

**On Variance Estimation for a Gini Coefficient Estimator
Obtained from Complex Survey Data**

Ahmed A. Hoque

Judith A. Clarke

*Department of Economics, University of Victoria
Victoria, B.C., Canada V8W 2Y2*

February 2014

Summary

Obtaining variances for the plug-in estimator of the Gini coefficient for inequality has preoccupied researchers for decades with proposed analytic formulae often cumbersome to apply, in addition to being obtained assuming an *iid* structure. Bhattacharya (2007, *Journal of Econometrics*) provides an (asymptotic) variance when data arise from a complex survey, a sampling design common with data frequently used in inequality studies. Under a complex survey sampling design, we prove that Bhattacharya's variance estimator is equivalent to an asymptotic version of the estimator derived by Binder and Kovačević (1995, *Survey Methodology*) more than a decade earlier. In addition, we show that Davidson's (2009, *Journal of Econometrics*) derived variance, for the *iid* case, is a simplification of that provided by Binder and Kovačević. These results are computationally useful, as the Binder and Kovačević variance estimator is straightforward to calculate in practice. To aid applied researchers, we show how auxiliary regressions can be used to generate the plug-in Gini estimator and its asymptotic variance, irrespective of the sampling design. Health data on the body mass index for Bangladeshi women is employed in an illustration.

Keywords: *Inequality; Asymptotic inference; Gini index; Complex survey*

Corresponding Author Contact:

Judith A Clarke, Department of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, B.C., Canada V8W 2Y2; e-mail: jaclarke@uvic.ca; Phone: (250) 721-8542; FAX: (250) 721-6214

1. INTRODUCTION

Arguably the best known and most widely employed measure of inequality, the Gini coefficient, proposed by Corrado Gini in 1914, has been the focus of many theoretical and empirical studies. For instance, the Gini coefficient is reported extensively as a way to rank countries in terms of income inequality by, for example, the United Nations, the World Bank and the Central Intelligence Agency. Two recent empirical studies that adopt the Gini coefficient, among the many that could be listed, are Nho (2006), who considers regional income inequality for Korean households, and Slater et al. (2009), who examine the prevalence of being overweight and obese in Canadian adults across a range of socio-economic and geographic groupings. Both of these cited works estimate Gini coefficients from sample data obtained from complex survey designs, a multistage sampling process, typically involving stratification and clustering used to guarantee representation of groups of interest as well as to keep costs as low as possible^{1,2}. The process used to obtain the sample, which clearly is not a randomly drawn *iid* sample, needs to be accounted for when forming both the estimator and an associated variance. Such applications motivate our work.

The problem of obtaining a variance for a Gini coefficient estimator has been investigated by many researchers including, but not limited to, Hoeffding (1948), Glasser (1962), Sendler (1979), Beach and Davidson (1983), Gastwirth and Gail (1985), Schechtman and Yitzhaki (1987), Sandström et al. (1985, 1988), Nygård and Sandström (1989), Yitzhaki (1991), Shao (1994), Binder and Kovačević (1995), Bishop et al. (1997), Karagiannis and Kovačević (2000), Giles (2004), Modarres and Gastwirth (2006), Bhattacharya (2007), Xu (2007), Davidson (2009) and Qin et al. (2010). Some of these studies propose analytic (asymptotic) variances while others adopt resampling methods, with the latter works often claiming that such tools are preferable, as they avoid the mathematical and coding complexities associated with

¹ See, for instance, Cochran (1977), Skinner et al. (1989) and Cameron and Trivedi (2005, Chapter 24).

² Nho (2006)'s work, for example, uses household data from the Korean National Survey of Household Incomes and Expenditures, which stratifies the country according to geographical regions and administrative districts and then forms clusters based on Census enumeration areas. Clusters are selected using probability sampling proportional to size and then two segments of five households are randomly selected from the sampled clusters.

the available analytic expressions^{3,4}. The extent of studies, even given our incomplete list, suggests that there should be nothing more to say about a seemingly simple task of providing a variance for a sample statistic.

However, the research that provides analytic asymptotic variances has adopted several different methods, with links between them and the resulting formulae often unclear. For instance, some obtain asymptotic variances based on U -statistics (e.g., Hoeffding, 1948; Yitzhaki, 1991; Bishop et al. (1997); Xu, 2007) whereas others use L -statistic theory (e.g., Nygård and Sandström, 1989; Shao, 1994). Via the use of Taylor-series expansions, Binder and Kovačević (1995) and Davidson (2009) provide approximation expressions for Gini coefficient estimators, from which they obtain variances; Binder and Kovačević (1995) allow for complex survey sample data whereas Davidson (2009) assumes an *iid* sample. The use of estimating equations underlies the work of Binder and Kovačević (1995) (see also Kovačević and Binder, 1997), who, after generating an appropriate asymptotic approximation expression, appeal to standard survey theory to provide a so-called linearization variance. Coding for this variance is, in our view, not complicated, especially with access to software that accounts for survey design⁵.

Although Davidson (2009) limits attention to an *iid* random sample, he approaches variance estimation in a similar fashion by also deriving an approximation for his Gini estimator, which he uses to suggest a variance estimator. Bhattacharya (2007), based on his earlier, more general, paper (Bhattacharya, 2005) frames estimation of the Gini coefficient, with complex survey sample data, within generalized method of moments (GMM) theory, and appeals to available GMM results to show the consistency and asymptotic normality of a plug-in estimator. Using sample empirical process theory and the functional delta method, Bhattacharya (2007) derives an

³ Ogwong (2000, p123) provides one such example, stating, to justify his jackknife approach, that “standard errors of the Gini index that have been suggested so far ... are either mathematically very complicated or require heavy computation which cannot be conveniently undertaken using commonly available regression software packages”.

⁴ Empirical work has also often followed such recommendations; for instance, Nho (2006, p341) states “The standard error of the gini index is estimated by jackknife and bootstrap methods rather than the traditional delta method because of the complexity of the latter.”

⁵ For econometricians, Stata is likely the best known such package. Jenkins (2006) provides a Stata add-on entitled *svylorenz* that readily produces Binder and Kovačević’s variance estimator.

expression for the asymptotic variance of his Gini estimator, taking account of the sample design. One of the many notable contributions of Bhattacharya's work is his breakdown of the variance into three components: the estimate of the variance without taking the sample design into account; the impact of clustering on the variance; and the stratum effect on the variance. The resulting variance formula, to quote Davidson (2009, p30), "is however not at all easy to implement"⁶.

We focus on these last three cited works: Binder and Kovačević (1995), Bhattacharya (2007) and Davidson (2009). First, we show that the approximation for the Gini estimator obtained by Bhattacharya (2007) is algebraically equivalent to that provided by Binder and Kovačević (1995), over a decade earlier. Second, we show that Davidson's (2009) approximation for the Gini estimator is a special example of Binder and Kovačević's (1995) expression. This latter result is not surprising, as Davidson notes (p32) that the random variable in his approximation "can with some effort be shown to be the same as that used by Bhattacharya (2007)"; the difference here is that Davidson's approximation follows from that of Binder and Kovačević's (1995) with very little effort. We believe that it is important to note the connections of the recent econometrics works (Bhattacharya, 2007; Davidson, 2009) with an earlier work, from the sample survey literature, which seems to have slipped their attention.

We then consider variance estimators, showing that Bhattacharya's and Binder and Kovačević's variance estimators are equivalent, at least asymptotically, with the asymptotic analysis referring to the number of sampled clusters in each stratum going to infinity at the same rate. This outcome is expected given the equivalence of the approximations for the estimator of the Gini coefficient. That the two works lead to the same variance is useful, as calculating Binder and Kovačević's variance is straightforward, especially with packages designed to account for complex surveys. Furthermore, we show that Davidson's variance estimator is a special example of that derived by Binder and Kovačević.

Another recent study that also stresses the lack of connection of branches of the literature with regard to estimation of the sampling variance for the Gini estimator is

⁶ In addition, our reading of Bhattacharya's work suggests some typographical errors that further complicate use of the results. We have attempted to correct these in our presentation.

Langel and Tillé (2013)⁷. These authors provide an excellent survey of the evolution of papers on variance estimation, highlighting some methodological issues that have arisen over time. In particular, they concentrate on linking various linearization techniques used to approximate the variance. The papers we consider – Binder and Kovačević (1995), Bhattacharya (2007) and Davidson (2009) – are mentioned by Langel and Tillé (2013), as they adopt linearization tools, but Langel and Tillé do not algebraically provide the results contained in our work. We believe that our study provides a natural complement to Langel and Tillé (2013), being of particular use to applied economists, who are likely to be more aware of the research in economics and econometrics journals than in statistics journals. It is our view that the variance formula derived by Bhattacharya (2007), although algebraically equivalent to that obtained by others as we show, is useful for understanding the implications of complex survey sampling, compared with simple random sampling, a feature not highlighted by Langel and Tillé. However, we believe that Bhattacharya’s formula is simply not friendly to code, whereas other formulations, such as that of Binder and Kovačević (1995), lead to a linearization variance, under complex survey sampling, that is far easier to compute. Our work also suggests easy ways to obtain the linearization variance using artificial regressions, not considered by Langel and Tillé, which we believe applied researchers will find helpful⁸.

These auxiliary regressions can be used to obtain the Gini estimator and its associated variance without the need for specialized survey software. Although using artificial regressions for estimating the Gini coefficient has been considered in the literature (e.g., Ogwong, 2000; Giles, 2004; Davidson, 2009), this has been limited to data presumed to arise under an *iid* assumption, whereas our regressions account for sample data obtained from a complex survey design. Additionally, we show how auxiliary regressions can be employed to compute variances, often regarded as burdensome.

This paper is organized as follows. Section 2 outlines the sampling design framework considered in our theoretical work, along with considering sampling

⁷ We thank Thomson Ogwong for informing us of this study after he read a draft version of our paper.

⁸ We note that the discussion in Langel and Tillé (2013) on regression-based variance estimation, their section 5, is not related to our proposed regressions.

weights. For comparability, we adopt Bhattacharya's (2005, 2007) structure, but we explain how this setup does not limit the applicability of results for researchers using other commonly considered complex survey data. We explore estimation of the Gini coefficient and approximations for the estimator in Section 3. Our main results are presented in Section 4, where we examine asymptotic variance (hereafter, variance) estimation of the usual plug-in estimator of the Gini coefficient using data obtained under a complex survey design. In addition, variance estimation with a random *iid* sample is considered. Section 5 suggests ways to obtain estimators in practice. We detail results from a small empirical study in Section 6. Our illustration shows how the theoretical results can be applied under a somewhat different sampling design than that examined by Bhattacharya. We conclude in Section 7 and provide algebraic proofs in an Appendix.

2. FRAMEWORK

The effect of the sampling design on estimation of a population parameter is discussed in standard statistics texts, such as Cochran (1977) and Wolter (2007). Our focus is on multistage sampling, often adopted when obtaining household (or individual) data, a sampling design that may involve one or more combinations of sampling techniques, with the key outcome being that the sample cannot be regarded as *iid*. Ignoring how the sample has been constructed (such as behaving as if the sample is *iid*) can lead to inconsistent population parameter estimators and inconsistent variance estimators. Some common sampling methods include stratification, clustering, double sampling, multiple frames, poststratification and so on (see, e.g., Wolter, 2007). We focus here on a design that consists of first stage stratification⁹ and second stage clustering, as it turns out that further sampling stages, after the initial level of clustering, do not affect the (asymptotic) variance estimator, which is computed from quantities formed from the ultimate cluster observations (see, e.g., Skinner et al., 1989, p47)¹⁰.

⁹ Should the sampling design involve several stages of stratification (e.g., urban/rural, then province/state), prior to the first level of clustering, then it is the "ultimate" level of (initial) stratification that needs to be considered.

¹⁰ Because any additional correlation between units, introduced from further stratification and clustering, is already accounted for, as the "primary" cluster level units are assumed to be correlated.

Stratifying divides a population into relatively homogenous subgroups before sampling (e.g., area of residence, gender, or race) with sample selection then proceeding separately for each stratum. Such a design typically breaks the *identical* part of an *iid* assumption. The *independent* component, on the other hand, is usually violated with clustering, which splits the population into contiguous groupings; e.g., villages in rural areas and blocks or enumeration areas in urban areas. Given this contiguity, units within clusters are usually correlated.

For comparability, we adopt Bhattacharya's (2007) framework of a *household* survey with interest in inequality for a well-being variable that is at the *individual* level, but with the feature that the value of the well-being variable is the same for all members of a household; e.g., per capita annual household consumption expenditure. This implies that the unit being sampled is the *household*, but the relevant sampling weight (discussed shortly) is for the *individual*; the number of sampled units is the total number of households. Such a structure, although commonly of interest, differs from that explored by other theoretical studies. For instance, Biewen and Jenkins (2006) and Clarke and Roy (2012) consider survey designs where the ultimate unit is the *individual* (rather than the household) so that the total number of observations equals the number of individuals, with the *individual* sampling weight of relevance. In contrast, Binder and Kovačević (1995) illustrate their theoretical results using a *household* level variable (family income) for *households*, so the adopted sampling weight is for the *household*. It turns out that such specifics are not important. With appropriate changes in the sampling weight and the number of units being sampled, our presented theoretical results carry through. To illustrate, we purposely examine an application where the well-being variable is measured at the *individual* level, with *individuals* (ever-married women in a household) being the ultimate unit of interest.

Following Bhattacharya's setup, our notation with respect to the design follows. Let $\{\mathcal{U}[hci]: h = 1, \dots, H; c = 1, \dots, N_h; i = 1, \dots, M_{hc}\}$ be a finite population stratified into H strata, with N_h clusters or primary sampling units within each stratum so that the population consists of $\mathcal{N} = \sum_{h=1}^H N_h$ clusters. In cluster c , within stratum h , there are M_{hc} units, which we assume to be households, leading to $\mathcal{M}_h = \sum_{c=1}^{N_h} M_{hc}$ households in stratum h ; the population number of households is

then $\mathcal{M} = \sum_{h=1}^H \sum_{c=1}^{N_h} M_{hc}$. A sample of n_h clusters is drawn from the h^{th} stratum via simple random sampling, with the total number of sampled clusters being $N = \sum_{h=1}^H n_h$ ¹¹. From cluster c , within stratum h , we suppose that k households are selected¹², using simple random sampling, such that the total number of sampled households is $M = k \sum_{h=1}^H n_h = kN$. Let s_{hci} be the number of members in the i^{th} household in the c^{th} cluster within the h^{th} stratum.

Given the complex survey's design, along with the common practice of oversampling particular subgroups to ensure stable estimates, household members in the population likely will not have the same probability of being included in the sample, a feature that is accounted for by a sampling weight for each member that denotes how many individuals this observation represents in the population. Under the described sampling framework, the weight, which is inversely proportional to selection probability, is $W_{hci} = \frac{M_{hc} N_h}{k n_h} s_{hci}$, which is often normalized so that $\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k W_{hci}$ is equal to the number of sampled households or individuals; we assume that the weights have been scaled to ensure that $\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k W_{hci} = M$, the number of sampled households. In addition, it simplifies the algebra to adopt standardized weights $w_{hci} = \frac{W_{hci}}{\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k W_{hci}}$ such that $\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} = 1$ ¹³.

3. GINI COEFFICIENT ESTIMATION AND ASYMPTOTIC APPROXIMATIONS

In subsection 3.1, we define the Gini coefficient and provide its natural plug-in estimator when sample data are obtained from a multistage complex survey. We also give the formula for the estimator when the sample is regarded as *iid*, randomly drawn from an underlying population. We follow, in subsection 3.2, with a

¹¹ Clusters are sometimes selected with probability proportional to size, which alters the sampling weight but not the fundamental results presented herein.

¹² The assumption that the same number of households is sampled from each cluster can be relaxed.

¹³ Sampling weights also usually account for other features such as different nonresponse rates of interviews. Should clusters have been selected with probability proportional to size, then $W_{hci} = \frac{\sum_{c=1}^{N_h} M_{hc}}{k n_h} s_{hci}$.

consideration of approximations for the estimator, showing that Binder and Kovačević (1995) and Bhattacharya (2007) obtain equivalent expressions. We also show that the approximation of Davidson (2009) is a straightforward special example of that derived by Binder and Kovačević (1995).

3.1. Gini Coefficient and Plug-in Estimator

The Gini coefficient, bounded by zero and one, is typically defined as twice the area between the 45° –line and the Lorenz (1905) curve. The Lorenz curve graphically illustrates the distribution of the well-being variable (e.g., household income, education attainment, per capita mean household consumption expenditure) by displaying the cumulative share of the well-being variable against its recipient share. Recall that our framework considers an individual level characteristic of well-being. Specifically, for a random well-being variable $y \in [0, \infty)$ with cumulative distribution function $F(y)$ and finite non-zero mean $\mu \equiv \int_0^\infty y dF(y)$, the Lorenz curve is

$$L(p) = \frac{1}{\mu} \int_0^{z_p} y dF(y) , \quad (3.1)$$

where $z_p = F^{-1}(p) = \inf\{y | F(y) \geq p\}$ is the p th quantile or fractile of the distribution function with $p = F(z_p) = \int_0^{z_p} dF(y)$, $0 \leq p \leq 1$. On the 45° –line, the line of equality, $p = L(p)$, whereas there is inequality when $p > L(p)$. Given expression (3.1), the Gini coefficient is then commonly defined as

$$G = 1 - 2 \int_0^1 L(p) dp = \frac{2}{\mu} \int_0^\infty y F(y) dF(y) - 1 . \quad (3.2)$$

This summary measure of the degree of inequality, is zero (one) for a perfectly equal (unequal) distribution. To proceed, let $\hat{F}(y_{hci})$ be the empirical distribution function for y_{hci} , which given the sampling design, is

$$\begin{aligned}
\hat{F}(y_{hci}) &= \sum_{r=1}^H \sum_{s=1}^{n_r} \sum_{t=1}^k w_{rst} I[y_{rst} \leq y_{hci}] \\
&= \sum_{d=1}^M w_d I[y_{(d)} \leq y_{hci}] \quad ,
\end{aligned} \tag{3.3}$$

where $y_{(d)}$ is the d^{th} order statistic in the full sample and w_d is the associated sampling weight. Although not necessary, it is sometimes helpful to write expressions in terms of order statistics, as it provides consistency with some of the related research. Then, let $\hat{z}(p)$ be the estimated sample quantile (i.e., $\hat{z}(p) = \inf\{y_{hci} | \hat{F}(y_{hci}) \geq p\}$) and $\hat{\mu} = \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} y_{hci}$ be the estimator of μ , such that we obtain the following estimator for $L(p)$, allowing for the sampling design:

$$\hat{L}(p) = \frac{1}{\hat{\mu}} \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} y_{hci} I(y_{hci} \leq \hat{z}(p)) = \frac{\hat{\alpha}(p)}{\hat{\mu}} \quad ,$$

with $\hat{\alpha}(p)$ being an estimator of $\alpha(p) = E(YI(Y \leq z(p)))$. Using this, we have the following estimator of G :

$$\begin{aligned}
\hat{G} &= 1 - 2 \int_0^1 \hat{L}(p) dp = 1 - \frac{2}{\hat{\mu}} \int_0^1 \left[\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} y_{hci} I(y_{hci} \leq \hat{z}(p)) \right] dp \\
&= 1 - \frac{2}{\hat{\mu}} \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} y_{hci} \left[\int_0^1 I(y_{hci} \leq \hat{z}(p)) dp \right] \quad ,
\end{aligned}$$

which can be estimated by

$$\begin{aligned}
\hat{G} &= 1 - \frac{2}{\hat{\mu}} \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} y_{hci} \left[\sum_{d=1}^M w_d I(y_{(d)} \geq y_{hci}) \right] \\
&= 1 - \frac{2}{\hat{\mu}} \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} y_{hci} (1 - \hat{F}(y_{hci})) \quad .
\end{aligned}$$

This is the estimator adopted by Bhattacharya (2007). Re-arranging, we arrive at Binder and Kovačević's (1995) estimator:

$$\hat{G} = \frac{2}{\hat{\mu}} \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} y_{hci} \hat{F}(y_{hci}) - 1 \quad . \tag{3.4}$$

For an *iid* randomly drawn sample of individuals of size n_0 , this result suggests using

$$\hat{G} = \frac{2}{M\hat{\mu}} \sum_{d=1}^M y_d \hat{F}(y_d) - 1 \quad , \quad (3.5)$$

where $\hat{F}(y_d) = \frac{1}{M} \sum_{j=1}^M I(y_j \leq y_d)$. However, noting that the second expression in (3.2) can be equivalently written as $G = \frac{1}{\mu} \int_0^\infty y d(F(y))^2 - 1$, and using order statistics, Davidson (2009) suggests the following estimator for G , which overcomes the problem of defining the empirical distribution function as being either right- or left-continuous:

$$\begin{aligned} \hat{G} &= \frac{1}{\hat{\mu}} \sum_{d=1}^M y_{(d)} \left[\left(\frac{d}{M} \right)^2 - \left(\frac{d-1}{M} \right)^2 \right] - 1 \\ &= \frac{2}{\hat{\mu}M^2} \sum_{d=1}^M y_{(d)} \left(d - \frac{1}{2} \right) - 1 \quad . \end{aligned} \quad (3.6)$$

In this formula, an average of the lower and upper limits is used for \hat{F} . Expression (3.6) is equivalent to that considered by, amongst others, Sandler (1979), Nygård and Sandström (1989), Ogwong (2000) and Giles (2004).

3.2 Approximations for $(\hat{G} - G)$

We now turn to obtaining approximate expressions for $(\hat{G} - G)$ from which we can obtain variance estimators. Bhattacharya (2007) frames estimation as a method of moments problem, showing that an approximate expression for the Lorenz share at a fixed percentile p , with $\theta = (z(p), \alpha(p), \mu)$ is given by

$$\hat{L}(p) - L(p) \approx \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} m(y_{hci}, \theta) \quad , \quad (3.7)$$

where

$$\begin{aligned} m(y_{hci}, \theta) &= \frac{1}{\mu} (y_{hci} I(y_{hci} \leq z(p)) - \alpha(p)) + \frac{1}{\mu} z(p) (p - I(y_{hci} \leq z(p))) \\ &\quad + \frac{\alpha(p)}{\mu^2} (\mu - y_{hci}) \\ &= \frac{1}{\mu} \left[y_{hci} I(y_{hci} \leq z(p)) + z(p) (p - I(y_{hci} \leq z(p))) - \frac{\alpha(p)}{\mu} y_{hci} \right] \quad . \end{aligned}$$

Then, the approximate expression for $(\hat{G} - G)$ is

$$\begin{aligned}
\hat{G} - G &\approx -2 \int_0^1 dp \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} m(y_{hci}, \theta) \\
&= -2 \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} \left[\int_0^1 m(y_{hci}, \theta) dp \right] \\
&= \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} \Psi_{hci} .
\end{aligned} \tag{3.8}$$

An estimator of Ψ_{hci} is:

$$\begin{aligned}
\hat{\Psi}_{hci} &= -2 \sum_{d=1}^M w_d \left\{ \frac{1}{\hat{\mu}} \left[y_{hci} I(y_{hci} \leq y_{(d)}) + y_{(d)} (\hat{F}(y_{(d)}) - I(y_{hci} \leq y_{(d)})) \right. \right. \\
&\quad \left. \left. - \frac{\hat{\alpha}(y_{(d)})}{\hat{\mu}} y_{hci} \right] \right\} ,
\end{aligned}$$

where $\hat{\alpha}(y_{(d)}) = \sum_{r=1}^H \sum_{s=1}^{n_r} \sum_{t=1}^k w_{rst} y_{rst} I(y_{rst} \leq y_{(d)})$ is used to estimate $\hat{\alpha}(p)$. We then have Bhattacharya's (2007) approximation:

$$\begin{aligned}
\hat{G} - G &\approx \\
\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} &\left(-2 \sum_{d=1}^M w_d \left\{ \frac{1}{\hat{\mu}} \left[\begin{aligned} &y_{hci} I(y_{hci} \leq y_{(d)}) + \\ &y_{(d)} \hat{F}(y_{(d)}) - y_{(d)} I(y_{hci} \leq y_{(d)}) - \frac{\hat{\alpha}(y_{(d)})}{\hat{\mu}} \end{aligned} \right] \right\} \right) \tag{3.9}
\end{aligned}$$

It should be noted that we have written Bhattacharya's approximate expression using the traditional summing over stratum, clusters, and households, in contrast to Bhattacharya, who rearranges the summing to take account of his asymptotic analysis, which is with respect to the number of clusters¹⁴. Specifically, let $a_h = \frac{n_h}{N}$ and $I(c \in h) = 1$ when cluster c is in stratum h , 0 otherwise. Then, let $W_{hci}^* =$

¹⁴ As household surveys often have significantly more clusters sampled per stratum than the number of strata or the number of households sampled per cluster, it seems sensible to consider an asymptotic framework with the number of clusters going to infinity, holding the number of strata and households per cluster fixed and finite. This is Bhattacharya's framework. In contrast, Krewski and Rao (1981), Rao and Wu (1985), Binder and Kovačević (1995) and Kovačević and Binder (1997), amongst others, consider asymptotic analysis with the number of strata tending to infinity, assuming the number of clusters per stratum is fixed.

$NW_{hci} = (I(c \in h)) \times \frac{M_{hc}N_h}{ka_h} s_{hci}$ and $w_{hci}^* = W_{hci}^* / \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k W_{hci}$ so that we write (3.8) as

$$\hat{G} - G \approx -\frac{2}{N} \sum_{c=1}^N \sum_{h=1}^H \sum_{i=1}^k w_{hci}^* \left[\int_0^1 dp \left\{ \frac{1}{\mu} \left[y_{hci} I(y_{hci} \leq z(p)) + z(p) (p - I(y_{hci} \leq z(p))) - \frac{\alpha(p)}{\mu} y_{hci} \right] \right\} \right].$$

Rearranging, we obtain Bhattacharya's (2007, p685) explicit expression:

$$\sqrt{N}(\hat{G} - G) \approx -\frac{2}{\sqrt{N}} \sum_{c=1}^N \Psi_i \quad (3.10)$$

where

$$\Psi_i = \int_0^1 dp \left\{ \sum_{h=1}^H \sum_{i=1}^k w_{hci}^* \left(\frac{1}{\mu} \left[y_{hci} I(y_{hci} \leq z(p)) + z(p) (p - I(y_{hci} \leq z(p))) - \frac{\alpha(p)}{\mu} y_{hci} \right] \right) \right\}.$$

We now consider Binder and Kovačević's (1995) approximation for $(\hat{G} - G)$. Binder and Kovačević (1995) approach estimation using estimating equations¹⁵, a general way to estimate population parameters. Some examples of methods that lead naturally to estimating equations are maximum likelihood, method of moments and least squares. To illustrate, following Binder and Kovačević (1995), suppose interest lies in estimation of a finite population parameter λ that can be written as the solution to

$$\int u(y, \lambda) dF(y) = 0 \quad ,$$

where $F(y)$ is the finite population distribution function. Then, with $\hat{F}(y)$ being a consistent empirical distribution function, the estimating equations estimator of λ is that value of $\hat{\lambda}$ such that

¹⁵ First proposed by Godambe (1960); see also Godambe and Thompson (1978, 1984), Binder (1991) and Binder and Patak (1994).

$$\int \hat{u}(y, \hat{\lambda}) d\hat{F}(y) = 0 \quad ,$$

where $\hat{u}(y, \hat{\lambda})$, the estimating equation, is an estimate (or approximation) of $u(y, \lambda)$. Given these definitions, it is straightforward to see how moment equations and likelihood equations are examples of estimating equations. Not surprisingly, sometimes more than one estimating equation is required. Using this method, Binder and Kovačević (1995) provide the following two estimating equations for estimating the Lorenz curve ordinate and the 100 p th percentile of the distribution:

$$u_1(y, L(p)) = I(y \leq z(p))y - L(p)y \quad \& \quad u_2(y) = I(y \leq z(p)) - p \quad .$$

Using these estimating equations and approximations, based on theorems from Francisco and Fuller (1991), Binder and Kovačević (p141) obtain an approximation for the Lorenz share at a fixed percentile p :

$$\hat{L}(p) - L(p) \approx \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} u(y_{hci}, \theta) \quad ,$$

where

$$u(y_{hci}, \theta) = \frac{1}{\mu} [(y_{hci} - z(p))I(y_{hci} \leq z(p)) + pz(p) - y_{hci}\alpha(p)/\mu] \quad .$$

By simple inspection, $u(y_{hci}, \theta) = m(y_{hci}, \theta)$. Given the equivalence of these approximations, the approximate expressions of Binder and Kovačević (1995) and Bhattacharya (2007) for \hat{G} are also the same. It is, nevertheless, useful to show this result, as we find Binder and Kovačević's (1995) expression to be practically more convenient. Binder and Kovačević (1995, p143) provide the sample estimating equation for the Gini coefficient:

$$\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} [(2\hat{F}(y_{hci}) - 1)y_{hci} - \hat{G}y_{hci}] = 0 \quad ,$$

and the estimator for their approximation of $(\hat{G} - G)$ is:

$$\hat{G} - G \approx$$

$$\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} \left\{ \frac{2}{\hat{\mu}} \left[y_{hci} \left(\hat{F}(y_{hci}) - \frac{(\hat{G} + 1)}{2} \right) + B(y_{hci}) - \frac{\hat{\mu}}{2} (\hat{G} + 1) \right] \right\} \quad (3.11)$$

$$= \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} u_{hci}^* , \quad (3.12)$$

where

$$B(y_{hci}) = \sum_{r=1}^H \sum_{s=1}^{n_r} \sum_{t=1}^k w_{rst} y_{rst} I(y_{rst} \geq y_{hci}) .$$

The equivalence of expressions (3.9) and (3.11) is shown in the Appendix under the Proof of Result 1. When calculating variances, considered in the next section, practitioners need to generate either u_{hci}^* or $\hat{\Psi}_{hci}$. Although equivalent terms, having already formed $\hat{\mu}$, $\hat{F}(y_{hci})$ and \hat{G} , it is our belief that it is computationally easier to code u_{hci}^* rather than $\hat{\Psi}_{hci}$, as $B(y_{hci})$ is the only additionally required term.

To end this subsection, we simplify Binder and Kovačević's (1995) expression (3.11) for the case of a randomly drawn *iid* sample, showing that it straightforwardly gives the expression derived by Davidson (2009). In the *iid* case, expression (3.11) becomes:

$$\hat{G} - G \approx \frac{1}{M} \sum_{j=1}^M \frac{2}{\hat{\mu}} \left[y_j \left(\hat{F}(y_j) - \frac{(\hat{G} + 1)}{2} \right) + B(y_j) - \frac{\hat{\mu}}{2} (\hat{G} + 1) \right] , \quad (3.13)$$

where $B(y_j) = \frac{1}{M} \sum_{i=1}^M y_i I(y_i \geq y_j)$. Using Davidson's notation, $B(y_j) = \hat{\mu} - \hat{m}(y_j)$, $\frac{\hat{\mu}}{2} = \frac{(\hat{G}+1)}{2}$ and $\hat{I} = \frac{\hat{\mu}}{2} (\hat{G} + 1)$, where $I = \int_0^\infty yF(y)dy$. Making these substitutions, with some minor algebraic manipulations, we obtain

$$\hat{G} - G \approx \frac{1}{M} \sum_{d=1}^M \frac{2}{\hat{\mu}} \left[-\frac{\hat{I}}{\hat{\mu}} (y_d - \hat{\mu}) + y_d \hat{F}(y_d) - \hat{m}(y_d) - (2\hat{I} - \hat{\mu}) \right], \quad (3.14)$$

which is Davidson's (p32) estimator of the approximation for $(\hat{G} - G)$. Note that we can equivalently write (3.13) as:

$$\hat{G} - G \approx \frac{1}{M} \sum_{d=1}^M u_d^*,$$

where

$$\begin{aligned} u_d^* &= \frac{2}{\hat{\mu}} \left[y_d \left(\hat{F}(y_d) - \left(\frac{\hat{G} + 1}{2} \right) \right) + B(y_d) - \frac{\hat{\mu}}{2} (\hat{G} + 1) \right] \\ &= \frac{(\hat{Z}_d - \bar{Z})}{\hat{\mu}}, \end{aligned} \quad (3.15)$$

in Davidson's adopted notation, with $\hat{Z}_d = -(\hat{G} + 1)y_d + 2(y_d \hat{F}(y_d) - \hat{m}(y_d))$ and $\bar{Z} = \hat{\mu}(\hat{G} - 1)$.

In this section, we have shown that recent derivations of approximate expressions for \hat{G} by Bhattacharya (2007) and Davidson (2009) are either equivalent to or a special example of the expression obtained by Binder and Kovačević (1995), over a decade earlier. We now turn to variances for \hat{G} .

4. VARIANCE ESTIMATORS

We first present the variance estimator of Binder and Kovačević (1995), which is based on standard survey theory, and then consider its limiting form as the number of clusters goes to infinity. We then examine Bhattacharya's (2007) variance estimator, showing that formula is equivalent to the limiting form of Binder and Kovačević's (1995) variance estimator, a formula that we believe is easier to use. Finally, we show that Davidson's (2009) variance formula, suggested for the *iid* case, is easy to obtain from the complex survey variance of Binder and Kovačević (1995). That this follows is expected from our results in the previous section.

Using expression (3.12), an estimator of the variance of \hat{G} is

$$\widehat{Var}(\hat{G}) = Var \left(\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} u_{hci}^* \right),$$

which is merely the variance of a survey weighted total, a well-discussed estimation problem in the survey literature. For instance, following Skinner et al. (1989, p47)¹⁶, the standard nonparametric estimator is:

$$\widehat{Var}(\hat{G}) = \sum_{h=1}^H \left(\frac{n_h}{n_h - 1} \right) \sum_{c=1}^{n_h} (u_{hc}^* - \bar{u}_h^*)^2, \quad (4.1)$$

where $u_{hc}^* = \sum_{i=1}^k w_{hci} u_{hci}^*$ and $\bar{u}_h^* = (1/n_h) \sum_{c=1}^{n_h} u_{hc}^*$. As

$(n_h/(n_h - 1)) = (a_h/(a_h - (1/n))) \rightarrow 1$ as $n_h, n \rightarrow \infty$, an asymptotically¹⁷ equivalent estimator is

$$\widehat{Var}(\hat{G}) = \sum_{h=1}^H \sum_{c=1}^{n_h} (u_{hc}^* - \bar{u}_h^*)^2. \quad (4.2)$$

Bhattacharya (2007), on the other hand, based on his general paper Bhattacharya (2005), provides the asymptotic variance estimator:

$$\begin{aligned} \widetilde{Var}(\hat{G}) &= Var \left(\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} \hat{\Psi}_{hci} \right) \\ &= \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci}^2 \hat{\Psi}_{hci}^2 \\ &\quad + \sum_{h=1}^H \sum_{c=1}^{n_h} \left(\sum_{i=1}^k \sum_{j \neq i}^k w_{hci} w_{hcj} \hat{\Psi}_{hci} \hat{\Psi}_{hcj} \right) \\ &\quad - \sum_{h=1}^H \frac{1}{n_h} \left(\sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} \hat{\Psi}_{hci} \right)^2, \end{aligned} \quad (4.3)$$

where, from Section 3,

¹⁶ This assumes that primary sampling units, the clusters, are selected with replacement.

¹⁷ As per Bhattacharya (2005, 2007), the asymptotic behaviour is with respect to the number of sampled clusters for each stratum going to infinity at the same rate, leading to a_h remaining fixed.

$$\hat{\Psi}_{hci} = -2 \sum_{d=1}^M w_d \left\{ \frac{1}{\hat{\mu}} \left[y_{hci} I(y_{hci} \leq y_{(d)}) + y_{(d)} \left(\hat{F}(y_{(d)}) - I(y_{hci} \leq y_{(d)}) \right) - \frac{\hat{\alpha}(y_{(d)})}{\hat{\mu}} y_{hci} \right] \right\}.$$

This is equivalent to u_{hci}^* (recall Result 1 in the Appendix). Bhattacharya's expression (4.3) is useful for understanding the impact of the sampling design on variance estimation; the first term is the estimator of the variance under a simple random sampling design (or *iid* assumption) with weights, which we denote by $\widehat{Var}_{SRS}(\hat{G})$, the second term is the effect on the variance from clustering and the third term is the impact of stratification on the variance. As we would expect a positive covariance between values obtained from the same cluster, the cluster effect is expected to be positive, raising the variance over that which would arise under an *iid* with weights assumption. Stratification reduces the variance; the more homogeneous are the units within a stratum and the more heterogeneous are units across strata, the higher would be this stratification effect.

Despite the benefits of writing the variance estimator as expression (4.3), it is not especially friendly for practitioners to implement, whereas the form of expression (4.1) (or (4.2)) is simpler to practically calculate. Not surprisingly, given our earlier findings, Bhattacharya's variance (4.3) is equivalent to the limiting form of Binder and Kovačević's variance, expression (4.2). We show this as Proof of Result 2 in the Appendix. Aside from showing the equivalence of Bhattacharya's recent result with that of one derived over a decade earlier, this outcome is beneficial for practitioners as it turns out to be relatively easy to estimate expression (4.2) (and (4.1) should a researcher wish to adopt the standard survey sampling variance). Estimation is discussed in the next section.

Prior to doing so, we consider the case of an *iid* sample. The natural estimator from expression (4.2) is

$$\widehat{Var}(\hat{G}) = \frac{1}{M^2} \sum_{d=1}^M (u_d^* - \bar{u}^*)^2 = \frac{1}{M^2} \sum_{d=1}^M u_d^{*2} \quad (4.4)$$

where u_d^* is as defined in expression (3.15) and using that $\bar{u}^* = \frac{1}{M} \sum_{d=1}^M u_d^* = 0$ ¹⁸. In terms of Davidson's (2009) notation, this is equivalent to $\widehat{Var}(\hat{G}) = \frac{1}{(M\hat{\mu})^2} \sum_{d=1}^M (\hat{Z}_d - \bar{Z})^2$, which is explicitly his equation (19) on p32. That is, as expected, Davidson's proposed variance estimator is a special example of the variance estimator suggested by Binder and Kovačević (1995).

5. CALCULATING ESTIMATES IN PRACTICE

In this section we discuss straightforward ways of practically calculating \hat{G} and its associated variance estimator. In subsection 5.1 we examine obtaining \hat{G} for both the complex survey and *iid* sample and in subsection 5.2 we consider estimating $Var(\hat{G})$ with a complex survey sample. We do not discuss calculating $\widehat{Var}(\hat{G})$ for an *iid* sample as this is just a scaled sum of squares.

5.1 Computing \hat{G}

When data are from an *iid* random sample, it has been show that \hat{G} can be easily obtained from an artificial ordinary least squares (OLS) regression; see, e.g., Ogwong (2000), Giles (2004) and Davidson (2009). Specifically, with unordered data and denoting v as a random error, we estimate the artificial regression:

$$(2\hat{F}(y_d) - 1)\sqrt{y_d} = \theta\sqrt{y_d} + v_d, \quad d = 1, \dots, M$$

by OLS to yield $\hat{\theta} = \hat{G}$. If data are ordered with \hat{F} computed using the average of the lower and upper limits, as advocated by Davidson (2009), the artificial regression:

$$\left(\frac{2d}{M} - \frac{1}{M} - 1\right)\sqrt{y_{(d)}} = \theta\sqrt{y_{(d)}} + v_d, \quad d = 1, \dots, M$$

estimated by OLS results in $\hat{\theta} = \hat{G}$ as defined by expression (3.6).

A similar approach can be adopted with a complex survey. For the auxiliary regressions with complex survey data, the sampling weights are assumed to be normalized such that $\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} = 1$. Should a researcher be using a

¹⁸ This also holds for the complex survey; i.e., $\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} u_{hci}^* = 0$.

software package not explicitly designed to handle complex survey data, the OLS estimator of θ using the artificial regression:

$$(2\hat{F}(y_{hci}) - 1)\sqrt{w_{hci}y_{hci}} = \theta\sqrt{w_{hci}y_{hci}} + v_{hci}, \quad h = 1, \dots, H; c = 1, \dots, n_h; i = 1, \dots, k$$

leads to \hat{G} given in expression (3.4); the data need not be ordered for this regression. That is, we simply estimate the OLS regression over all data ignoring the sampling design. If access is available to software that accounts for survey design, having declared appropriate elements of the sampling design, the artificial regression:

$$(2\hat{F}(y_{hci}) - 1)\sqrt{y_{hci}} = \theta\sqrt{y_{hci}} + v_{hci}, \quad h = 1, \dots, H; c = 1, \dots, n_h; i = 1, \dots, k$$

yields $\hat{\theta} = \hat{G}$.

5.2 Computing Variances with a Complex Survey Sample

If a researcher is using a package designed for surveys¹⁹, then it is easy to calculate Binder and Kovačević's (1995) variance estimator

$$\widehat{Var}(\hat{G}) = \sum_{h=1}^H \left(\frac{n_h}{n_h - 1} \right) \sum_{c=1}^{n_h} (u_{hc}^* - \bar{u}_h^*)^2,$$

as this is the variance for a survey weighted total based on the approximation

$$\hat{G} - G \approx \sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} u_{hci}^*.$$

For such software packages, the variance (often called the linearization variance) is easily generated using an appropriate “total” command after forming the series u_{hci}^* , along with specifying the weight series and declaring strata and cluster identification variables. When each stratum contains a large number of clusters (as would be the case for many household surveys) there will be little difference between this estimator and its asymptotic version (i.e., that associated with the number of sampled clusters going to infinity at the same rate):

$$\widetilde{Var}(\hat{G}) = \sum_{h=1}^H \sum_{c=1}^{n_h} (u_{hc}^* - \bar{u}_h^*)^2,$$

¹⁹ Examples include Stata (StataCorp), SPSS (SPSS Inc.), SUDAAN (Research Triangle Institute) and the “survey” package developed by Lumley for R (see Lumley, 2010).

which, we recall, is the same as the variance derived by Bhattacharya (2007) (as shown in the Proof of Result 2 in the Appendix). However, should stratum contain few clusters, then it may be preferable to use Binder and Kovačević's (1995) variance estimator, as the scaling factor $(n_h/(n_h - 1))$ may then be useful.

When access is not available to survey software, one way to obtain Bhattacharya's variance estimator is to estimate the artificial regression:

$$u_{hc}^* = \sum_{h=1}^H \beta_h D_{hc} + v_{hc} \quad ,$$

where $D_{hc} = 1$ if cluster c is in stratum h , 0 otherwise. Let SSR be the sum of squared residuals from this regression. It follows that $\widehat{Var}(\hat{G}) = \sum_{h=1}^H \sum_{c=1}^{n_h} (u_{hc}^* - \bar{u}_h^*)^2 = SSR$. To undertake this regression, a researcher needs to initially generate $u_{hc}^* = \sum_{i=1}^k w_{hci} u_{hci}^*$, which only requires a few lines of code. Finally, Binder and Kovačević's (1995) variance estimator, $\widehat{Var}(\hat{G}) = \sum_{h=1}^H \left(\frac{n_h}{n_h-1}\right) \sum_{c=1}^{n_h} (u_{hc}^* - \bar{u}_h^*)^2$, can be generated without survey software, as the SSR from the artificial regression:

$$\left(\frac{n_h}{n_h-1}\right)^{0.5} u_{hc}^* = \sum_{h=1}^H \beta_h D_{hc}^* + v_{hc} \quad ,$$

where $D_{hc}^* = \sqrt{(n_h/(n_h - 1))} = 1$ if cluster c is in stratum h , 0 otherwise. It may seem surprising that $\widehat{Var}(\hat{G}) = SSR$ from an auxiliary regression, as the SSR does not go to zero asymptotically under a traditional LS framework, which is required here for consistency. However, recalling that $u_{hc}^* = \sum_{i=1}^k w_{hci} u_{hci}^*$ with $w_{hci} = \frac{W_{hci}}{\sum_{h=1}^H \sum_{c=1}^{n_h} \sum_{i=1}^k W_{hci}} = \frac{W_{hci}}{M}$ and $M = kN$, it is clear that $SSR \rightarrow 0$ as $N \rightarrow \infty$.

6. EMPIRICAL EXAMPLE: BMI INEQUALITY AMONG BANGLADESHI WOMEN

To illustrate results, we examine health inequality amongst Bangladeshi women employing body mass index (BMI) as our gauge of well-being. Data are from three Bangladesh Demographic and Health Surveys (BDHS): 2004, 2007 and 2011; NIPORT et al. (2005, 2009, 2013). Our adoption of such a measure of health status,

as an indicator of welfare, is part of a growing body of research that studies inequality using metrics of well-being aside from income or consumption expenditure, recognizing that attributes of an individual's life, such as education, health and longevity, may be just as important as income or consumption. Some studies that have considered inequality in BMI for developing countries are Araar et al. (2009), Sahn and Younger (2009) and Molini et al. (2010). Araar et al. (2009) study a sample of Namibian individuals, including children; Sahn and Younger (2009) compare inequality across 36 developing countries, based on Demographic Health Surveys and Living Standards Measurement Studies, with data from several surveys for each country; and Molini et al. (2010) contrast female BMIs for a cross-section of countries and male and female BMIs for Vietnam.

An individual's BMI is given by his or her weight in kilograms scaled by height in metres squared. Low adult BMI suggests inadequate access to food and possible diseases, leading to an increased likelihood of morbidity and mortality, birth difficulties and poor health in delivered infants, and reduced ability to work productively; see, for example, Molini et al. (2010) for discussion and the references cited therein. High adult BMI (obesity) may also be detrimental for an individual's well-being, being associated with high blood pressure, high cholesterol and triglycerides, type 2 diabetes, coronary artery disease, stroke, breathing problems such as sleep apnea and asthma, gallbladder disease and gallstones, osteoarthritis, and some cancers, among other conditions. That both low and high BMI are associated with a decline in well-being is an issue when calculating inequality indicators, as these measures assume a monotonic ranking of individuals. For two reasons we ignore this concern. First, as we shall see in section 6.1, although obesity is rising in Bangladesh, the proportion is still relatively small compared with the developed world, suggesting that inequality indicators will not be too distorted with the inclusion of obese women. Secondly, as discussed by, for instance, Molini et al. (2010), it is unlikely that the obese BMI observations will severely skew the BMI distribution, in contrast to high incomes in that distribution.

To construct Gini coefficients using our BMI data, we need to ensure comparability across women of different ages, as our sample comprises women aged

15-49. We allow for natural age effects by standardizing the BMIs using growth curves provided by the World Health Organization (WHO); see de Onis et al. (2007). Specifically, as the WHO finds that the distribution of BMIs for 19 year olds matches that for adult women, we use WHO growth curves to convert the BMIs for women aged 15 through 18 to an “equivalent” BMI for a 19 year old woman. By “equivalent”, we mean that the woman’s position in the age specific BMI distribution is maintained. Let BMI_i be the actual sample value for the i^{th} woman with associated cumulative distribution function $F_a(BMI_i)$ for age a , calculated using the WHO growth curve. Then, let $F_{19}(BMI_{19,i})$ be the cumulative distribution function from the WHO growth curve for a 19 year old woman with standardized BMI given by $BMI_{19,i}$. For those women with $a = 15, \dots, 18$, we generate $BMI_{19,i}$ such that $F_a(BMI_i) = F_{19}(BMI_{19,i})$, our notion of “equivalent”, with these “standardized” BMIs being used in statistics generated for the sample²⁰. The number of “standardized” observations is 1295 (954; 1147) for the 2011 (2007; 2004) datasets. Hereafter, we use “BMI” to denote the series that includes the “standardized” BMI numbers for those women younger than 19 years of age.

In addition to estimating the Gini coefficient and its variance for ever-married women in each dataset, we also report Gini coefficients when data are divided by region of residence (urban; rural) and by wealth category (poorest; poorer; middle; richer; richest). We consider region of residence as urbanization is often cited as being a contributing factor to increasing BMIs in developing countries, with urban residents often cited to having predominantly higher BMIs than that of rural residents; e.g., see Nubé et al. (1998), Popkin and Gordon-Larson (2004), Mendez et al. (2005), Shafique et al. (2007), Neuman et al. (2013). This may result in differing inequality in BMI between urban and rural women. Subdividing data by wealth category enables us to consider the impact of a common measure of socioeconomic status on inequality in health and well-being, as measured by BMI; e.g., Mendez et al. (2005), Corsi et al. (2011), Subramanian et al. (2011), Razak et al. (2013). A women’s health

²⁰ See, for example, de Onis et al. (2007) and Araar et al. (2009) for descriptions of the methods we use to calculate the age specific distribution functions with relevant, so-called, LMS coefficients, obtained from http://www.who.int/childgrowth/standards/bmi_for_age/en/index.html, last accessed 14 January 2014.

likely varies substantially with wealth, affecting, for instance, her ability to acquire adequate and nutritious food supplies, have access to appropriate health services, safe sources of water and safe sanitation facilities, and her knowledge of diseases; see, for instance, Banerjee et al. (2004), Araar et al. (2009), Neuman et al. (2011).

6.1 Data, Survey Designs and Summary Statistics

The raw BMI data are from the 2004, 2007 and 2011 Bangladesh Demographic and Health Surveys (BDHS), nationally representative sample surveys that aim to detail statistics useful for ascertaining changes in key areas of development, including maternal and child health, domestic violence, education and poverty reduction; see, NIPORT et al. (2005, 2009, 2013). These are the fourth, fifth and sixth such surveys for this developing country, conducted through collaboration between the National Institute of Population Research and Training (NIPORT) of the Ministry of Health and Family Welfare, ICF International/Macro International/ORC Macro, and Mitra and Associates. The 2011 (2007; 2004) survey consists of information from ever-married women aged 12-49 (15-49; 10-49). The 2007 survey originally included ever-married women aged 10-49 but those aged 10-14 were dropped due to the small number of women in this age range. Accordingly, for consistency across surveys, we limited attention to those ever-married women aged 15-49.

Data were obtained using stratified multistage cluster sampling with somewhat similar designs across surveys. We provide a brief outline here, referring the reader to NIPORT et al. (2005, 2009, 2013) for full details. Bangladesh is comprised of administrative divisions²¹ that are split into zilas with zilas further subdivided into upazilas. In rural regions, upazilas are split into union parishads, with these districts further divided into mouzas. Urban region upazilas contain wards, which are subdivided into mahallas. These divisions enable the country to be stratified into rural and urban areas, with enumeration areas (EAs) used (mostly) as the primary sampling units (PSUs). The EAs correspond to mahallas in urban districts and mouzas in rural regions, with 100-120 households, on average, in each EA; large EAs

²¹ For the 2004 and 2007 surveys, there were six administrative divisions (Barisal, Chittagong, Dhaka, Khulna, Rajshahi and Sylhet), with an additional one existing at the time of the 2011 survey (Rangpur).

were further segmented with these segmented areas forming the PSU. Having stratified the country into rural and urban areas for each division, urban areas are further stratified into city corporations and other than city corporations for the 2011 survey, and statistical metropolitan areas, municipality areas and other urban areas for the 2004 and 2007 surveys. This resulted in 20 strata for the 2011 survey and 22 strata for the 2004 and 2007 surveys²².

The desired number of clusters was then selected, independently across strata, with probability proportional to the EA household size. Equal probability systematic selection was used to draw 30 households from each sampled cluster. As most of the population reside in rural areas, urban households were over-sampled to obtain a similar level of statistical precision to the rural regions²³. All ever-married women who were members of the selected household, and any ever-married woman who slept in the household the night before the survey, were eligible to be interviewed. Hence, the sampling design of our empirical example differs from the framework considered in the theoretical sections; here the clusters and households are chosen by other than simple random sampling, and we have potentially more than one ever-married woman from a household in the sample. Thus, the “i” (or “t”, as appropriate) subscript, in the expressions, is now over all ever-married women in the sampled cluster, rather than over households. The total number of sample observations is now the number of ever-married women. Such changes do not matter, as long as the “right” sampling weight is considered. Here we use the individual weight from the ever-married women survey file. Given the setup, only nonresponse leads to a difference between the individual sampling weight and the household sampling weight.

Wealth of a woman is determined by her household’s assignment in a categorical ranking: poorest, poorer, middle, richer and richest. These divisions are calculated via a wealth index, constructed using principal components analysis, which combines

²² We detected several coding mistakes in the raw 2007 BDHS data with respect to matching the strata with rural/urban classifications; we amended these and used the corrected data file throughout our study.

²³ In addition, as one of the BDHS objectives is to obtain estimates for each division, as well as for the country as a whole, some divisions were also oversampled. Such features mean that the BDHS samples are not self-weighting.

weighted values for indicator variables that reflect economic status, including household assets (e.g., television, bicycle, refrigerator, telephone, car, motor cycle, radio, livestock), utility services (e.g., electricity, sources of water, sanitation facilities, type of cooking fuel), materials used to construct housing (e.g., type of flooring, roofing material), persons per sleeping room, ownership of agricultural land, and whether the household has one or more domestic servants; see Rutstein and Johnson (2004) for details.

Turning to our sample data, in Table 1 we report BMI summary statistics, including the percentages for each sample that fall within traditional WHO classifications of BMI. In particular, a BMI < 18.5 represents chronic energy deficiency. Evidently, BMI has increased over time with slightly more variation, as

Table 1. Summary statistics for the three BDHS surveys: ever-married women

	2004	2007	2011
Number of observations	11,166	10,836	17,309
Number of clusters (rural; urban)	(239;122)	(228;136)	(393;207)
Range of number of clusters in each stratum (smallest; largest; number of strata)	(2; 57; 22)	(3; 46; 22)	(5; 61; 20)
Mean BMI (std. error)	20.26 (0.06)	20.73 (0.07)	21.45 (0.06)
BMI std. dev.	3.32	3.52	3.80
Underweight (<18.50)	31.5%	27.2%	22.4%
Severe thinness (<16.00)	5.2%	4.2%	3.4%
Moderate thinness (16.00-<17.00)	7.5%	6.4%	5.4%
Mild thinness (17.00-<18.50)	18.8%	16.6%	13.6%
Normal weight (18.50-24.99)	57.9%	58.9%	59.8%
Overweight (≥ 25.00)	10.6%	13.9%	17.8%
Pre-obese (25.00-<30.00)	8.8%	11.6%	14.6%
Obese (≥ 30.00)	1.8%	2.3%	3.2%
Obese class I (30.00-<35.00)	1.6%	1.9%	2.7%
Obese class II (35.00-<40.00)	0.1%	0.3%	0.4%
Obese class III (≥ 40.00)	0.1%	0.1%	0.1%

Notes: All statistics are calculated using the standardized BMI data. The mean statistics account for the survey design, with the common survey-based estimator (see, e.g., Skinner et al., 1989, p.47) employed to estimate variances. This variance estimator is equivalent to expression (4.1) applied to the BMI data.

indicated by the increase in the standard deviation. The mean BMIs are statistically significantly higher from each survey to the next, at common choices of significance level using a one-sided test. Over time, we see proportionally fewer women in each of the thinness categories, an indication that perhaps some gains have been made on malnutrition, suggesting that Bangladesh has made progress on Millennium Development Goal 1, particularly Target 1C.²⁴ However, that 1 in 5 women remain underweight is troubling, as is also noted in a report on Bangladesh’s progress on

Table 2. Summary statistics for the three BDHS surveys: by urban and rural divisions

	2004	2007	2011
Urban Women			
Number of observations	3,816	4,118	6,024
Mean BMI (std. error)	21.59 (0.14)	22.29 (0.17)	23.03 (0.12)
BMI std. dev.	5.00	5.42	4.86
Underweight (<18.50)	24.0%	19.5%	15.0%
Normal weight (18.50-<25.00)	56.3%	57.0%	56.9%
Overweight (≥ 25.00)	19.7%	23.5%	28.2%
Rural Women			
Number of observations	7,350	6,718	11,285
Mean BMI (std. error)	19.87 (0.06)	20.27 (0.07)	20.90 (0.06)
BMI std. dev.	2.72	2.83	3.27
Underweight (<18.50)	35.4%	32.0%	26.5%
Normal weight (18.50-<25.00)	58.8%	60.1%	61.3%
Overweight (≥ 25.00)	5.8%	8.0%	12.2%

Notes: All statistics are calculated using the standardized BMI data. The mean statistics account for the survey design, with the common survey-based estimator (see, e.g., Skinner et al., 1989, p.47) employed to estimate variances. This variance estimator is equivalent to expression (4.1) applied to the BMI data.

reaching the Millennium Development Goals (Bangladesh Planning Commission, 2012). Nearly 60% of women are in the normal BMI range. On the other hand, although only a small percentage of women are obese, the proportion of women who

²⁴ Goal 1 is “Eradicate extreme poverty and hunger” and Target 1C is “Halve between 1990 and 2015, the proportion of people who suffer from hunger.”

are overweight or obese has nearly doubled over the time span, being, in 2011, not much less than the proportion of women who are underweight. This feature of simultaneously observing a significant percentage of the population being underweight and overweight coincides with evidence from other developing countries; e.g., for Namibia (Arrar et al. 2009) and for Kenya (Jayne et al. 2011), and see also Caballero (2005), Prentice (2006) and Razak et al. (2013). Although an investigation of the specific causes of this trend for Bangladeshi women is beyond our scope, we anticipate that growing urbanization, more sedentary lifestyles, and higher consumption of energy-dense foods, forms part of the story, as has been hypothesized for other developing countries; e.g., Martorell et al. (2000). The double burden of having women who are underweight and overweight is a dilemma for health authorities and the health care system.

Summary statistics are provided in Tables 2 and 3 when the sample data are divided by region of residence (urban and rural) and wealth category. We detail percentages in three broad BMI categories: underweight ($BMI < 18.5$), normal ($18.50 \leq BMI < 25.00$) and overweight ($BMI \geq 25.00$), along with means and standard deviations. Mean BMIs are higher for urban women than for rural women, with the mean BMIs increasing across time for both rural and urban women. At a nominal 5% significance level, with a one-sided test, the mean BMIs are statistically significantly higher for urban women than for rural women, and higher for each successive survey for a given region of residence, except when comparing 2011 with 2007 for rural women. Chronic energy deficiency ($BMI < 18.5$) is more widespread for rural women than for women residing in urban regions, whereas overweight and obesity ($BMI > 25.0$) is significantly more prevalent for women living in urban areas. Although there have been gains in reducing chronic energy deficiency for both urban and rural women, it is troubling that one in four rural women are still underweight.

The proportion of women in the BMI classifications across the different wealth categories varies quite substantially, irrespective of survey. Over or close to a third of the “poor” women are underweight, with only modest declines in the proportion of underweight “poor” women over the period covered by the three surveys. Only a small percentage of the poorest and poorer women are overweight, but this proportion

has more than doubled over the seven years. Although a woman in the “middle” wealth group is less likely to be underweight in contrast with her poorer counterpart, there is still a substantial prevalence of chronic energy deficiency for the “middle” wealth category women. In contrast, “rich” women are now about equally likely to be overweight as underweight, contrary to the situation in 2004 when three times as many women were underweight than overweight. Nearly 40% of the richest women were overweight in 2011, with less than 10% underweight. The degree of spread of BMI, as measured by the standard deviations, increases with wealth, for a given survey.

Mean BMIs have statistically significantly increased (nominal 5% significance level, one-sided test) from one survey to the next for each of the wealth categories, and the mean BMIs are also statistically significantly higher (nominal 5% significance level) for women in each wealth category compared with those in the poorer adjacent category (e.g., middle compared with poorer). Overall, these findings suggest that there is a strong positive correlation between wealth and BMI with low BMI being concentrated in the “poor” and high BMI being concentrated in the “rich” – malnutrition does not fall randomly across the population. Also, mean BMI is increasing over time, for each wealth group. Nevertheless, as it is feasible for a woman to be underweight in a “rich” household and overweight in a “poor” household, it is more than just wealth status that determines nutritional status. We also note that more than half of women have a BMI within the normal category, irrespective of wealth group and survey date, suggesting that malnutrition (underweight or overweight) is of concern for just under half of the women. We now turn to estimating Gini coefficients.

Table 3. Summary statistics for the three BDHS surveys: by wealth group

	2004	2007	2011
Poorest			
Number of observations	1,979	1,751	3,013
Mean BMI (std. error)	19.04 (0.06)	19.35 (0.09)	19.68 (0.07)
BMI std. dev.	2.37	2.42	2.81
Underweight (<18.50)	45.3%	41.7%	38.3%
Normal weight (18.50-<25.00)	52.7%	54.9%	56.7%
Overweight (≥ 25.00)	2.0%	3.4%	5.0%
Poorer			
Number of observations	1,997	1,964	3,235
Mean BMI (std. error)	19.39 (0.07)	19.76 (0.08)	20.34 (0.06)
BMI std. dev.	2.47	2.57	2.94
Underweight (<18.50)	39.5%	34.3%	28.3%
Normal weight (18.50-<25.00)	57.8%	61.5%	64.5%
Overweight (≥ 25.00)	2.8%	4.2%	7.2%
Middle			
Number of observations	2,092	2,070	3,328
Mean BMI (std. error)	19.88 (0.08)	20.24 (0.08)	21.02 (0.07)
BMI std. dev.	2.76	2.97	3.25
Underweight (<18.50)	34.6%	30.6%	24.0%
Normal weight (18.50-<25.00)	60.4%	62.4%	64.0%
Overweight (≥ 25.00)	5.0%	7.0%	12.0%
Richer			
Number of observations	2,230	2,173	3,670
Mean BMI (std. error)	20.54 (0.10)	20.99 (0.10)	21.94 (0.08)
BMI std. dev.	3.25	3.34	3.78
Underweight (<18.50)	29.8%	25.2%	18.7%
Normal weight (18.50-<25.00)	61.0%	62.9%	61.1%
Overweight (≥ 25.00)	9.2%	11.9%	20.2%
Richest			
Number of observations	2,868	2,878	4,063
Mean BMI (std. error)	22.38 (0.11)	23.15 (0.13)	23.96 (0.10)
BMI std. dev.	4.51	4.74	4.44
Underweight (<18.50)	15.6%	12.8%	8.4%
Normal weight (18.50-<25.00)	57.5%	54.0%	53.6%
Overweight (≥ 25.00)	27.0%	33.2%	38.0%

Notes: All statistics are calculated using the standardized BMI data. The mean statistics account for the survey design, with the common survey-based estimator (see, e.g., Skinner et al., 1989, p.47) employed to estimate variances. This variance estimator is equivalent to expression (4.1) applied to the BMI data.

6.2 Gini Coefficients and Variance Estimates: All Women

In the previous subsection, we discussed health status (as measured by BMI) in terms of means and the proportion of women with a BMI above or below a specified threshold. Such statistics ignore much of the BMI information. Here we estimate Gini coefficients for BMI; a value of $G = 0$ results when there is perfect equality, that is, all women have exactly the same BMI. In contrast, a value of $G = 1$ corresponds to the unrealistic case of perfect inequality; i.e., one woman has all the BMI. Table 4 provides the estimated Gini coefficients for all women in each survey, along with estimated variances using expressions (4.1) and (4.3); the former corresponds with that obtained by Binder and Kovačević (1995) and the latter to that derived by Bhattacharya (2007). Recall that the difference between them is a correction factor for the finite number of clusters in each stratum. We also report the breakdown of the variance estimates, using expression (4.3), into the three component parts: the variance arising from naively assuming a simple random sample (SRS) design with weights and the effects of clustering and stratifying on the variance. In addition, we provide Wald statistics, and associated Chi-square p-values, for testing equality of the Gini coefficients using the variances corrected for the complex survey sampling design, along with the incorrect Wald statistics (and p-values) that would result if we employed the SRS with weights variance formula.²⁵

The Gini estimates show that there is inequality in health, as measured by BMI, with inequality growing statistically significantly over the three surveys. As there will always be variations in women's BMI, even in a healthy population with equality in health, as measured by the absence of malnutrition (under- or over-weight), we calculated Gini estimates for those Bangladeshi women whose standardized BMI was in the "normal" range (18.50-<25.00). For the three survey years, these Gini estimates are, respectively, 0.048, 0.047 and 0.046; the observed estimates for all ever-married women are approximately double these values. Malnutrition is resulting in significant increases in inequality. How do these estimates compare with other countries? Interestingly, the absolute magnitudes of the Gini estimates for

²⁵ We assume that the samples across surveys are independent, a reasonable assumption given that clusters are randomly selected.

Bangladesh are approximately thirty percent lower than the Gini estimate reported by Araar et al. (2009) for Namibian females, but this may be due to the fact that their sample includes children, aside from any inequality concerns. Further, the Bangladeshi indices are lower than those reported by Contoyannis and Wildman (2007) for females from two developed countries, Canada and England.

Table 4. Gini coefficient estimates, estimated variances, design effects and hypothesis tests: ever-married women

Survey	\hat{G}	$\widehat{Var}(\hat{G})$	$\widetilde{Var}(\hat{G})$	$\widetilde{Var}_{SRS}(\hat{G})$	Design effect	Stratum Effect	Cluster Effect
2004	0.088	0.880	0.806	0.529	1.524	0.644	0.922
2007	0.092	1.114	1.056	0.591	1.787	0.471	0.936
2011	0.097	0.607	0.589	0.369	1.596	0.170	0.390
Hypothesis Tests							
Null hypothesis	Change in \hat{G}	Using $\widetilde{Var}(\hat{G})$		Using $\widetilde{Var}_{SRS}(\hat{G})$			
		Wald statistic	p-value	Wald statistic	p-value		
2007=2004	0.004	8.593	0.003	14.286	0.000		
2011=2007	0.005	15.198	0.000	26.042	0.000		
2011=2007=2004	n.a.	59.096	0.000	92.297	0.000		

Notes: Variance estimates and associated components have been scaled by 10^6 . $\widehat{Var}(\hat{G})$ is Binder and Kovačević's (1995) estimator, see expression (4.1). $\widetilde{Var}(\hat{G})$ is Bhattacharya's (2007) estimator, see expression (4.3). $\widetilde{Var}_{SRS}(\hat{G})$ is the first term of expression (4.3), the variance estimator under an assumption of SRS with weights. The "design effect" provides the ratio of $\widetilde{Var}(\hat{G})$ to $\widetilde{Var}_{SRS}(\hat{G})$. The "stratum effect" and "cluster effect" are, respectively, the second and third terms of expression (4.3).

Turning to variance estimates, we see some differences from using Bhattacharya's (2007) and Binder and Kovačević's (1995) formulae, with $\widehat{Var}(\hat{G})$ being between 3% and 10% higher than $\widetilde{Var}(\hat{G})$. As some strata have few clusters, the correction factor $(n_h/(n_h - 1))$ in expression (4.1) can markedly differ from unity for these strata, implying that for samples that do have relatively few clusters in some of the strata, the asymptotic variance, $\widetilde{Var}(\hat{G})$, may understate the variance. Then, in practice, it may be preferable to adopt $\widehat{Var}(\hat{G})$.

Further, allowing for the survey design²⁶ leads to more than a 50% increase in the variance than under an incorrect assumption of a simple random sample with weights. The magnitude of the design effects is consistent with those for other statistics from the BDHS surveys, as reported in the appendices of the final reports (see NIPORT 2005; 2009; 2013). These results show the importance of accounting for the survey design when estimating variances. We see the appreciatively large contribution of clustering and the smaller offset gain in precision by stratifying. The impact of increasing the sample size and the number of clusters is evident when comparing the variances, and the design components, from the 2011 survey with the earlier surveys – over 50% more women were surveyed for the 2011 survey than with the earlier surveys and the number of clusters sampled nearly doubled. The larger number of women in each stratum for the 2011 survey likely reduces homogeneity, resulting in smaller gains in precision as shown by a relatively smaller stratum effect, compared with the earlier surveys. In contrast, sampling many more clusters likely adds useful information that is reflected in a smaller cluster effect for the 2011 survey. As expected, the SRS variance decreases with the larger sample size.

For our cases, taking account of the survey design on the variances for the Gini coefficient estimators does not alter the outcomes of the hypothesis tests. Here, we reject that the Gini coefficients are equal from one survey to the next, and that they are jointly equal across the three surveys. There is a significant increase in BMI inequality over the three surveys. This contrasts to changes in Gini coefficient estimates for inequality in income over the same period, which has shown a small decline.²⁷ We next consider results when the data are subdivided by broad region of residence.

²⁶ In some studies, the design effect is relative to the SRS variance without weights; i.e., that assumes each observation has the same probability of being selected. As SRS with weights is commonly undertaken in inequality work, we have allowed for the survey weights when forming $\tilde{var}_{SRS}(\hat{G})$. Such an approach also enables a straightforward decomposition of the overall variance.

²⁷ As reported by The World Bank; see <http://data.worldbank.org/indicator/SI.POV.GINI/countries?display=default>, last accessed 14 January 2014.

6.3 Gini Coefficients and Variance Estimates: Rural and Urban Regions

This subsection reports Gini coefficient estimates and estimated variances when we subdivide the population into the region (urban/rural) of residence for each woman. For space reasons, we report only $\widehat{Var}(\hat{G})$, $\widetilde{Var}(\hat{G})$ and $\widetilde{Var}_{SRS}(\hat{G})$, concentrating our attention on hypothesis tests for equality of Gini coefficients, which use $\widetilde{Var}(\hat{G})$, across regions for a given survey and across surveys for a given region. Results are provided in Tables 5 and 6. Nutritional inequalities are higher for urban women than for rural women, practically and statistically significantly. Interestingly, inequality has increased amongst women living in rural regions across the surveys, whereas it remained relatively stable amongst urban women, even declining marginally from 2007 to 2011.

Table 5. Gini coefficient estimates, estimated variances and design effects: by region of residence

Survey	\hat{G}		$\widehat{Var}(\hat{G})$		$\widetilde{Var}(\hat{G})$		$\widetilde{Var}_{SRS}(\hat{G})$		Design effect	
	U	R	U	R	U	R	U	R	U	R
2004	0.104	0.081	2.422	1.214	1.981	1.188	1.806	0.691	1.097	1.719
2007	0.104	0.085	2.402	1.733	2.169	1.689	1.659	0.843	1.307	2.004
2011	0.101	0.092	1.814	0.750	1.736	0.738	1.307	0.470	1.328	1.570

Notes: Variance estimates have been scaled by 10^6 . “U” and “R” denote Urban and Rural, respectively; $\widehat{Var}(\hat{G})$ is Binder & Kovačević’s (1995) estimator (expression (4.1)), $\widetilde{Var}(\hat{G})$ is Bhattacharya’s (2007) estimator (expression (4.3)); $\widetilde{Var}_{SRS}(\hat{G})$ is the first term of expression (4.3), the variance estimator under an assumption of SRS with weights. The “design effect” provides the ratio of $\widetilde{Var}(\hat{G})$ to $\widetilde{Var}_{SRS}(\hat{G})$.

Variance estimates are substantially higher for the urban Gini coefficient estimators than the rural ones, most likely due to the smaller number of clusters (and number of women) in each urban stratum, along with the variance effect from nearly twice the number of rural women, over urban women, being sampled in each survey. In addition, design effects are relatively larger for rural areas than urban areas,

suggesting, *ceteris paribus*, that there is a higher positive intracluster correlation for rural women than for urban women with respect to BMI.²⁸

Two-sided hypothesis tests for urban women, reported in Table 6, suggest equality of Gini coefficients between 2004 and 2007, with the change in inequality between 2007 and 2011 being significantly different from zero, at least at the nominal 10% significance level. Overall, however, we support that the urban Gini coefficients are jointly equal across the three surveys. In contrast, the changes in Gini coefficients between surveys for rural women are statistically significantly different from zero. We also reject that the rural Gini coefficients are equal across the three surveys. In addition, results suggest that the urban and rural Gini coefficients are unequal for each survey. Such features are not evident from the summary statistics reported in Table 2, indicating the usefulness of examining inequality measures, along with commonly reported summary measures when considering BMI.

Table 6. Hypothesis tests for equality of Gini coefficients: by region of residence

Test	Across surveys		
	Urban Wald statistic (p-value)	Rural Wald statistic (p-value)	
2004=2007	0.101 (0.751)	5.561 (0.018)	
2007=2011	2.783 (0.095)	19.943 (0.000)	
2004=2007=2011	3.259 (0.196)	65.196 (0.000)	
Test	Across regions		
	2004 Wald statistic (p-value)	2007 Wald statistic (p-value)	2011 Wald statistic (p-value)
urban=rural	161.886 (0.000)	96.514 (0.000)	32.348 (0.000)

Notes: Statistics are formed using $\widehat{Var}(\hat{G})$.

²⁸ All else the same, the size of the design effect is also impacted by the variability of the sampling weights. For our cases variability of the sampling weights is higher for the urban subsamples than for rural ones.

6.4 Gini Coefficients and Variance Estimates: Wealth Categories

In this subsection we provide, in Table 7, estimated Gini coefficients, and associated variances²⁹, corrected for the survey design, for women’s BMI when the women are sorted by wealth category: poorest, poorer, middle, richer and richest. To illustrate hypothesis testing, we report Wald statistics, and corresponding Chi-square p-values, that examine for equality of Gini coefficients across the three surveys by wealth category in Table 8. The results in Table 7 highlight the usefulness of examining subgroups when considering inequality – irrespective of survey, inequality in women’s BMI increases with socioeconomic status, as measured by wealth asset status. This is as expected.

Table 7. Gini coefficient estimates and estimated variances: by wealth category

Wealth category	$\hat{G} (\widehat{var}(\hat{G}); \widetilde{var}(\hat{G}))$		
	2004	2007	2011
Poorest	0.069 (2.309;2.200)	0.073 (3.066;2.998)	0.081 (1.933;1.905)
Poorer	0.073 (2.584;2.527)	0.075 (2.570;2.487)	0.081 (1.483;1.457)
Middle	0.077 (1.931;1.868)	0.081 (3.301;3.202)	0.088 (1.573;1.543)
Richer	0.088 (3.966;3.826)	0.089 (2.823;2.713)	0.095 (1.761;1.718)
Richest	0.100 (2.237;1.937)	0.101 (2.648;2.518)	0.098 (2.130;2.055)

Notes: Variance estimates have been scaled by 10^6 . $\widehat{var}(\hat{G})$ ($\widetilde{var}(\hat{G})$) is Binder & Kovačević’s (1995) (Bhattacharya’s (2007)) estimator.

In addition, for a given wealth subgroup (except for the richest subgroup), inequality has numerically increased over the surveys, especially between 2007 and 2011, with most of the changes, but not all, being statistically significantly different.³⁰ We see, for instance, using the hypothesis test outcomes reported in Table 8, that the Gini coefficient has statistically increased for the “poorest” and “middle” subgroups across all surveys but not for the “richest” subgroup. And, although there is no

²⁹ The breakdowns of the complex survey variances into the component parts (SRS, cluster, stratum effects) are available on request.

³⁰ At the nominal 10% significance level. Note that for some of the cases, we would conclude statistical significance in the change in the Gini coefficient had we examined an appropriate one-sided alternative hypothesis.

statistical change in the Gini coefficient between 2004 and 2007 for the “poorer” and “richer” women, the changes are statistically significantly different from zero when we compare the 2007 and 2011 Gini coefficients for these subgroups of women. Such features, along with the summary statistics reported in Table 3, suggest that, typically, the prevalence of overweight is increasing at a faster rate than the decline in the prevalence of underweight, leading to higher inequality. In other words, although mean BMIs have increased over time for women in each wealth category, the weight gain is distributed unevenly leading to a rise in inequality. This undesirable outcome suggests that Bangladesh faces rising issues associated with women being overweight and obese, along with difficulties in decreasing chronic energy deficiency for a large proportion of women.

Table 8. Hypothesis tests for equality of Gini coefficients: by wealth category

Hypothesis Test	Wealth Category				
	Poorest	Poorer	Middle	Richer	Richest
2004=2007	3.078 (0.079)	0.798 (0.372)	3.156 (0.076)	0.153 (0.696)	0.224 (0.636)
2007=2011	13.053 (0.000)	9.128 (0.003)	10.327 (0.001)	8.125 (0.004)	1.968 (0.161)
2004=2007=2011	36.550 (0.000)	18.835 (0.000)	36.486 (0.000)	12.603 (0.002)	2.107 (0.349)

Note: The table reports Wald statistics for testing the appropriate null hypothesis of equality with Chi-square p-values in parentheses. Statistics are formed using $\widehat{Var}(\hat{G})$.

7. CONCLUSIONS

In this paper, we examine variance estimation for Gini coefficients calculated from samples obtained under complex survey designs. We show that a relatively recently proposed variance estimator (Bhattacharya, 2007) is equivalent to an estimator derived by Binder and Kovačević (1995), in an article published in the survey literature over a decade earlier. A key advantage of this equivalence result is that the variance formula provided by Binder and Kovačević, along with the approximation for the Gini estimator, is, we believe, far easier to practically calculate, of importance

to practitioners who often resort to resampling methods for variance estimation, under the belief that it is too computationally burdensome to estimate a variance obtained from asymptotic approximations. As an *iid* sample can be treated as a special case of a complex survey sample, we also link Davidson's (2009) work to the earlier research undertaken by Binder and Kovačević, showing that Davidson's derived approximation for the Gini estimator and his proposed variance estimator also follow directly from Binder and Kovačević's results.

In addition to linking econometric research with past survey literature research, we believe that our work dismisses the folklore that asymptotic variances for Gini indices, especially with complex survey data, are computationally burdensome to calculate. This is not the case; asymptotic variances can be readily calculated, even for those without access to specialized complex survey software.

We provide applied researchers with easily implementable ways to calculate both a Gini coefficient estimator and an estimator of its associated variance.

Our empirical application, using BMI data for Bangladeshi women, illustrates the importance of accounting for the complex survey design when forming variances. Corrected variances are much higher than would have been obtained under a simple random sampling (with weights) assumption. Overall, for all ever-married women we detect increasing BMI inequality across the surveys. Urban inequality, typically, has remained (statistically) unchanged, whereas rural inequality has (statistically) increased. There still remains, nevertheless, higher inequality in BMI for urban women than for rural women. BMI inequality increases with wealth, but has also been increasing over time for all wealth categories, aside from the "richest" cohort. Our findings suggest that Bangladesh faces growing issues associated with dealing with the dual health problems of chronic low BMI and increasing high BMI.

ACKNOWLEDGMENTS

We thank David Giles, Laura Cowen and Nilanjana Roy for helpful comments and suggestions, and we gratefully acknowledge Macro International for allowing us access to the data for this research. Useful comments were obtained from participants at a meeting of the Canadian Economics Association. This paper is based on work in Ahmed Hoque's Ph.D. dissertation.

REFERENCES

- Araar, A., S. Levine and J.-Y. Duclos (2009). Body mass index, poverty and inequality in Namibia. Paper prepared for the Namibian Central Bureau of Statistics.
- Banerjee, A., A. Deaton and E. Duflo (2004). Wealth, health, and health services in rural Rajasthan. *The American Economic Review* 94, 326-30.
- Bangladesh Planning commission (2012). *The Millennium Development Goals: Bangladesh Progress Report 2011*. General Economics Division, Bangladesh Planning Commission, Government of the People's Republic of Bangladesh, February 2012.
- Bhattacharya, D. (2005). Asymptotic inference from multistage samples. *Journal of Econometrics* 126, 145-71.
- Bhattacharya, D. (2007). Inference on inequality from household survey data. *Journal of Econometrics* 137, 674-707.
- Beach, C.M. and R. Davidson (1983). Distribution-free statistical inference with Lorenz curves and income shares. *The Review of Economic Studies* 50, 723-35.
- Biewen, M. and S.P. Jenkins (2006). Variance estimation for generalized entropy and Atkinson inequality indices: the complex survey data case. *Oxford Bulletin of Economics and Statistics* 68, 371-83.
- Binder, D.A. (1991). Use of estimating functions for interval estimation from complex surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 34-42.
- Binder, D.A. and M.S. Kovačević (1995). Estimating some measures of income inequality from survey data: an application of the estimating equations approach. *Survey Methodology* 21, 137-45.

- Binder, D.A. and Z. Patak (1994). Use of estimating functions for interval estimation from complex surveys. *Journal of the American Statistical Association* 89, 1035-44.
- Bishop, J.A., J.P. Formby and B. Zheng (1997). Statistical index and the Sen index of poverty. *International Economic Review* 38, 381-7.
- Caballero, B. (2005). A nutrition paradox – underweight and obesity in developing countries. *New England Journal of Medicine* 352, 1514-6.
- Cameron, A.C. and P.K. Trivedi (2005). *Microeconometrics: Methods and Applications*, New York: Cambridge University Press.
- Clarke, J.A. and N. Roy (2012). On statistical inference for inequality measures calculated from complex survey data. *Empirical Economics* 43, 499-524.
- Cochran, W.G. (1977). *Sampling Techniques 3rd Edition*, New York: John Wiley & Sons.
- Contoyannis, P. And J. Wildman (2007). Using relative distributions to investigate the Body Mass Index in England and Canada. *Health Economics* 16, 929-44.
- Corsi, D.J., H. H. Kyu and S.V. Subramanian (2011). Socioeconomic and geographic patterning of under- and over-nutrition among women in Bangladesh. *Journal of Nutrition* 141, 631-8.
- Davidson, R. (2009). Reliable inference for the Gini index. *Journal of Econometrics* 150, 30-40.
- de Onis, M., A.W. Onyango, E. Borghi, A. Siyam, C. Nishida and J. Siekmann (2007). Development of a WHO growth reference for school-aged children and adolescents, *Bulletin of the World Health Organization* 85, 660-7.
- Francisco, C.A. and W.A. Fuller (1991). Quantile estimation with a complex survey design. *Annals of Statistics* 19, 454-69.
- Gastwirth, J.L. and M. Gail (1985). Simple asymptotically distribution-free methods for comparing Lorenz curves and Gini indices obtained from complete data. In R. Basmann and G. Rhodes (Eds.), *Advances in Econometrics, Volume 4*, 229-43. Greenwich, CT: JAI Press.
- Giles, D.E.A. (2004). Calculating a standard error for the Gini coefficient: some further results. *Oxford Bulletin of Economics and Statistics* 66, 425-33.

- Gini, C. (1914). Sulla misura della concentrazione e della variabilità dei caratteri. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti LXXIII*, 1203-1248. Republished in English as Gini, C. (2005). On the measurement of concentration and variability of characters. *Metron LXIII*, 3-38.
- Glasser, G.J. (1962). Variance formulas for the mean difference and coefficient of concentration. *Journal of the American Statistical Association* 57, 648-54.
- Godambe, V.P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* 31, 1208-11.
- Godambe, V.P. and M. Thompson (1978). Some aspects of the theory of estimating equations. *Journal of Statistical Planning and Inference* 2, 95-104.
- Godambe, V.P. and M. Thompson (1984). Robust estimation through estimating equations. *Biometrika* 71, 115-25.
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* 19, 293-325.
- Jayne, J., A.G. Scrimgeour, M.E. Polhemus, L. Otieno and M.E. Bouill (2011). Dietary and socio-economic correlates of nutritional status in rural adult Kenyan population. *African Journal of Food, Agriculture, Nutrition and Development* 11 (4). Available from <http://www.ajfand.net/Volume11/No4/index4.html>, last accessed 16 January 2014.
- Jenkins, S.P. (2006). Estimation and interpretation of measures of inequality, poverty and social welfare using Stata. Presentation at the North American Stata User's Group Meeting 2006, Boston Massachusetts.
- Karagiannis, E. and M.S. Kovačević (2000). A method to calculate the jackknife variance estimator for the Gini coefficient. *Oxford Bulletin of Economics and Statistics* 62, 119-22.
- Kovačević, M.S. and D.A. Binder (1997). Variance estimation for measures of income inequality and polarization – the estimating equations approach. *Journal of Official Statistics* 13, 41-58.
- Krewski, D. and J.N.K. Rao (1981). Inference from stratified samples: properties of linearization, jackknife and balanced repeated replication methods. *Annals of Statistics* 9, 1010-9.
- Langel, M. and Y. Tillé (2013). Variance estimation of the Gini index: revisiting a result several times published. *Journal of the Royal Statistical Society (A)* 176, 521-40.

- Lorenz, M.O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association* 9, 209-19.
- Lumley, T. (2010). *Complex Surveys: A Guide to Analysis using R*, New York: John Wiley & Sons.
- Martorell, R., L. Kettel Khan, M.L. Hughes and L.M. Grummer-Strawn (2004). Obesity in women from developing countries. *European Journal of Clinical Nutrition* 54, 247-52.
- Mendez, M.A., C.A. Monteiro and B.M. Popkin (2005). Overweight exceeds underweight among women in most developing countries. *American Journal of Clinical Nutrition* 81, 714-21.
- Modarres, R. and J. Gastwirth (2006). A cautionary note on estimating the standard error of the Gini index of inequality. *Oxford Bulletin of Economics and Statistics* 68, 385-90.
- Molini, V., M. Nubé and B. Van den Boom (2010). Adult BMI as a health and nutritional inequality measure: applications at macro and micro levels. *World Development* 38, 1012-23.
- National Institute of Population Research and Training (NIPORT), Mitra and Associates, and ORC Macro (2005). *Bangladesh Demographic and Health Survey 2004*, Dhaka, Bangladesh and Calverton, Maryland (USA): National Institute of Population Research and Training, Mitra and Associates and ORC Macro.
- National Institute of Population Research and Training (NIPORT), Mitra and Associates, and Macro International (2009). *Bangladesh Demographic and Health Survey 2007*, Dhaka, Bangladesh and Calverton, Maryland (USA): National Institute of Population Research and Training, Mitra and Associates and Macro International.
- National Institute of Population Research and Training (NIPORT), Mitra and Associates, and ICF International (2013). *Bangladesh Demographic and Health Survey 2011*, Dhaka, Bangladesh and Calverton, Maryland (USA): National Institute of Population Research and Training, Mitra Associates and ICF International.
- Neuman, M., J.E. Finlay, G.D. Smith and S.V. Subramanian (2011). The poor stay thinner: stable socioeconomic gradients in BMI among women in lower- and middle-income countries. *The American Journal of Clinical Nutrition* 94, 1348-57.

- Neuman, M., I. Kawachi, S. Gortmaker and S.V. Subramanian (2013). Urban-rural differences in BMI in low- and middle-income countries: the role of socioeconomic status. *The American Journal of Clinical Nutrition* 97, 428-36.
- Nho, E. (2006). Statistical tests on the regional income inequality in Korea. *The Korean Economic Review* 22, 341-65.
- Nubé, M., W.K. Asenso-Okyere and G.J.M. van den Boom (1998). Body mass index as indicator of standard of living in developing countries. *European Journal of Clinical Nutrition* 52, 136-44.
- Nygård, F. and A. Sandström (1989). Inequality measures based on sample surveys. *Journal of Econometrics* 42, 81-95.
- Ogwang, T. (2000). A convenient method of computing the Gini index and its standard error. *Oxford Bulletin of Economics and Statistics* 62, 123-8.
- Popkin, B.M. and P. Gordon-Larson (2004). The nutrition transition: worldwide obesity dynamics and their determinants. *International Journal of Obesity* 28 (S3), S2-S9.
- Prentice, A.M. (2006). The emerging epidemic of obesity in developing countries. *International Journal of Epidemiology* 35, 93-9.
- Qin, Y., J.K.N. Rao and C.F. Wu (2010). Empirical likelihood confidence intervals for the Gini measure of income inequality. *Economic Modelling* 27, 1429-35.
- Rao, J.N.K. and C.F. Wu (1985), Inference from stratified samples: second-order analysis of three methods for nonlinear statistics. *Journal of the American Statistical Association* 80, 620-30.
- Razak, F., D.V. Corsi and S.V. Submamanian (2013). Change in the body mass index distribution for women: analysis of surveys from 37 low- and middle-income countries. *PLoS Medicine* 10: e1001367. DOI: 10.1371/journal.pmed.100367.
- Rutstein, S.O. and K. Johnson (2004). *The DHS Wealth Index*, Calverton, Maryland: ORC Macro.
- Sahn, D.E. and S.D. Younger (2009). Measuring intra-household health inequality: explorations using the body mass index. *Health Economics* 18, S13-S36.
- Sandström, A., J.H. Wretman and B. Waldén (1985). Variance estimators of the Gini coefficient: simple random sampling. *Metron* 43, 41-70.

- Sandström, A., J.H. Wretman and B. Waldén (1988). Variance estimators of the Gini coefficient: probability sampling. *Journal of Business & Economic Statistics* 6, 113-9.
- Schechtman, E. and S. Yitzhaki (1987). A measure of association based on Gini's mean difference. *Communications in Statistics – Theory and Methods* 16, 207-31.
- Sendler, W. (1979). On statistical inference in concentration measurement. *Metrika* 26, 109-22.
- Shafique, S., N. Akhter, G. Stallkamp, S. De Pee, D. Pangides and M.W. Bloem (2007). Trends of under- and overweight among rural and urban poor women indicate the double burden of malnutrition in Bangladesh. *International Journal of Epidemiology* 36, 449-57.
- Shao, J. (1994). L-statistics in complex survey problems. *Annals of Statistics* 22, 946-67.
- Skinner, C., D. Holt and T. Smith (1989). *Analysis of Complex Surveys*, New York: John Wiley & Sons.
- Slater, J., C. Green, G. Sevenhuyes, J. O'Neil and B. Edginton (2009). Socio-demographic and geographic analysis of overweight and obesity in Canada adults using the Canadian Community Health Survey (2005). *Chronic Diseases in Canada* 30, 3-14.
- Subramanian, S.V., J.M. Perkins, E. Özaltin and G.D. Smith (2011). Weight of nations: a socioeconomic analysis of women in low- to middle-income countries. *American Journal of Clinical Nutrition* 93, 413-21.
- Wolter, K.M. (2007). *Introduction to Variance Estimation 2nd Edition*, New York: Springer Verlag.
- Xu, K. (2007). U-statistics and their asymptotic results for some inequality and poverty measures. *Econometric Reviews* 26, 567-77.
- Yitzhaki, S. (1991). Calculating jackknife variance estimators for parameters of the Gini method. *Journal of Business & Economic Statistics* 9, 235-9.

APPENDIX: PROOFS OF RESULTS

Proof of Result 1: Here we show equivalence of $\hat{\Psi}_{hci}$ and u_{hci}^* , implying that Bhattacharya's (2007) and Binder and Kovačević's (1995) approximate expressions for \hat{G} are the same. We have

$$\begin{aligned} \hat{\Psi}_{hci} &= -2 \sum_{d=1}^M w_d \left\{ \frac{1}{\hat{\mu}} \left[y_{hci} I(y_{hci} \leq y_{(d)}) + y_{(d)} \left(\hat{F}(y_{(d)}) - I(y_{hci} \leq y_{(d)}) \right) \right. \right. \\ &\quad \left. \left. - \frac{\hat{\alpha}(y_{(d)})}{\hat{\mu}} y_{hci} \right] \right\} \\ &= -\frac{2}{\hat{\mu}} \left[\sum_{d=1}^M w_d y_{hci} I(y_{hci} \leq y_{(d)}) - \sum_{d=1}^M w_d y_{(d)} I(y_{hci} \leq y_{(d)}) + \sum_{d=1}^M w_d y_{(d)} \hat{F}(y_{(d)}) \right. \\ &\quad \left. - \sum_{d=1}^M w_d \left(\frac{\hat{\alpha}(y_{(d)})}{\hat{\mu}} \right) y_{hci} \right] . \end{aligned}$$

Now:

- $\sum_{d=1}^M w_d y_{(d)} \hat{F}(y_{(d)}) = \frac{\hat{\mu}}{2} (\hat{G} + 1)$
- $\sum_{d=1}^M w_d y_{(d)} I(y_{hci} \leq y_{(d)}) = B(y_{hci})$
- $\sum_{d=1}^M w_d y_{hci} I(y_{hci} \leq y_{(d)}) = y_{hci} - y_{hci} \hat{F}(y_{hci})$
- $\sum_{d=1}^M \frac{w_d y_{hci} \hat{\alpha}(y_{(d)})}{\hat{\mu}} = \frac{y_{hci}}{\hat{\mu}} \sum_{r=1}^H \sum_{s=1}^{n_r} \sum_{t=1}^k w_{rst} y_{rst} \sum_{d=1}^M w_d I(y_{rst} \leq y_{(d)}) = y_{hci} - \frac{(\hat{G}+1)}{2} y_{hci}$.

Using these results, we have

$$\begin{aligned} \hat{\Psi}_{hci} &= -\frac{2}{\hat{\mu}} \left[y_{hci} - y_{hci} \hat{F}(y_{hci}) - B(y_{hci}) + \frac{\hat{\mu}}{2} (\hat{G} + 1) - y_{hci} + \frac{(\hat{G} + 1)}{2} y_{hci} \right] \\ &= u_{hci}^* \quad \# \end{aligned}$$

Proof of Result 2: Here we show that $\widehat{Var}(\hat{G})$ as given by expression (4.2) is equivalent to $\widetilde{Var}(\hat{G})$ as defined in expression (4.3). From (4.2)

$$\widehat{Var}(\hat{G}) = \sum_{h=1}^H \sum_{c=1}^{n_h} (u_{hc}^* - \bar{u}_h^*)^2$$

$$= \sum_{h=1}^H \sum_{c=1}^{n_h} (u_{hc}^{*2} - 2\bar{u}_h^* u_{hc}^* + \bar{u}_h^{*2}) . \quad (\text{A.1})$$

We have,

$$\begin{aligned} & \sum_{h=1}^H \sum_{c=1}^{n_h} \left(\sum_{i=1}^k w_{hci} u_{hci}^* \right)^2 \\ &= \sum_{h=1}^H \sum_{c=1}^{n_h} \left(\sum_{i=1}^k w_{hci}^2 u_{hci}^{*2} \right. \\ & \quad \left. + \sum_{i=1}^k \sum_{j \neq i}^k w_{hci} w_{hck} u_{hci}^* u_{hck}^* \right) , \end{aligned} \quad (\text{A.2})$$

and,

$$\begin{aligned} \sum_{h=1}^H \sum_{c=1}^{n_h} (-2\bar{u}_h^* u_{hc}^* + \bar{u}_h^{*2}) &= \sum_{h=1}^H \left(-2\bar{u}_h^* \sum_{c=1}^{n_h} u_{hc}^* + \sum_{c=1}^{n_h} \bar{u}_h^{*2} \right) \\ &= - \sum_{h=1}^H n_h \bar{u}_h^{*2} \\ &= - \sum_{h=1}^H \frac{1}{n_h} \left(\sum_{c=1}^{n_h} \sum_{i=1}^k w_{hci} u_{hci}^* \right)^2 . \end{aligned} \quad (\text{A.3})$$

Substituting (A.2) and (A.3) into (A.1) and using that $u_{hci}^* = \widehat{\Psi}_{hci}$ gives the desired result that

$$\widehat{Var}(\widehat{G}) = \widehat{Var}(\widehat{G}). \quad \#$$