# Logistic Generalized Regression (LGREG) Estimator in Cluster Samples

Timothy L. Kennel*† Richard Valliant‡

**Abstract**

Statistical models are often used to assist estimation of descriptive statistics from surveys. Perhaps the most common estimator is the Generalized Regression Estimator (GREG) which is design consistent, uses a linear assisting model, and results in a set of calibrated weights. However, when the variable of interest is binary, binomial, or multinomial, it may be more appropriate to use a logistic assisting model instead of the standard linear model. In this paper we develop point and variance estimators for totals of finite population characteristics from a clustered sample assisted by a logistic regression model. Using a national Public Use Microdata set we compare the design-based properties of the new estimator to the GREG and the Horvitz-Thompson estimator under two clustered sample designs.

**Key Words:** Generalized Regression Estimator, Logistic Regression, Calibration, Clustered Sample Designs

## 1. Introduction

Data collected from surveys are often organized into discrete categories. Analyzing such categorical data from a complex survey often requires specialized techniques. In this paper, we describe one technique used to estimate the total of a categorical variable from a complex survey with clustering. The method we propose is an extension of the Generalized REGression estimator (GREG).

Generalized regression is a popular design-based method used in the production of descriptive statistics from survey data. Generalized regression is attractive because it results in a common set of weights that can be used for all variables in a dataset, estimated totals from the survey can be made to match known population controls, and often the sampling variance of an estimator is reduced through borrowing strength from a linear assisting model.

Although the GREG is design-consistent regardless of the form of the assisting model, the sampling error of the GREG is a function of the assisting model. In fact, assisting models that fit the data well generally result in estimators that have lower sampling variance than GREGs based on poorly fit assisting models.

In classical statistics, it is common to use logistic regression to model data with dichotomous outcomes. Such models are dominant because they assure that the predicted probability of an event is bounded between 0 and 1. Moreover, logistic models tend fit binary data better than models based on linear regression.

Thus, it seems natural to use a logistic assisting model instead of a linear assisting model when the variable of interest is categorical. Model-assisted estimators motivated by a logistic model have rarely been explored or used, despite their potential advantages over the GREG. In this paper, we aim to expand previous research by providing a new variance estimator for the Logistic GREG (LGREG) in single stage with-replacement samples.

*U. S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

†This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on methodological and operational issues are those of the author and not necessarily those of the U.S. Census Bureau.

‡University of Michigan, 1218 Lefrak Hall, College Park, MD 20742

Moreover, we also aim to explore point and variance estimators of the LGREG in clustered samples.

## 2. Previous Literature

Logistic regression is often used in the analysis of categorical data and has been well studied for decades from the model-based framework; however, very little has been written about how to use logistic regression models to estimate finite population quantities under complex sample designs. In this section, we highlight some of the main points of the model-based literature and discuss the small body of literature that uses logistic regression to assist finite population estimation. After motivating and defining logistic regression, we describe some of the technical details regarding how logistic regression models are built and how parameters are estimated from the model-based framework. Lastly, we briefly review how previous authors have suggested using logistic regression to assist estimating finite population quantities.

### 2.1 Introduction to Logistic Regression

Logistic regression is a popular method used to analyze binary, binomial, percent, and multinomial response data. It is widely used in medical and epidemiological studies, economics, survey methodology, and a host of other fields. Unlike linear regression, logistic regression is well suited to the analysis of binary and binomial data because predicted values are bounded, the interpretation of coefficients is closely linked to the odds ratio, and the variance of the observations does not need to be independent of the mean.

Numerous textbooks and papers devote attention to the model fitting, parameter estimation, and interpretation of logistic regression (see Agresti (2002), Bishop et al. (2007), McCullagh and Nelder (1999), Hosmer and Lemeshow (2000), Hilbe (2009), and Shao (2003)). All of these introductory texts focus on estimating superpopulation parameters, such as $\boldsymbol{\beta}$. However, none of them discuss how logistic regression can be used to make inference to descriptive statistics such as finite population totals, quartiles, and means.

Hilbe (2009)[p. 270] argues that the term *Logistic Regression* is used to describe several different kinds of models that can be characterized by the distribution of the response variable. In this paper, we provide results for binary logistic regression, binomial logistic regression, and multinomial logistic regression. Logistic regression is also often used to describe ordered categorical data, such as responses from Likert scales. The motivation, computation, and analysis of such data is different enough from the unordered cases that we do not consider ordered logistic regression in this paper.

### 2.2 Binary Logistic Regression

In binary logistic regression, the response variable is a Bernoulli random variable. Binary logistic regression is commonly used when the response can take one of two values. For example, it can be used to model the presence or absence of a disease, whether an elementary school student is proficient at math or not, whether a person has been a victim of a violent crime, whether a housing unit is vacant or not, whether a sample unit will respond to a survey request or not, and whether someone was satisfied with a product or not. In logistic regression, the response variable, $y_k$, can take on one of two values, usually written as $0$ for failure or $1$ for success.

### 2.3 Binomial Logistic Regression

In binomial logistic regression, the response variable, $y_k$, is a binomial random variable that can be any natural number from 0 to $z_k$. The binomial distribution is characterized by the number of successful events that occurred in a fixed number of independent trials. The total number of trials, $z_k$, can be different from one sample unit to another, but must be a known nonrandom quantity. For example, if school enrollment is fixed and known, the total number of students receiving a free or reduced lunch can be modeled with the binomial distribution. If the total number of mailable households in every Census tract is known, then the total number of households that would not participate in a mail census can be modeled by a binomial distribution. Binary logistic regression is a special case of binomial logistic regression when the total number of trials for all units in the population is fixed at 1.

### 2.4 Multinomial Logistic Regression

Multinomial logistic regression can be used to model a response vector where each element in the random vector is an independent Poisson random variable. The multinomial distribution is the joint distribution of all of the Poisson random variables conditional on the sum of the Poisson variables (McCullagh and Nelder (1999)).

Thus, one way to conceptualize the multinomial distribution is to consider $\mathtt{C}$ independent Poisson random variables. The response for the $k^{\text{th}}$ unit is a $\mathtt{C}$-valued column vector where each element of the random vector is a Poisson random variable, denoted $y_{k\mathtt{c}}$. The multinomial distribution is the multivariate distribution of $\mathbf{y}_k$ conditional on the sum of the Poisson random variables, $\sum_{\mathtt{c}=1}^{\mathtt{C}} y_{k\mathtt{c}} = z_k$.

An alternative way to conceptualize the multinomial distribution is to define $\mathtt{C}$ mutually exclusive and exhaustive categories indexed by the letter $\mathtt{c}$. Notice that $\mathtt{c}$ indexes categories; while $k$ indexes units. For the $k^{\text{th}}$ unit, we measure how many times $z_k$ items fall into each of the $\mathtt{c}$ categories. The result of this measurement is a $\mathtt{C}$-valued column vector for the $k^{\text{th}}$ unit, called $\mathbf{y}_k$.

Questions with multinomial outcomes are quite common in surveys. Questions where respondents must select one in a series of options can be modeled by a multinomial distribution. For example, the American Community Survey asks "Which FUEL is used MOST for heating this house, apartment, or mobile home?" followed by nine response options. In this case, $\mathtt{C} = 9$ and $z_k = 1$. Often $z_k$ is fixed to be 1 so that $\mathbf{y}_k$ is a vector with $\mathtt{C} - 1$ elements equal to 0 and exactly one element equal to 1.

In studying time usage, we could divide the day into minutes and categorize each minute into one of three categories: eating, sleeping, and other. In this case, the total number of minutes in the day is known and fixed at 1,440. If all respondents place the total number of minutes they spend in each category, then the vector of length three containing the hours spent in each category is an example of a multinomial random variable. One example of $\mathbf{y}_k$ is

$$\mathbf{y}_k = \begin{bmatrix} y_{eating,k} \\ y_{sleeping,k} \\ y_{other,k} \end{bmatrix} = \begin{bmatrix} 90 \\ 480 \\ 870 \end{bmatrix}$$

Since $z_k = \sum_{\mathtt{c}=1}^{\mathtt{C}} y_{k\mathtt{c}}$, one of the responses is usually removed from $\mathbf{y}_k$ to make it independent of $z_k$. Thus, $\mathbf{y}_k$ is often recoded to be a $\mathtt{C} - 1$ dimensional vector.

### 2.5 Logistic Model

The most important difference between linear regression and logistic regression is that in logistic regression a nonlinear transformation of the expected value of the response variable is related to explanatory variables while in linear regression the expected value of the observed response variable is linearly related to explanatory variables. The simple **linear** regression model for the $k^{\text{th}}$ unit can be written as

$$\mathscr{E}(y_k) = \mu_k = \mathbf{x}_k^\top \boldsymbol{\beta} \tag{1}$$

where $\mathbf{x}_k$ and $\boldsymbol{\beta}$ are $p$-dimensional vectors. For **binary** logistic regression, the corresponding regression model is

$$\mathscr{E}(y_k) = \mu_k = \frac{e^{\mathbf{x}_k^\top \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_k^\top \boldsymbol{\beta}}} \tag{2}$$

For **binomial** logistic regression, the corresponding regression model is

$$\mathscr{E}(y_k) = \mu_k = \frac{z_k e^{\mathbf{x}_k^\top \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_k^\top \boldsymbol{\beta}}} \tag{3}$$

For **multinomial** logistic regression, the corresponding regression model is

$$\mathscr{E}(\mathbf{y}_k) = \boldsymbol{\mu}_k = \frac{z_k e^{\mathbf{x}_k^\top \boldsymbol{\beta}}}{1 + \sum_{\mathtt{C}=1}^{\mathtt{C}-1} e^{\mathbf{x}_k^\top \boldsymbol{\beta}}} \tag{4}$$

In the multinomial case, $\boldsymbol{\beta}$ is a $p$ by $\mathtt{C}$ matrix and $\boldsymbol{\mu}_k$ is a row vector of length $\mathtt{C}$. Elementwise division is performed in (4). For example, suppose we wish to model the three time-usage categories with a model that includes an intercept and the person's age. In this case,

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_{intercept,eat} & \widehat{\beta}_{intercept,sleeping} & \widehat{\beta}_{intercept,other} \\ \widehat{\beta}_{age,eat} & \widehat{\beta}_{age,sleeping} & \widehat{\beta}_{age,other} \end{bmatrix}$$

In the design-based framework, inference is often made to finite population quantities instead of superpopulation parameters. Thus, the finite population models are identical to the models above, with the exception that $\boldsymbol{\beta}$ is replaced by $\mathbf{B}$, where $\mathbf{B}$ is the estimate of $\boldsymbol{\beta}$ that would be obtained if the entire finite population was in the sample.

### 2.6 Point Estimators of a Total

Surveys are often used to estimate totals of a finite population characteristic. One of the earliest and most studied estimators of a finite population total is the *Horvitz-Thompson estimator*,

$$\widehat{t}_y^{\pi} = \sum_s w_k y_k$$

where $s$ denotes the sample, $\pi_k$ is the probability of selecting unit $k$, $w_k = \frac{1}{\pi_k}$ denotes the sampling weights, $y_k$ is the characteristic of interest, and $k$ indexes units in the sample. Although the Horvitz-Thompson estimator is design-unbiased, it can be quite inefficient. Thus, estimates from a single sample may be far from the true value, especially if the

probabilities of selection are negatively correlated with the characteristic of interest (see Basu (1971) and Little (2004)).

From the model-based framework, sometimes the *projective estimator* is used. The projective estimator is the sum of predicted values for the complete population. That is,

$$\widehat{t}_y^{pro} = \sum_{\mathscr{U}} \widehat{\mu}_k$$

where $\widehat{\mu}_k$ is the fitted value from a model and $\mathscr{U}$ is the set of all population units. In general $\widehat{t}_y^{pro}$ is a model-based estimator and not design-consistent. However, under certain conditions and estimation procedures, Firth and Bennett (1998) show that $\widehat{t}_y^{pro}$ can be design-consistent. Valliant (1985) studied a closely related model-based estimator called the prediction estimator for binary regression.

The GREG is an alternative estimator that uses a model to assist design-based estimation. Särndal et al. (1992) discuss GREG estimators of the general form,

$$\widehat{t}_y^{g} = \widehat{t}_y^{pro} + \sum_{s} w_k e_{ks}$$

where $e_{ks} = y_k - \widehat{\mu}_k$. When the GREG is written in this form, we can easily see that the GREG is the projective estimator of the finite population total with a weighted adjustment based on residuals. Robinson and Särndal (1983) show that the GREG is design-consistent and asymptotically design-unbiased in single stage samples. Moreover, Särndal et al. (1992, p. 226) argue that the single-stage GREG often has lower variance than estimators that are not assisted by a model. In official statistics, the GREG is often used because it has the calibration property. That is, the weighted sum of the explanatory variables are forced to equal known population totals. Särndal (2007) reviews many of the advantages to using the GREG over design-based methods that are not assisted by a model.

If the response variable is a Bernoulli, binomial, or multinomial random variable, it is more natural to use a logistic assisting model than a linear assisting model. Lehtonen and Veijanen (1998) provide one design-consistent method that uses an assisting logistic model. Their estimator, called the LGREG, has not been developed in complex samples with clustering. In one stage of sampling, the LGREG for a binary response is written as

$$\widehat{t}_y^{LG} = \sum_{k=1}^{N} \widehat{\mu}_k + \sum_{k=1}^{n} w_k e_k \tag{5}$$

where $e_k = y_k - \widehat{\mu}_k$. If $\widehat{\mathbf{B}}$ is calculated using weighted pseudo maximum likelihood estimating equations, then $\widehat{t}_y^{LG}$ will be a design-consistent estimator of the population total under a variety of sample designs, including multiple stage samples. Since the first summation is over the entire universe, $\mathbf{x}_k$ must be known for all units in the population. For this reason, using the LGREG requires a sampling frame complete with all explanatory variables used in the assisting model for all units in the population. Many address-based sampling frames, business registers, and trade association lists contain a wealth of covariates.

### 2.7 Variance Estimator of LGREG in Unclustered Samples

Lehtonen and Veijanen (1998) recommend estimating the variance of $\widehat{t}_y^{LG}$ with

$$v_{LV} = \sum\sum_{s} \frac{\Delta_{kl}}{\pi_{kl}} \left(\frac{e_k}{\pi_k}\right) \left(\frac{e_l}{\pi_l}\right)$$

where $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$. However, this simple variance estimator will generally underestimate the sampling error in clustered samples because it does not account for the correlation between clusters. Moreover, in small samples, it may poorly estimate the variability of $\widehat{t}_y^{LG}$ because it estimates the asymptotic variance of $\widehat{t}_y^{LG}$ rather than the exact variance of $\widehat{t}_y^{LG}$. The variance estimator proposed by Lehtonen and Veijanen (1998) also requires knowledge of joint inclusion probabilities, which often are impossible to compute or unavailable to data analysts.

## 3. New Estimators

### 3.1 Alternative Variance Estimators of LGREG in Unclustered Samples

Commonly, with-replacement variance estimators are used even when the first stage sample is selected without-replacement. As long as the sampling fraction is relatively small, the bias of using a with-replacement variance estimator is relatively small. Furthermore, any bias in the with-replacement variance estimator tends to be positive, thus making the with-replacement variance estimator conservative. Särndal et al. (1992, sec 4.6) discuss the classic with-replacement variance estimator of a total and provide some limitations for using the with-replacement variance estimator for samples selected without-replacement. For estimating the variance of the Horvitz-Thompson estimator, the with-replacement variance estimator is

$$\upsilon_{wr,\pi} = \frac{1}{n\,(n-1)} \sum_{k=1}^{n} \left( \frac{y_k}{p_k} - \widehat{t}_y^{\pi} \right)^2 \tag{6}$$

where $p_k = \frac{\pi_k}{n}$ is the probability of drawing the $k^{\text{th}}$ unit in single draw and $n$ is the total number of sample units. We can modify (6) for the LGREG by replacing $\frac{y_k}{p_k}$ with $\widehat{t}_{yk}^{LG} = \sum_{\mathscr{U}} \widehat{\mu}_k + \frac{e_k}{p_k}$ and $\hat{t}_y^{\pi}$ by $\hat{t}_y^{LG} = \frac{1}{n} \sum_{k=1}^{n} \hat{t}_{yk}^{\pi}$ which equals the LGREG in (5). For the LGREG, the with-replacement variance estimator is

$$\begin{aligned}
\upsilon_{wr} &= \frac{1}{n\,(n-1)} \sum_{k=1}^{n} \left[ \sum_{k=1}^{N} \widehat{\mu}_k + \frac{e_k}{p_k} - \left( \sum_{k=1}^{N} \widehat{\mu}_k + \sum_{k=1}^{n} w_k e_k \right) \right]^2 \\
&= \frac{1}{n\,(n-1)} \sum_{k=1}^{n} \left( \frac{e_k}{p_k} - \widehat{t}_e^{\pi} \right)^2 \\
&= \frac{n}{(n-1)} \sum_{k=1}^{n} \left( \frac{e_k}{\pi_k} - \widehat{\bar{e}} \right)^2
\end{aligned}$$

where

$$\begin{aligned}
\widehat{t}_e^{\pi} &= \sum_{k=1}^{n} w_k e_k \\
\widehat{\bar{e}} &= \frac{\widehat{t}_e^{\pi}}{n}
\end{aligned}$$

A second alternative variance estimator uses implicit differentiation. First described by Binder (1983), implicit differentiation uses linearization and estimating equations to produce design-consistent estimators of finite population parameters. Implicit differentiation is especially useful when the parameter of interest cannot be solved explicitly in closed form. Both Binder (1983) and Särndal et al. (1992)[section 13.4] give several examples

of how implicit differentiation can be used to construct design-consistent estimators of $\mathbf{B}$ from a logistic regression model. An advantage of implicit differentiation is that variance estimators can easily be computed from the estimating equations,

$$\widehat{\mathbf{W}}(\boldsymbol{\theta}) = \mathbf{0} \tag{7}$$

where

$$\underset{(q+1)\times 1}{\boldsymbol{\theta}} = \begin{bmatrix} t_y^{LG} \\ \mathbf{B} \end{bmatrix} \tag{8}$$

and

$$\underset{(q+1)\times 1}{\widehat{\mathbf{W}}_k(\boldsymbol{\theta})} = \begin{bmatrix} w_k(y_k - \mu_k) - \left(t_y^{LG} - \sum_{k=1}^{N} \mu_k\right) \\ w_k(y_k - \mu_k) x_{1k} \\ \vdots \\ w_k(y_k - \mu_k) x_{qk} \end{bmatrix} \tag{9}$$

$$\underset{(q+1)\times 1}{\widehat{\mathbf{W}}(\boldsymbol{\theta})} = \sum_{k=1}^{n} \widehat{\mathbf{W}}_k(\boldsymbol{\theta}) \tag{10}$$

The value of $\boldsymbol{\theta}$ that solves the estimating equations, $\widehat{\mathbf{W}}(\boldsymbol{\theta})$, is denoted $\widehat{\boldsymbol{\theta}}$.

Simultaneously solving for $\mathbf{B}$ and $t_y^{LG}$ has the advantage that it simplifies variance estimation. Moreover, it results in the complete covariance matrix containing the estimated covariances between $\hat{t}_y^{LG}$ and $\widehat{\mathbf{B}}$. The variance estimator obtained from implicit differentiation has the form,

$$v\left(\widehat{\boldsymbol{\theta}}\right) = \left[\widehat{\mathbf{J}}^{-1}\left(\widehat{\boldsymbol{\theta}}\right)\right]\left[\widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{\theta}}\right)\right]\left[\widehat{\mathbf{J}}^{-1}\left(\widehat{\boldsymbol{\theta}}\right)\right]^{\top}$$

where

$$\widehat{\mathbf{J}}\left(\widehat{\boldsymbol{\theta}}\right) = \frac{\partial}{\partial \widehat{\boldsymbol{\theta}}} \widehat{\mathbf{W}}\left(\widehat{\boldsymbol{\theta}}\right)$$

and $\widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{\theta}}\right)$ is an estimate of the design-based variance of $\widehat{\mathbf{W}}$. Assuming with-replacement sampling,

$$\widehat{\boldsymbol{\Sigma}} = \frac{n}{n-1} \sum_{k=1}^{n} \left[\widehat{\mathbf{W}}_k - \overline{\widehat{\mathbf{W}}}\right]\left[\widehat{\mathbf{W}}_k - \overline{\widehat{\mathbf{W}}}\right]^{\top}$$

The variance estimator for $\hat{t}_y^{LG}$ is the element in the upper left-hand corner of $v\left(\widehat{\boldsymbol{\theta}}\right)$.

For example, suppose we wish to estimate the total number of students in the country who are receiving a free or reduced lunch. We formulate an assisting binomial logistic model with the poverty rate of the neighborhood around the school as a covariate. Our theoretical model for school $k$ is,

$$\widehat{\mu}_k = \frac{z_k e^{\widehat{\alpha} + x_k \widehat{B}}}{1 + e^{\widehat{\alpha} + x_k \widehat{B}}}$$

where $z_k$ is the school enrollment and $x_k$ is the poverty rate around the school. We have three parameters to estimate,

$$\boldsymbol{\theta} = \begin{bmatrix} t_y^{LG} \\ \alpha \\ B \end{bmatrix}$$

and the estimating equations are

$$\widehat{\mathbf{W}}\left(\widehat{\boldsymbol{\theta}}\right) = \begin{bmatrix} \sum_{k=1}^{n} \left\{ w_k \left( y_k - \frac{z_k e^{\hat{\alpha}+\hat{B}x_k}}{1+e^{\hat{\alpha}+\hat{B}x_k}} \right) - \left( \widehat{t}_y^{LG} - \sum_{k=1}^{N} \frac{z_k e^{\hat{\alpha}+\hat{B}x_k}}{1+e^{\hat{\alpha}+\hat{B}x_k}} \right) \right\} \\ \sum_{k=1}^{n} w_k \left( y_k - \frac{z_k e^{\hat{\alpha}+\hat{B}x_k}}{1+e^{\hat{\alpha}+\hat{B}x_k}} \right) \\ \sum_{k=1}^{n} w_k \left( y_k - \frac{z_k e^{\hat{\alpha}+\hat{B}x_k}}{1+e^{\hat{\alpha}+\hat{B}x_k}} \right) x_k \end{bmatrix}$$

The Jacobian of the estimating equations is

$$\widehat{\mathbf{J}}\left(\widehat{\boldsymbol{\theta}}\right) = \sum_{k=1}^{n} \begin{bmatrix} -1 & -w_k \frac{z_k e^{\hat{\alpha}+\hat{B}x_k}}{\left(1+e^{\hat{\alpha}+\hat{B}x_k}\right)^2} + \sum_{k=1}^{N} \frac{z_k e^{\hat{\alpha}+\hat{B}x_k}}{\left(1+e^{\hat{\alpha}+\hat{B}x_k}\right)^2} & -w_k \frac{z_k e^{\hat{\alpha}+\hat{B}x_k}}{\left(1+e^{\hat{\alpha}+\hat{B}x_k}\right)^2} x_k + \sum_{k=1}^{N} \frac{z_k e^{\hat{\alpha}+\hat{B}x_k}}{\left(1+e^{\hat{\alpha}+\hat{B}x_k}\right)^2} x_k \\ 0 & -w_k \frac{z_k e^{\hat{\alpha}+\hat{B}x_k}}{\left(1+e^{\hat{\alpha}+\hat{B}x_k}\right)^2} & -w_k \frac{z_k e^{\hat{\alpha}+\hat{B}x_k}}{\left(1+e^{\hat{\alpha}+\hat{B}x_k}\right)^2} x_k \\ 0 & -w_k \frac{z_k e^{\hat{\alpha}+\hat{B}x_k}}{\left(1+e^{\hat{\alpha}+\hat{B}x_k}\right)^2} x_k & -w_k \frac{z_k e^{\hat{\alpha}+\hat{B}x_k}}{\left(1+e^{\hat{\alpha}+\hat{B}x_k}\right)^2} x_k^2 \end{bmatrix}$$

### 3.2 LGREG in Clustered Samples

In clustered samples, the LGREG is

$$\begin{aligned} \widehat{t}_{y,II}^{LG} &= \sum_{i=1}^{M} \sum_{k=1}^{N_i} \widehat{\mu}_{ik} + \sum_{i=1}^{m} \sum_{k=1}^{n_i} w_{ik} \left( y_{ik} - \widehat{\mu}_{ik} \right) \\ &= \sum_{k=1}^{N} \widehat{\mu}_k + \sum_{k=1}^{n} w_k \left( y_k - \widehat{\mu}_k \right) \end{aligned}$$

where $M$ is the total number of clusters in the population, $m$ is the total number of clusters in the sample, $N_i$ is the number of elements in cluster $i$, $n_i$ is the number of sample elements in cluster $i$, $n = \sum_{i=1}^{m} n_i$, and $N = \sum_{i=1}^{M} N_i$.

### 3.3 Variance Estimators of LGREG in Clustered Samples

The with-replacement variance estimator for a cluster sample is similar to the with-replacement variance estimator for unclustered samples. The one exception is that weighted cluster totals are used instead of unit responses. For the LGREG, the with-replacement variance estimator is

$$v_{wr,II} = \frac{m}{(m-1)} \sum_{i=1}^{m} \left( \widehat{e}_i - \widehat{\overline{e}} \right)^2$$

where

$$\begin{aligned} e_k &= y_k - \widehat{\mu}_k \\ \widehat{e}_i &= \sum_{k=1}^{n_i} \frac{e_k}{\pi_k} \\ \widehat{\overline{e}} &= \frac{1}{m} \sum_{i=1}^{m} \widehat{e}_i \end{aligned}$$

The implicit variance estimator for clustered samples is similar to the implicit variance estimator for unclustered samples. In fact $\widehat{\mathbf{J}}\left(\widehat{\boldsymbol{\theta}}\right)$ is the same for both estimators. The key difference is that $\widehat{\boldsymbol{\Sigma}}\left(\boldsymbol{\theta}\right)$ must be estimated with respect to the sample design. Assuming a with-replacement sample of clusters gives

$$\widehat{\boldsymbol{\Sigma}} = \frac{m}{m-1}\sum_{i=1}^{m}\left[\widehat{\mathbf{W}}_i - \overline{\widehat{\mathbf{W}}}\right]\left[\widehat{\mathbf{W}}_i - \overline{\widehat{\mathbf{W}}}\right]^{\top}$$

where

$$\widehat{\mathbf{W}}_i = \sum_{k=1}^{n_i}\widehat{\mathbf{W}}_k$$

That is, $\widehat{\mathbf{W}}_i$ is the sum of $\widehat{\mathbf{W}}_k$ over all sample units in cluster $i$. Recall that $\widehat{\mathbf{W}}_k$ was previously defined in (9).

## 4. Methodology

### 4.1 Pseudo-Populations

We conducted two sets of simulations to test how the LGREG and the two new variance estimators performed in clustered samples. In the first set of simulations, we evaluated how the LGREG and the variance estimators performed in large samples under ideal conditions. The second set of simulations were designed to evaluate the LGREG and the variance estimators under more realistic conditions.

### 4.2 Ideal Population

For the first set of simulations, we generated a clustered population of binary, binomial, and multinomial random variables.

First we generated $M = 30{,}000$ clusters of size $N_i = 11+\lambda_i$ where $\lambda_i$ is a random draw from an exponential distribution with parameter 0.25. To assure that $N_i$ was an integer, we rounded $\lambda_i$ to the nearest whole number. Overall, the pseudo population contained $N = 450{,}265$ units.

Next, we generated our auxiliary variable using a hierarchial process to simulate a clustering effect. For each unit, we created an auxiliary variable using the model $x_k = \delta_i + \varepsilon_k$ where $\delta_i$ was a draw for the $i^{th}$ cluster from a standard normal distribution and $\varepsilon_k$ was a draw from a normal distribution with mean of 0 and a standard deviation of 0.1.

Using the explanatory variable, we generated random response variables. For the binary response, we drew a random number from a Bernoulli distribution with a parameter of $\pi_k = \frac{e^{.5+3x_k}}{1+e^{.5+3x_k}}$. For the binomial response variable, we drew a random number from a binomial distribution with a probability of success $\pi_k$ in $z_k = 10 + \lambda_k$ trials where $\lambda_k$ was a draw from an exponential distribution with parameter of 0.01. To assure that $z_k$ was an integer, we rounded $\lambda_k$ to the nearest whole number. Lastly, for the multinomial response, we generated a random vector of length 3 using the `rmultinomial()` function from the mc2d package in R. The probabilities generating the multinomial random vector were set to be: $\pi_{1k} = \frac{e^{.5+3x_k}}{1+e^{.5+3x_k}+e^{-.5+2x_k}}$, $\pi_{2k} = \frac{e^{-.5+2x_k}}{1+e^{.5+3x_k}+e^{-.5+2x_k}}$, and $\pi_{3k} = 1 - (\pi_{1k} + \pi_{2k})$. The sum of the three random elements was set to be $z_k$.

### 4.3 Realistic Population

The second pseudo-population was derived from Census 2000 data. We downloaded Census 2000 housing unit and population data from Summary File 3 for California, Florida, and New York from the US Census Bureau's website. We then subset the data to block groups with at least one occupied housing unit and one person. Furthermore, all "orphan" tracts, tracts containing only one valid block group, were removed. We read in the following variables: Total Number of Occupied Housing Units in the block group (H007001), Total Number of Housing Units being rented in the block (H007003), and the Percent of persons living at or below the poverty line (P088002 + P088003 + P088004) / P088001.

We used the dataset to estimate the total number of rental housing units in California, Florida, and New York. If one has a complete address list of housing units for these three states, one can use the LGREG to improve estimation over the basic Horvitz-Thompson estimator.

The motivating sample design is to select tracts in the first stage of sampling. Then, within sample tracts, a set of block groups is selected. The block group is treated as the ultimate sampling unit. A survey is then conducted within the sample block groups to determine the total number of rental units in each sample block group.

Overall, this population has 13,135 primary sampling units (tracts) and 44,032 ultimate sampling units (block groups).

We used a binomial logistic regression assisting model to estimate the total number of rental units in California, Florida, and New York. The assisting model contained an intercept and the percent of persons living at or below the poverty line in each block group.

### 4.4 Sample Design

We used the `UPrandomsystematic()` function in the `sampling` package of R to select all the samples (Tillé and Matei 2009). This function selects a randomized systematic sample by sorting the population into a random order and then selecting a sample with probabilities proportional to a size measure.

We tested how the LGREG preformed under two realistic sample designs: simple random sampling without replacement ($srswor$) and probabilities proportional to size sampling ($\pi ps$). For the $srswor$ design, we first selected a simple random sample of clusters without replacement. From the list of sample clusters, we selected a simple random sample of units. Because the clusters varied in size, the $srswor$ design resulted in unequal probabilities of selection at the unit level. For the $\pi ps$ design, we selected clusters with probabilities proportional to the number of elements in each cluster. We selected the first stage sample without replacement. Within each cluster, we selected a simple random sample without replacement of units. Such a sample design is common in area frame sampling and results in a sample of units with equal probabilities of selection.

In the ideal population, we selected a sample of 1,500 clusters. From each cluster, a sample of 2 units were selected. In the realistic population, we selected two different samples; one design to investigate the LGREG in large samples, and the other to investigate the LGREG in smaller samples. The first sample contained 1,500 clusters; while the second sample contained 20 clusters. From the sample cluster, two units were randomly selected. We selected 2,000 samples from each of the sample designs. Table 1 summarizes the different designs used to select the samples.

**Table 1**: Simulation Design for LGREG

| Population | First Stage Sample Size | Second Stage Sample Size | First Stage Sample Design | Samples Selected |
|---|---|---|---|---|
| Ideal | 1,500 | 2 | $srswor$ | $2,000$ |
| Ideal | 1,500 | 2 | $\pi ps$ | $2,000$ |
| Realistic | 1,500 | 2 | $srswor$ | $2,000$ |
| Realistic | 1,500 | 2 | $\pi ps$ | $2,000$ |
| Realistic | 20 | 2 | $srswor$ | $2,000$ |
| Realistic | 20 | 2 | $\pi ps$ | $2,000$ |

### 4.5 Estimation

We estimated the total of each response variable using the Horvitz-Thompson estimator, the GREG, and the LGREG. The one exeception is that we did not calculate the GREG for the multinomial response variable, because there was no clear multivariate extension to the GREG. Using the with-replacement estimator and implicit differentiation, we estimated the variance of the LGREG. Altogether, we estimated the statistics in Table 2. We repeated this process for all samples.

**Table 2**: Statistics of interest for Simulation

| Statistic | Description |
|---|---|
| $\hat{t}_y^{\pi}$ | Horvitz-Thompson Estimator |
| $\hat{t}_y^{g}$ | GREG |
| $\hat{t}_y^{LG}$ | LGREG |
| $\upsilon_{wr}\left(\hat{t}_y^{LG}\right)$ | With Replacement Variance Estimator of $\hat{t}^{LG}$ |
| $\upsilon_{I}\left(\hat{t}_y^{LG}\right)$ | Implicit Differentiation Variance Estimator of $\hat{t}^{LG}$ |

We used the `lm()` function in R with a weights option to predict the fitted values which we used in the GREG estimation.

To compute the LGREG, we first estimated $\boldsymbol{\beta}$, the superpopulation parameter associated with the assisting model, using the `glm()` function in R. Then, we used the value of $\widehat{\boldsymbol{\beta}}$ as a starting point to minimize the logistic pseudo-log likelihood. Table 3 shows the pseudo-maximum log-likelihood equations that were used to estimate $\mathbf{B}$. These estimating equations were solved numerically using the `optim()` function in R. One advantage of using the `optim()` function was that the numerical hessian, a major component of the implicit differentiation variance estimator, was automatically calculated. The solution to the pseudo-maximum likelihood equations was noted. Table 3 shows both the sample pseudo log-likelihood estimating equations as well as the derivative of them for the three different cases of logistic regression.

**Table 3**: Logistic Regression Estimating Equations

| Distribution of Response | Sample Pseudo Log Likelihood $\widehat{L}(\mathbf{B})$ | Gradient $\widehat{\ell}(\mathbf{B})$ |
|---|---|---|
| Ber $(\mathfrak{p}_k)$ | $\sum_s w_k \left[ y_k \left( \mathbf{x}_k^{\top} \mathbf{B} \right) - \ln \left( 1 + e^{\mathbf{x}_k^{\top} \mathbf{B}} \right) \right]$ | $\sum_s w_k \left( y_k - \frac{e^{\mathbf{x}_k^{\top} \mathbf{B}}}{1+e^{\mathbf{x}_k^{\top} \mathbf{B}}} \right) \mathbf{x}_k$ |
| Bin $(\mathfrak{p}_k; z_k)$ | $\sum_s w_k \left[ y_k \left( \mathbf{x}_k^{\top} \mathbf{B} \right) - z_k \ln \left( 1 + e^{\mathbf{x}_k^{\top} \mathbf{B}} \right) \right]$ | $\sum_s w_k \left( y_k - z_k \frac{e^{\mathbf{x}_k^{\top} B}}{1+e^{\mathbf{x}_k^{\top} B}} \right) \mathbf{x}_k$ |
| MN $(\mathfrak{p}_k; z_k)$ | $\sum_s w_k \left[ \mathbf{y}_k^{\top} \left( \mathbf{X}_k^{\top} \mathbf{B} \right) - z_k \ln \left( 1 + e^{\mathbf{X}_k^{\top} \mathbf{B}} \right) \right]$ | $\sum_s w_k \left( \mathbf{y}_k - z_k \frac{e^{\mathbf{X}_k^{\top} \mathbf{B}}}{1+e^{\mathbf{X}_k^{\top} \mathbf{B}}} \right) \mathbf{X}_k$ |

### 4.6 Measures

To compare the point estimators, we calculated the relative empirical bias, coefficient of variation, and the relative root empirical mean squared error for the point estimators. The relative root empirical mean squared error is,

$$\widetilde{\mathrm{RRMSE}}\left(\widehat{t}_y^{LG}\right) = 100 \cdot \frac{\sqrt{\frac{1}{2{,}000}\sum_{\nu=1}^{2{,}000}\left(\widehat{t}_{y\nu}^{LG} - t\right)^2}}{t}$$

where $t$ is the true population total and $\nu$ indexes the 2,000 simulation runs. We assessed how the design-based empirical mean squared error of the LGREG compared to the design-based empirical mean squared error of the Horvitz-Thompson estimator and the GREG using a linear assisting model containing the same explanatory variables as the logistic assisting model.

We also compared the two variance estimators to the empirical variance. To summarize the variance estimators, we provided descriptive statistics about the distribution of the estimated variances over the 2,000 simulation runs.

## 5. Results

Table 4 shows the results from the simulations for the ideal population.

**Table 4**: Summary of LGREG, GREG, and HT Point Estimators of Totals for Ideal Population

| Response Variable | Design | Estimator | Relative Bias | CV | Relative Root MSE |
|---|---|---|---|---|---|
| Binary | srs | LGREG | 0.0 | 1.2 | 1.2 |
| Binary | srs | GREG | 0.0 | 1.3 | 1.3 |
| Binary | srs | HT | 0.0 | 2.2 | 2.2 |
| Binary | pps | LGREG | 0.0 | 1.1 | 1.1 |
| Binary | pps | GREG | 0.0 | 1.2 | 1.2 |
| Binary | pps | HT | 0.0 | 2.0 | 2.0 |
| Binomial | srs | LGREG | 0.0 | 0.1 | 0.1 |
| Binomial | srs | GREG | 0.0 | 2.2 | 2.2 |
| Binomial | srs | HT | 0.0 | 2.8 | 2.8 |
| Binomial | pps | LGREG | 0.0 | 0.1 | 0.1 |
| Binomial | pps | GREG | -0.1 | 2.1 | 2.1 |
| Binomial | pps | HT | -0.2 | 2.6 | 2.6 |
| Multinomial category 1 | srs | LGREG | 0.0 | 2.4 | 2.4 |
| Multinomial category 1 | srs | HT | 0.0 | 2.8 | 2.8 |
| Multinomial category 2 | srs | LGREG | 0.0 | 2.3 | 2.3 |
| Multinomial category 2 | srs | HT | 0.0 | 2.4 | 2.4 |
| Multinomial category 1 | pps | LGREG | 0.1 | 2.2 | 2.2 |
| Multinomial category 1 | pps | HT | 0.0 | 2.6 | 2.6 |
| Multinomial category 2 | pps | LGREG | 0.1 | 2.2 | 2.2 |
| Multinomial category 2 | pps | HT | 0.1 | 2.2 | 2.2 |

* Numbers are in percents

Scanning down the relative bias column reveals that all of the estimators appear to be unbiased in large samples where the model holds nearly perfectly. In terms of variability, we see that the coefficient of variation for the LGREG tends to be smaller than the coefficient of variation for the GREG as well as the Horvitz-Thompson estimator. Thus, we see that the LGREG has potential to be much more efficient than the GREG and the Horvitz-Thompson estimator. The LGREG clearly outperforms the Horvitz-Thompson estimator and the GREG for the Binomial response variable. In the case of the Binary response variable, the LGREG also outperforms the Horvitz-Thompson estimator; but is similar to

the GREG. Lastly, in the case of the Multinomial response, the LGREG is similar to the Horvitz-Thompson estimator. Table 4 shows that there are situations where the LGREG can outperform the GREG and the Horvitz-Thompson estimator.

Table 5 also shows that the bias of the LGREG tends to be negligible in realistic situations. In fact, we found that the empirical bias of the LGREG is always less than one percent of the true value. Although we did find a small relative bias of -0.6 percent for the LGREG in small srs samples, we see that this bias tends to disappear as the number of sample clusters increases. Under a very simple model, containing only one covariate and an intercept, we see clear benefits to the LGREG over the GREG and the Horvitz-Thompson estimator.

**Table 5**: Summary of LGREG, GREG, and HT Point Estimators for Census Population

| Design | Sample Clusters | Estimator | Relative Bias | CV | Relative Root MSE |
|---|---|---|---|---|---|
| srs | 20 | LGREG | -0.6 | 12.9 | 12.9 |
| srs | 20 | GREG | 1.8 | 20.9 | 21.0 |
| srs | 20 | HT | 0.2 | 21.6 | 21.6 |
| pps | 20 | LGREG | -0.4 | 13.1 | 13.1 |
| pps | 20 | GREG | 1.3 | 20.6 | 20.6 |
| pps | 20 | HT | 0.2 | 20.9 | 20.9 |
| srs | 1,500 | LGREG | 0.0 | 1.5 | 1.5 |
| srs | 1,500 | GREG | 0.0 | 2.2 | 2.2 |
| srs | 1,500 | HT | 0.0 | 2.4 | 2.4 |
| pps | 1,500 | LGREG | 0.0 | 1.4 | 1.4 |
| pps | 1,500 | GREG | 0.1 | 2.2 | 2.2 |
| pps | 1,500 | HT | 0.1 | 2.3 | 2.3 |

* Numbers are in percents

Moreover, we also see major efficiency gains of the LGREG over the GREG and Horvitz-Thompson estimator from Table 5. This is especially true for the small samples, but there are also gains for larger samples as well. Although these efficiency gains are not guaranteed, it is promising to note that the LGREG has the potential to outperform the GREG by leaps and bounds. If careful attention is put into building the assisting model, the benefits to using the LGREG can be great, as seen by the large reductions in mean squared error.

In addition to calculating the LGREG, GREG, and Horvitz-Thompson estimators for each simulation, we also computed the with-replacement and implicit differentiation variance estimators for the LGREG. Table 6 summarizes the estimators for the ideal population.

In general, the implicit differentiation and with-replacement variance estimators have similar distributions. On average, both the implicit differentiation and with-replacement variance estimators tend to be close to the empirical variance. Moreover, the confidence interval coverage is close to the expected 95 percent for all cases. Thus, when the number of clusters is large and the model fits reasonably well, both the implicit differentiation and with-replacement variance estimators are about the same.

Table 7 summarizes the distribution of the LGREG variance estimators for the realistic population. For the small samples, the mean and median of the variance estimators tend to be within 31 percentage points of the empirical variance. For example, half of the implicit differentiation estimators for the srs sample underestimated the empirical variance by 25 percentage points or more. The with-replacement estimator was even worse because half of the samples underestimated the empirical variance by 29 percentage points or more. One good aspect of the two variance estimators is that the interquartile range of the estimators contain the empirical value, at least for the smaller sample. Estimating variance is challenging when the sample size is small. As we see, both variance estimators are highly variable when the number of clusters is small. In looking at the minimum and maximum estimated

**Table 6**: Summary Statistics for the Variance Estimator as a Percent of the Empirical Variance $\left(\frac{\text{variance estimator}}{\text{empirical variance}}\right)$ for Ideal Population

| Population | Design | Variance Estimator | Minimum | Quartile 1 | Median | Mean | Quartile 3 | Maximum | Coverage |
|---|---|---|---|---|---|---|---|---|---|
| Binary | srs | Binder | 0.80 | 0.93 | 0.97 | 0.97 | 1.00 | 1.21 | 94.35 |
| Binary | srs | wr | 0.79 | 0.93 | 0.97 | 0.97 | 1.00 | 1.16 | 94.25 |
| Binary | pps | Binder | 0.87 | 0.98 | 1.01 | 1.01 | 1.04 | 1.17 | 94.70 |
| Binary | pps | wr | 0.86 | 0.98 | 1.01 | 1.01 | 1.04 | 1.19 | 94.75 |
| Binomial | srs | Binder | 0.79 | 0.95 | 1.00 | 1.00 | 1.05 | 1.39 | 94.65 |
| Binomial | srs | wr | 0.77 | 0.95 | 0.99 | 1.00 | 1.04 | 1.38 | 94.55 |
| Binomial | pps | Binder | 0.81 | 0.97 | 1.01 | 1.02 | 1.05 | 1.26 | 94.75 |
| Binomial | pps | wr | 0.82 | 0.97 | 1.01 | 1.01 | 1.05 | 1.26 | 94.70 |
| Multinomial category 1 | srs | Binder | 0.69 | 0.91 | 0.96 | 0.97 | 1.03 | 2.73 | 94.15 |
| Multinomial category 1 | srs | wr | 0.71 | 0.91 | 0.96 | 0.97 | 1.02 | 3.02 | 94.20 |
| Multinomial category 1 | pps | Binder | 0.78 | 0.97 | 1.03 | 1.03 | 1.09 | 2.45 | 95.55 |
| Multinomial category 1 | pps | wr | 0.75 | 0.97 | 1.02 | 1.03 | 1.08 | 2.63 | 95.40 |
| Multinomial category 2 | srs | Binder | 0.74 | 0.92 | 0.97 | 0.98 | 1.03 | 1.89 | 95.00 |
| Multinomial category 2 | srs | wr | 0.74 | 0.92 | 0.97 | 0.98 | 1.03 | 4.07 | 95.10 |
| Multinomial category 2 | pps | Binder | 0.76 | 0.98 | 1.02 | 1.03 | 1.07 | 5.44 | 95.40 |
| Multinomial category 2 | pps | wr | 0.77 | 0.98 | 1.02 | 1.03 | 1.07 | 3.07 | 95.30 |

variances, we see that some samples can produce variance estimates that are less than 90 percent of what they should be; while, others are more than eight times what they should be. Of course, such extreme estimates are rare. The large variability of estimates presents many opportunities for further research.

**Table 7**: Summary Statistics for the Variance Estimator as a Percent of the Empirical Variance $\left(\frac{\text{variance estimator}}{\text{empirical variance}}\right)$ for Simulations of the Census Population

| Design | Sample Clusters | Variance Estimator | Minimum | Quartile 1 | Median | Mean | Quartile 3 | Maximum | Coverage |
|---|---|---|---|---|---|---|---|---|---|
| srs | 20 | Binder | 0.13 | 0.53 | 0.75 | 0.88 | 1.08 | 6.26 | 90.75 |
| srs | 20 | wr | 0.11 | 0.49 | 0.71 | 0.92 | 1.11 | 8.35 | 90.15 |
| pps | 20 | Binder | 0.09 | 0.52 | 0.73 | 0.84 | 1.04 | 4.28 | 90.60 |
| pps | 20 | wr | 0.09 | 0.46 | 0.69 | 0.87 | 1.04 | 6.53 | 90.30 |
| srs | 1,500 | Binder | 0.83 | 1.02 | 1.07 | 1.08 | 1.14 | 1.49 | 95.85 |
| srs | 1,500 | wr | 0.77 | 1.02 | 1.07 | 1.08 | 1.14 | 1.52 | 95.70 |
| pps | 1,500 | Binder | 0.86 | 1.05 | 1.11 | 1.11 | 1.16 | 1.44 | 95.85 |
| pps | 1,500 | wr | 0.82 | 1.05 | 1.11 | 1.11 | 1.17 | 1.53 | 95.80 |

The variability of the variance estimates appears to decrease as the sample size increases. The minimum and maximum values of the variance estimates are much closer to the empirical values than the samples with only 20 clusters. Furthermore, the confidence interval coverage is much closer to 95 percent with the larger sample size. Although the with-replacement and implicit differentiation variance estimators tend to overestimate the empirical variance in large samples, some slight overestimation leads to conservative inference.

## 6. Conclusion

In this paper, we constructed two new variance estimators for the LGREG in single stage samples. We also formulated the LGREG in cluster samples and constructed two variance estimators for it. Both variance estimators assume sampling with replacement and may not be desirable for all samples.

We used a simulation to empirically test how the LGREG performed in cluster samples.

We found that it can be more efficient than the GREG and the Horvitz Thompson estimators. We also compared the two new variance estimators for the LGREG in cluster samples to the empirical variance. On average, we found that the variance estimators tended to be close to the empirical variance; however, estimates from individual samples may be much larger or smaller than what they should be, especially if the number of clusters is small.

In summary, the LGREG has the potential to outperform the GREG and the Horvitz-Thompsom estimators. The preliminary simulation showed that the LGREG is worth further research. Estimating the variance of the LGREG is difficult and careful attention should be given to this topic in the future.

## References


Agresti, A. (2002), *Categorical Data Analysis*, John Wiley & Sons.

Basu, D. (1971), "An essay on the logical foundations of survey sampling, Part I," *Foundations of Statistical Inference*, 203–242.

Binder, D. A. (1983), "On the Variances of Asymptotically Normal Estimators from Complex Surveys," *International Statistical Review / Revue Internationale de Statistique*, 51, 279–292.

Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (2007), *Discrete multivariate analysis*, New York: Springer Verlag.

Firth, D. and Bennett, K. E. (1998), "Robust Models in Probability Sampling," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60, 3–21.

Hilbe, J. M. (2009), *Logistic regression models*, New York: Chapman & Hall/CRC Press.

Hosmer, D. W. and Lemeshow, S. (2000), *Applied logistic regression*, New York: Wiley-Interscience.

Lehtonen, R. and Veijanen, A. (1998), "Logistic generalized regression estimators," *Survey Methodology*, 24, 51–55.

Little, R. J. (2004), "To model or not to model? Competing modes of inference for finite population sampling," *Journal of the American Statistical Association*, 99, 546–556.

McCullagh, P. and Nelder, J. A. (1999), *Generalized linear models. (Monographs on Statistics and Applied Probability)*, Boca Raton: Chapman and Hall/CRC.

Robinson, P. M. and Särndal, C.-E. (1983), "Asymptotic properties of the generalized regression estimator in probability sampling," *Sankhyā Ser. B*, 45, 240–248.

Särndal, C.-E. (2007), "The calibration approach in survey theory and practice," *Survey Methodology*, 33, 99 – 119.

Särndal, C.-E., Swensson, B., and Wretman, J. H. (1992), *Model assisted survey sampling*, Springer Series in Statistics, New York: Springer-Verlag.

Shao, J. (2003), *Mathematical Statistics*, New York: Springer Verlag.

Tillé, Y. and Matei, A. (2009), *sampling: Survey Sampling*, r package version 2.3.

Valliant, R. (1985), "Nonlinear Prediction Theory and the Estimation of Proportions in a Finite Population," *Journal of the American Statistical Association*, 80, 631–641.