# ESTIMATING SOME MEASURES OF INCOME INEQUALITY FROM SURVEY DATA:
## An application of the estimating equation approach

David A. Binder and Milorad S. Kovačević, Statistics Canada, Ottawa K1A 0T6
David A. Binder, Statistics Canada, Ottawa K1A 0T6

**Keywords:** *Gini family index, Lorenz curve ordinate, low income measure, quantile share, Taylor linearization, variance estimation*

**Abstract** We give a brief review of the definitions of certain measures of income inequality: ordinates of the Lorenz curve, income shares, the family of Gini coefficients, and a low income measure. We summarize some salient aspects of the theory of estimation for finite populations. In particular, we discuss the problem of estimation of means and totals and extend this theory to estimating functions. We then apply this estimating function framework to the problem of estimating complex statistics, such as measures of income inequality and their mean squared errors, for a wide class of survey designs used in practice.

**1. Introduction** In general, a population distribution can be described by its cumulative distribution function, $F(y) = Pr\{Y \le y\}$, where $Y$ is the random variable corresponding to selecting one population unit at random. Throughout this paper, we assume that $Y$ is non-negative. If $Y$ represents income then we are interested in the properties of an income distribution, such as income concentration, income shares for different population segments, low income proportions, etc. We may be interested in the quantile function $\xi(p) = F^{-1}(p) = inf\{y \mid F(y) \ge p\}$, as well.

The Lorenz curve, for example, depicts the cumulative income against the population share. The formal definition of the ordinate of the Lorenz curve corresponding to the $100p$-th percentile of the population is

$$L(p) = \frac{1}{\mu_Y}\int_0^{\xi_p} y\,dF(y),$$

where $\int_0^{\xi_p} dF(y) = p$, and $\int_0^{\infty} y\,dF(y) = \mu_Y$. (1.1)

The income (quantile) share is defined as the percentage of total income shared by the population allocated to a quantile interval $[\xi_{p_1}, \xi_{p_2})$. It is equal to the difference of Lorenz curve ordinates

$$Q(p_1, p_2) = L(p_2) - L(p_1).$$

The formal definition of the Gini coefficient is

$$G = 1 - 2\int_0^1 L(p)\,dp = \frac{1}{\mu}\int_0^{\infty}[2F(y) - 1]\,y\,dF(y)$$

$$= \frac{1}{2\mu}\int_0^{\infty}\int_0^{\infty}|x - y|\,dF(x)\,dF(y).$$

We see from this expression that the Gini coefficient is related to the mean absolute difference between two units selected independently at random from the population.

A more general family of Gini coefficients, given in Nygård and Sandström (1981) is

$$G_J = \frac{1}{\mu_Y}\int_0^{\infty} J[F(y)]\,y\,dF(y). \quad (1.2)$$

For the usual Gini coefficient, $J(p) = 2p - 1$.

Another measure of income inequality is the Low Income Measure used by some economists. It is defined as the proportion of the population units whose income is less than half the median income for the population. Formally, this is

$$\theta = \int_0^{M/2} dF(y), \quad (1.3a)$$

where $M$ is the median defined by

$$\int_0^{M} dF(y) = \frac{1}{2}. \quad (1.3b)$$

For all these measures, we can write the parameter of interest, $\theta_o$, as the solution to

$$\int u(y,\theta_o)\,dF(y) = 0,$$

where $u(y,\theta)$ is the kernel of the estimating equation. This estimating equation formulation will be discussed in Section 2. In Sections 3, 4, and 5 we give the estimating equations for the above measures along with the approximation of their mean squared error estimates. In Section 6 we present estimators of these measures based on the complex sample design. Section 7 contains an illustration based on the Canadian Survey of Consumer Finance data.

**2. Use of Estimating Equations for Finite Populations** The theory for estimating means and totals from finite populations is now well established in the statistical literature. A formulation which encompasses most estimators used in practice is given in Särndal, Swennson, and Wretman (1992). We extend this slightly by incorporating the theory given in Rao (1979). In this section, we briefly review this theory and show how it can be applied to more complex statistics through the use of estimating equations, as described by Binder (1991).

Let the population total, $Y$, be defined as

$$Y = N\int y\,dF(y).$$

Note here that $F(y)$ is a step function corresponding to the distribution function for the finite population. We consider estimators of the form:

$$\hat{Y} = \sum_{i \in s} w_i(s) y_i = \sum_{i=1}^{N} w_i(s) Y_i,$$

where $w_i(s)$ is zero whenever the $i$-th unit is not in the sample.

If there exist constants, $c_1,\ldots,c_N$, such that $MSE(\hat{Y}) = E(\hat{Y}-Y)^2$ becomes zero when $y_i \propto c_i$, then denoting $\tilde{y}_i = y_i/c_i$, Rao (1979) has shown that

$$MSE(\hat{Y}) = -\frac{1}{2}\sum_{\substack{i\neq j\\ i,j=1}}^{N}\sum \Delta_{ij}(\tilde{y}_i-\tilde{y}_j)^2, \qquad (2.1)$$

where $\Delta_{ij} = 2c_ic_j\,E\{[w_i(s)-1][w_j(s)-1]\}$. A non-negative unbiased quadratic estimator of $MSE(\hat{Y})$ is necessarily of the form

$$mse(\hat{Y}) = -\frac{1}{2}\sum_{\substack{i\neq j\\ i,j=1}}^{N}\sum \check{\Delta}_{ij}(s)(\tilde{y}_i-\tilde{y}_j)^2, \qquad (2.2)$$

where $\sum_{s\ni i,j} p(s)\check{\Delta}_{ij}(s) = \Delta_{ij}$, $i<j$, and $p(s)$ is the probability function for the sample $s$.

We now review how this theory can be extended to estimating equations, as described by Binder (1991). An estimator for the distribution function is given by

$$N\,\hat{F}(y) = \sum_{i\in s} w_i(s)I(y_i\le y),$$

where

$$I(y_i\le y) = \begin{cases} 1 & \text{if } y_i\le y,\\ 0 & \text{if } y_i>y.\end{cases}$$

We note that $\hat{F}(y)$ is unbiased for $F(y)$, but it is not necessarily a true distribution function, unless

$$\sum_{i\in s} w_i(s) = N.$$

Suppose parameter $\theta_o$ is defined as the solution to

$$\int u(y,\theta_o)dF(y) = 0.$$

We define the estimating equation for $\theta_o$ as that value of $\hat{\theta}$ for which

$$\int \hat{u}(y,\hat{\theta})d\hat{F}(y) = 0, \qquad (2.3)$$

where $\hat{u}(y,\theta)$ is an estimate of $u(y,\theta)$.

We can rewrite (2.3) as

$$\begin{aligned} 0 &= \int \hat{u}(y,\hat{\theta})d\hat{F}(y)\\ &= \int[\hat{u}(y,\hat{\theta})-u(y,\theta_o)]dF(y) + \int u(y,\theta_o)d\hat{F}(y) + R, \end{aligned} \qquad (2.4)$$

where

$$R = \int[\hat{u}(y,\hat{\theta})-u(y,\theta_o)][d\hat{F}(y)-dF(y)].$$

The decomposition in (2.4) is the basic starting point for all the derivations of variance in the remainder of this paper. We will be assuming throughout that the remainder term, $R$, is asymptotically negligible.

Binder (1983) considered the case where $\hat{u}(y,\theta) = u(y,\theta)$ and where, for large samples,

$$\int[u(y,\hat{\theta})-u(y,\theta_o)]dF(y) = (\hat{\theta}-\theta_o)\left.\frac{\partial E\{u(y,\theta)\}}{\partial\theta}\right|_{\theta=\theta_o}.$$

Using these approximations, we have

$$\begin{aligned} \hat{\theta}-\theta_o &= -\left[\left.\frac{\partial E\{u(y,\theta)\}}{\partial\theta}\right|_{\theta=\theta_o}\right]^{-1}\int u(y,\theta_o)d\hat{F}(y)\\ &= \int u^*(y)d\hat{F}(y) \end{aligned}$$

where

$$u^*(y) = -\left[\left.\frac{\partial E\{u(y,\theta)\}}{\partial\theta}\right|_{\theta=\theta_o}\right]^{-1}u(y,\theta_o).$$

Once we have obtained the expression for $u^*(y)$, the derivation of the variance of $\hat{\theta}$ becomes straight-forward. Since we have approximated $\hat{\theta}-\theta_o$ as an estimator of a population total of $u^*(y_i)$'s, we can use the mean squared error calculations given by (2.1) and (2.2) above.

For example, for population quantiles, we have

$$\begin{aligned} u &= I(y\le\theta) - p,\\ u^* &= -\frac{1}{f(\theta)}[I(y\le\theta) - p], \end{aligned} \qquad (2.5)$$

which is an extension of the Bahadur representation for sample quantiles, as described by Francisco and Fuller (1991). Result (2.5) will be used for the ordinates of the Lorenz curve and for the Low Income Measure, which are discussed in Sections 4 and 5.

**3. Gini Family Coefficient** For the Gini family coefficient, given by (1.2), we can use

$$u(y,G_J) = J[F(y)]y - G_J y.$$

Ignoring the remainder term in (2.4), we have the following approximation:

$$\begin{aligned} 0 &= \int\{J[\hat{F}(y)]y-\hat{G}_J y\}d\hat{F}(y)\\ &= \int\{J[\hat{F}(y)]-J[F(y)]\}y\,dF(y) - (\hat{G}_J-G_J)\int y\,dF(y)\\ &\quad + \int\{J[F(y)]y-G_J y\}d\hat{F}(y). \end{aligned}$$

Letting

$$\int \{J[\hat{F}(y)] - J[F(y)]\}\, y\, dF(y)$$

$$= \int [\hat{F}(y) - F(y)] J'[F(y)] y\, dF(y)$$

and

$$\int \hat{F}(y) J'[F(y)] y\, dF(y) = \int\int_0^y J'[F(y)] y\, d\hat{F}(x) dF(y)$$

$$= \int \left[ \int_y^{\infty} J'[F(x)] x\, dF(x) \right] d\hat{F}(y),$$

we have that

$$\hat{G}_J - G_J = \int u^*(y) d\hat{F}(y),$$

where

$$u^* = \frac{1}{\mu_Y} \left[ \int_{F(y)}^1 J'(p) F^{-1}(p)\, dp + J[F(y)] y - G_J y - \mathrm{E}\{F(y) J'[F(y)] y\} \right] \tag{3.1}$$

For the case of independent and identically distributed observations, this yields the same variance result as described by Glasser (1962) and Sendler (1979). To estimate the variance, it is necessary to use estimates for $\mu_Y$, $F(y)$, and $G_J$ in the expression for $u^*$.

**4. Lorenz Curve Ordinate and Quantile Share** The ordinate of the Lorenz curve was defined in (1.1). In terms of estimating equations, the following two equations are required:

$$u_1(y, L(p)) = I\{y \le \xi_p\} y - L(p) y,$$

$$u_2(y) = I\{y \le \xi_p\} - p.$$

The second equation defines the $100p$-th percentile of the distribution; whereas the first equation defines the ordinate of the Lorenz curve in terms of the $100p$-th percentile. Ignoring the remainder term in (2.4), we have the following approximation:

$$0 = \int [I\{y \le \hat{\xi}_p\} - \hat{L}(p)] y\, d\hat{F}(y)$$

$$= \int_{\xi_p}^{\hat{\xi}_p} y\, dF(y) - [\hat{L}(p) - L(p)] \int y\, dF(y) + \int [I\{y \le \xi_p\} - L(p)] y\, d\hat{F}(y)$$

The first term of this expression can be further approximated as

$$\int_{\xi_p}^{\hat{\xi}_p} y\, dF(y) = (\hat{\xi}_p - \xi_p) \xi_p f(\xi_p),$$

and from (2.5) we see that

$$\xi_p - \xi_p = -\int \frac{1}{f(\xi_p)} [I\{y \le \xi_p\} - p] d\hat{F}(y), \tag{4.1}$$

so that

$$(\hat{\xi}_p - \xi_p) \xi_p f(\xi_p) = -\int \xi_p [I\{y \le \xi_p\} - p] d\hat{F}(y).$$

Therefore, to estimate the variance of the ordinate of the Lorenz curve, the appropriate linearization is given by using

$$u^*(y) = \frac{1}{\mu_Y} [(y - \xi_p) I\{y \le \xi_p\} + p \xi_p - y L(p)].$$

This yields the same result as described by Beach and Davidson (1983) for variances and covariances of ordinates of the Lorenz curve in the case of independent and identically distributed random variables. To estimate the variance it is necessary to use $\hat{\xi}_p$ and $\hat{L}(p)$ in the expression for $u^*(y)$.

To estimate the quantile share $Q(p_1, p_2)$ we need three equations

$$u_1(y, Q(p_1, p_2)) = I\{\xi_{p_1} < y \le \xi_{p_2}\} y - Q(p_1, p_2) y,$$

$$u_2(y) = I\{y \le \xi_{p_1}\} - p_1,$$

$$u_3(y) = I\{y \le \xi_{p_2}\} - p_2.$$

Using the same arguments as before, we arrive to

$$u^*(y) = \frac{1}{\mu_Y} [(y - \xi_{p_2}) I\{y \le \xi_{p_2}\} - (y - \xi_{p_1}) I\{y \le \xi_{p_1}\} + p_2 \xi_{p_2} - p_1 \xi_{p_1} - y Q(p_1, p_2)]$$

**5. Low Income Measure** The Low Income Measure was defined in (1.3). In terms of estimating equations, the following two equations are required:

$$u_1(y, \theta) = I\left\{y \le \frac{M}{2}\right\} - \theta,$$

$$u_2(y) = I\{y \le M\} - \frac{1}{2}.$$

The second equation defines the median of the distribution; whereas the first equation defines the Low Income Measure in terms of the median. Ignoring the remainder term in (2.4), we have the following approximation:

$$0 = \int \left( I\left\{y \le \frac{\hat{M}}{2}\right\} - \hat{\theta} \right) d\hat{F}(y)$$

$$\approx \frac{1}{2}(\hat{M} - M) f\left(\frac{M}{2}\right) - (\hat{\theta} - \theta) + \int \left( I\left\{y \le \frac{M}{2}\right\} - \theta \right) d\hat{F}(y).$$

Using result (4.1) to substitute for $\hat{M} - M$, and solving for $\hat{\theta} - \theta$, we obtain

$$\hat{\theta} - \theta = \int u^*(y) d\hat{F}(y),$$

where

$$u^* = -\frac{f\left(\frac{M}{2}\right)}{2f(M)} \left( I\{y \le M\} - \frac{1}{2} \right) + I\left\{y \le \frac{M}{2}\right\} - \theta. \tag{5.1}$$

The problem with applying this result to estimate the variance of the estimated Low Income Measure is that it is necessary to estimate $f(M)$ and $f(M/2)$. To accomplish this, we could use

$$\hat{f}(\xi) = \frac{\hat{F}\left(\xi + \frac{h}{2}\right) - \hat{F}\left(\xi - \frac{h}{2}\right)}{h},$$

for some suitably small $h$. Alternatively, we could perform the following calculations, as suggested by Francisco and Fuller (1991) for another problem. For a given value of p, we estimate the corresponding percentile, $\xi$. We then construct the Woodruff interval for that percentile. This is determined by first solving for $h_1$ and $h_2$ in

$$\inf_{h_1}\left[\frac{\int [I(y \le \xi - h_1) - p]\, d\hat{F}(y)}{\left[mse\left\{\int [I(y \le \xi) - p]\, d\hat{F}(y)\right\}\right]^{1/2}} \le -z_{1-\frac{\alpha}{2}}\right],$$

$$\inf_{h_2}\left[\frac{\int [I(y \le \xi + h_2) - p]\, d\hat{F}(y)}{\left[mse\left\{\int [I(y \le \xi) - p]\, d\hat{F}(y)\right\}\right]^{1/2}} \ge z_{1-\frac{\alpha}{2}}\right],$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$–th percentile from the standard normal distribution. Then we compute

$$\hat{f}(\xi) = \frac{2z_{1-\frac{\alpha}{2}}\left[mse\left\{\int [I(y \le \xi) - p]\, d\hat{F}(y)\right\}\right]^{1/2}}{h_1 + h_2}. \qquad (5.2)$$

This calculation uses the asymptotic equivalence of $\hat{\xi} - \xi$ and the estimated sum of the $u^*(y)$'s given by (2.5).

We see that the estimated variance for the Low Income Measure may be somewhat complex to compute. The estimating function framework has provided us with the appropriate formulae.

**6. Estimation with a Complex Survey** Let us assume a stratified multistage design with a considerably large number of strata, $H$, with a few primary sampling units (clusters), $n_h (\ge 2)$, sampled from each stratum. Let $w_{hci}$ be the normalized weight attached to the $i$–th ultimate unit in the $c$–th cluster of the $h$–th stratum. The appropriate estimator of the population mean and the consistent estimator of its mean squared error are

$$\hat{\mu} = \sum_s w_{hci} y_{hci}$$

$$mse(\hat{\mu}) = \sum_h \frac{n_h}{n_h - 1} \sum_c (u^*_{hc} - \bar{u}^*_h)^2 \qquad (6.1)$$

where $u^*_{hc} = \sum_i w_{hci}(y_{hci} - \hat{\mu})$ and $\bar{u}^*_h = \frac{1}{n_h}\sum_c u^*_{hc}$. We use $\sum_s = \sum_h \sum_c \sum_i$ to denote summation over all ultimate units in the sample incorporating all stages of sampling. We assumed that PSU's are selected with replacement.

An estimator of the finite population distribution function is

$$\hat{F}(y) = \sum_s w_{hci} I(y_{hci} \le y)$$

The consistent estimator of the approximation of the mean squared error of the distribution function estimated in y takes the form (6.1) where $u^*_{hc} = \sum_i w_{hci}\left[I(y_{hci} \le y) - \hat{F}(y)\right]$ and $\bar{u}^*_h = \frac{1}{n_h}\sum_c u^*_{hc}$.

The usual estimate of the finite population quantile is a sample quantile

$$\hat{\xi}_p = \inf_{hci \in s} \{y_{hci} : \hat{F}(y_{hci}) \ge p\}$$

which is indeed the solution of the estimation equation

$$\sum_s w_{hci}[I(y_{hci} \le \xi_p) - p] = 0$$

Accordingly, using result (2.5), the mean squared error of the quantile has form (6.1) with

$$u^*_{hc} = \frac{1}{\hat{f}(\hat{\xi}_p)} \sum_i w_{hci}[I(y_{hci} \le \hat{\xi}_p) - p].$$

If the expression (5.2) is used for the estimation of the density function $f(\xi)$, the *MSE* of the quantile $\hat{\xi}_p$ is estimated as

$$mse_a(\hat{\xi}_p) = \left[\frac{D_a(\xi_p)}{z_{1-\alpha/2}}\right]^2 \qquad (6.2)$$

where $D_a(\xi_p) = \frac{1}{2}(h_1 + h_2) = \frac{1}{2}(\hat{\xi}_U - \hat{\xi}_L)$ is the half length of the $100(1-\alpha)\%$ confidence interval for the $\xi_p$ and $h_1$ and $h_2$ are obtained as solutions for

$$\hat{\xi}_L = \xi_p - h_1 = \inf_{y_{hci}} \{\hat{F}(y_{hci}) \ge p - z_{1-\alpha/2}\sqrt{mse\{\hat{F}(\hat{\xi}_p)\}}$$

$$\hat{\xi}_U = \xi_p + h_2 = \inf_{y_{hci}} \{\hat{F}(y_{hci}) \ge p + z_{1-\alpha/2}\sqrt{mse\{\hat{F}(\hat{\xi}_p)\}}$$

The estimator (6.2) was also used by Francisco and Fuller (1991). Generally speaking the motivation for (5.2) and consequently for (6.2) comes from Woodruff's (1952) confidence interval for individual quantiles. Francisco and Fuller (1986) and Rao and Wu (1987) used these intervals to derive variance estimators. Though the estimator depends on the confidence coefficient, they showed that it is asymptotically consistent for any significance level $\alpha$. Rao and Wu (1987) studied the standard errors of quantiles for the cluster samples estimated in this manner. Their Monte Carlo results suggest that 95% confidence interval works well as a basis for extracting the standard error. Binder (1991) obtained a similar form of the variance estimator by using the estimating equations approach.

The estimate of the usual Gini coefficient is the solution of the following estimation equation

$$\sum_s w_{hci}\{[2F(y_{hci}) - 1]y_{hci} - G y_{hci}\} = 0$$

and takes the form

$$\hat{G} = \frac{2}{\hat{\mu}} \sum_s \hat{F}(y_{hci}) w_{hci} y_{hci} - 1$$

where $\hat{\mu} = \sum_s w_{hci} y_{hci}$.

The estimate of the *MSE* of the Gini coefficient can be computed using expression (6.1) by substituting $u^*_{hc}$, originally defined by (3.1), with the complex survey form obtained as

$$u^*_{hc} = \frac{2}{\hat{\mu}} \sum_s w_{hci} \left[ A(y_{hci}) y_{hci} + B(y_{hci}) - \frac{\hat{\mu}}{2}(\hat{G}+1) \right]$$

where

$$A(y) = \hat{F}(y) - \frac{\hat{G}+1}{2} \text{ and}$$

$$B(y) = \sum_s w_{hci} y_{hci} I\{y_{hci} \geq y\}.$$

The Lorenz curve ordinates could be obtained by solving a system of estimating equations

$$\sum_s w_{hci} [I\{y_{hci} \leq \xi_p\} y_{hci} - L(p) y_{hci}] = 0$$

$$\sum_s w_{hci} [I\{y_{hci} \leq \xi_p\} - p] = 0$$

The resulting estimate is

$$\hat{L}(p) = \frac{1}{\hat{\mu}} \sum_s w_{hci} y_{hci} I\{y_{hci} \leq \hat{\xi}_p\}$$

To estimate the mean squared error of the Lorenz curve ordinates we simply use the values of $u^*_{hc}$ defined by (6.3) in (6.1)

$$u^*_{hc} = \frac{1}{\hat{\mu}} \sum_s w_{hci} [(y_{hci} - \hat{\xi}_p) I\{y_{hci} \leq \hat{\xi}_p\} + p\hat{\xi}_p - y_{hci}\hat{L}(p)] \qquad (6.3)$$

Similarly, the *MSE* of the quantile share

$$\hat{Q}(p_1, p_2) = \frac{1}{\hat{\mu}} \sum_s w_{hci} y_{hci} I\{\hat{\xi}_{p_1} < y_{hci} \leq \hat{\xi}_{p_2}\}$$

is approximated by (6.1) using

$$u^*_{hc} = \frac{1}{\hat{\mu}} \sum_s w_{hci} \Big[ (y_{hci} - \hat{\xi}_{p_2}) I\{y_{hci} \leq \hat{\xi}_{p_2}\} - (y_{hci} - \hat{\xi}_{p_1}) I\{y_{hci} \leq \hat{\xi}_{p_1}\}$$

$$+ p_2 \hat{\xi}_{p_2} - p_1 \hat{\xi}_{p_1} - y_{hci} \hat{Q}(p_1, p_2) \Big]$$

The Low Income Measure defined by (1.3) is estimated as

$$\hat{\theta} = \hat{F}(\hat{M}/2) = \sum_s w_{hci} I\{y_{hci} \leq \hat{M}/2\}$$

The mean squared error of the low income measure can be estimated approximately by the expression (6.1), where, (from the equation (5.1))

$$u^*_{hc} = -\frac{\hat{f}(\hat{M}/2)}{2\hat{f}(\hat{M})} \sum_s w_{hci} \left[ I\{y_{hci} \leq \hat{M}\} - \frac{1}{2} \right]$$

$$+ \sum_s w_{hci} \left[ I\{y_{hci} \leq \hat{M}/2\} - \hat{\theta} \right]$$

**7. Illustration** The methodology above is illustrated with an application to the family income data collected in the Canadian Survey of Consumer Finance in 1988 (SCF–88). We use the file on the Disposable Income of Economic Families obtained for the province Ontario. The disposable income is defined as the total income reduced by the tax reported in the survey. The SCF uses framework of the Canadian Labour Force Survey which is based on a stratified, multistage design. The Ontario sample contains 7474 households, situated in 525 clusters (PSU's), allocated in 91 strata. For more details on the sample design see *Methodology of the Canadian Labour Force Survey*, Catalogue 71–526, Statistics Canada. To each record a survey weight which is an adjusted sample weight is attached. We estimated the median $M$, the Gini coefficient $G$, the Low Income Measure $\theta$, and the quintile shares.

Their standard errors are obtained using the proposed methodology and the jackknife 'delete–one cluster' method.

It is known that the jackknife variance estimator performs poorly for quantiles due to its inconsistency (Kovar, Rao and Wu, 1988). There are some recent results (Shao and Wu, 1989, Rao and Yue 1992) suggesting that the 'delete $d$' jackknifing and 'delete one cluster', under certain conditions may have desirable asymptotic properties for the variance estimation of non–smooth statistics like quantiles or low income measure. On the other side for statistics like Gini family coefficient, Lorenz curve ordinates and quantile shares, the jackknife estimator of the asymptotic variance is consistent (Shao 1992).

Unlike the jackknifing, the estimating equation approach is not computationally intensive. It provides formulas for asymptotic variance which are easy to programming despite their complicated look.

With a single sample there is no way of serious comparison of the applied methods for the variance estimation. Therefore, the purpose of the example is to point out the differences in standard errors and coefficients of variation obtained by the estimating equation approach and a computationally intensive method like the jackknifing. Results are summarized in Table below.

**8. Summary** The problem of estimating the variance of complex statistics such as measures of income inequality, have eluded statisticians for years. Replication methods such as the jackknife are often suggested for estimation. The advantage of the linearization approach is that it can be used under a wide class of sampling designs and does not suffer from the need for intensive computations, which methods such as the bootstrap entail. Through the method of estimating functions and the decomposition given in (2.4), we find that some difficult problems can be solved more easily. Of course, we have ignored here the conditions under which asymptotic normality is achieved. These must be established before the results given here should be used.

**Acknowledgment** The authors are deeply grateful to H.J.Mantel for his careful review of an earlier version of the paper where he pointed out some technical errors.

References

Beach, C.M. and Davidson, R. (1983). Distribution-free statistical inference with Lorenz curves and income shares. ***Review of Economic Studies***, 50, 723-735.

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. ***International Statistical Review*** 51, 279-292.

Binder, D.A. (1991). Use of estimating functions for interval estimation from complex surveys. ***Proceedings of the American Statistical Association, Survey Research Methods Section***, 34-42.

Francisco, C.A. and Fuller, W.A. (1991). Quantile estimation with a complex survey design. ***Annals of Statistics*** 19, 454-469.

Glasser, G.J. (1962). Variance formulas for the mean difference and coefficient of concentration. ***Journal of the American Statistical Association*** 57, 648-654.

Kovar, J.G., Rao, J.N.K. and Wu, C.F.J. (1988). Bootstrap and other methods to measure errors in survey estimates. ***The Canadian Journal of Statistics*** 16, Supplement, 25-45.

Nygård, F. and Sandström, A. (1981). ***Measuring Income Inequality***, Stockholm: Almqvist & Wiksell International.

Rao, J.N.K. (1979). On deriving mean square errors and their non-negative unbiased estimators in finite population sampling. ***Journal of the Indian Statistical Association, 17***, 125-136.

Shao, J. and Rao, J.N.K. (1992) Standard Errors for Low Income Proportions Estimated from Stratified Multi-Stage Samples. ***Preprint.***

Särndal, C.-E., Swennson, B., and Wretman, J. (1992). ***Model Assisted Survey Sampling***. New York: Springer-Verlag.

Sendler, W. (1979). On statistical inference in concentration measurement. ***Metrika*** 26, 109-122.

Shao, J. and Wu, C.W.J. (1989). A general Theory for Jackknife Variance Estimation. ***Annals of Statistics, 17***,1176-1197

Shao, J. (1992). Inferences Based on L-statistics in Survey Problems: Lorenz Curve, Gini Family and Poverty Proportion. ***In Proceedings of the Workshop on Statistical Issues in Public Policy Analysis***, Carleton University and University of Ottawa, 1993

Woodruff, R.S. (1952). Confidence intervals for medians and other position measures. ***Journal of the American Statistical Association, 47***,635-646.


**Measures of Income Inequality and Their Standard Errors**

| Measure | Estimate | Standard Error | |
|---|---|---|---|
| | | Estimating Equation Approach | Jackknifing 'Delete one Cluster' |
| **Median** | 31705 | 303.3 | 569.8 |
| **Gini** | 0.3482 | 0.005 | 0.005 |
| **Low Income Measure** | 19.804 | 0.586 | 0.613 |
| **Quintile Shares** | | | |
| **Q(0, 0.2)** | 5.608 | 0.123 | 0.167 |
| **Q(0.2, 0.4)** | 11.858 | 0.126 | 0.221 |
| **Q(0.4, 0.6)** | 17.752 | 0.136 | 0.282 |
| **Q(0.6, 0.8)** | 24.607 | 0.119 | 0.337 |
| **Q(0.8, 1.0)** | 40.174 | 0.393 | 0.451 |