional to KMM, a bias correction estimate ("sample_s" rows in the tables) is obtained using 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.

4.2.1 Non-Response Bias

The Monte Carlo experiment shows that KMM can effectively increase the accuracy of *WTP* estimates especially when the non-response bias is large as shown in *Table 1.2*. The parameter estimates will be consistent when there is only non-response bias but no sample selection bias. KMM does not correct the parameter estimates itself and even increased 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 need to be multiplied with the biased sample mean of x_2, which can magnify the bias to an unacceptable scale as shown in *Table 1.2*. This is the situation when KMM become a powerful tool for bias correction. Replacing sample mean of x_2 with population mean has the similar bias correction effect as KMM, but is less efficient as shown in *Table 1.1, 1.2.*

TABLE 1.1 Simulation Results for Non-Response Bias[1]

(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

TABLE 1.2 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

[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 1.3 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 1.4 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

4.2.2 Sample Selection Bias

KMM can lower the variance of *WTP* when sample size is small, but cannot fix the inconsistency problem 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 unrelated to covariate shift need to be dealt with by other methods like bivariate probit (Eklöf and Karlsson, 1999). When *N=200* the convergence is already reached so we did not report the results at *N=400, N=800.*

TABLE 2.1 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 2.2 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

4.2.3 Sample Selection Bias Combined with Non-Response Bias

The results are similar with 4.2.1 when there is only Non-response bias except that there is a shift of sample selection bias. KMM can only improve the accuracy of *WTP* estimate, but not the parameter estimates.

TABLE 3.1 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 3.2 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

4.2.4 Skewed Bid Design

KMM can correct the bias caused by skewed bid design, even the variable on which KMM is used is not correlated with the bid variable, which is utterly randomly assigned in contingent valuation.

In the simulations, a repetition of bid vector {140, 205} is used which fail 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 4.2.* For these cases KMM perform very well in bias correction. *Table 4.3* verifies KMM's robustness to variation of bid design.

TABLE 4.1 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 4.2 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 4.3 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 4.4 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

4.2.5 Misspecification

To simulate misspecifications, we use

$$y_{i2} = 1.4136 - 0.008561 bid_i + 0.00372\sqrt{x_{2i}} + \varepsilon_{i2} \tag{6}$$

To generate the data, and use the linear form (5b) for estimation. The goal is to identify the best fit linear approximation model. The true *WTP* mean would be 169.5 per equation (6). 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 5.1* and *Table 5.2,* we see that non-response bias can be magnified severely when misspecification is present. More specifically, the bias of *beta3* parameter estimate increases significantly when there are non-responses correlated with x_2 . KMM does not correct *beta3* estimate, but it can correct *WTP* estimate, especially when non-response bias is also present.

TABLE 5.1 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 5.2 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

5. Discussion

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 to correct sample selection bias and cannot 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).

In real world applications, variations from 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 4.2.4, the sampling bias become trivial compared with the bias caused by skewed bid design. Another risk is the complication of interactions between different bias sources. As in 4.2.5, the covariate shift has a multiplier effect on the bias of linear approximation. The use of bias correction techniques need to be cautiously validated and benchmarking methods should be applied for affirmation.

Reference

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*. Retrieved from

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. doi:10.1023/a:1008349525868

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., Kristr, xf, & m, B. (1993). Sensitivity of Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Valuation Models: Comment. *Land Economics, 69*(2), 199-202. doi:10.2307/3146522

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. doi:10.2307/3146296

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. doi:10.2307/2657318

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. doi:10.1007/s10640-008-9255-3

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. doi:10.1023/a:1012625929384

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.