On Estimating Variances for Gini Coefficients with Complex Surveys: Theory and Application

by

Ahmed A. Hoque
BSS, Shahjalal University of Science and Technology, Bangladesh, 1996
MSS, Shahjalal University of Science and Technology, Bangladesh, 1997
MA, University of Manitoba, 2007

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in Interdisciplinary Studies
in the areas of Econometrics and Statistics

University of Victoria

Supervisory Committee

On Estimating Variances for Gini Coefficients with Complex Surveys: Theory and Application

by

Ahmed A. Hoque
BSS, Shahjalal University of Science and Technology, Bangladesh, 1996
MSS, Shahjalal University of Science and Technology, Bangladesh, 1997
MA, University of Manitoba, 2007

Supervisory Committee

Dr Judith Clarke (Department of Economics)
Co-Supervisor

Dr Laura Cowen (Department of Mathematics and Statistics)
Co-Supervisor

Dr David Giles (Department of Economics)
Departmental Member

Dr Nilanjana Roy (Department of Economics)
Departmental Member

Abstract

Supervisory Committee

Dr Judith Clarke (Department of Economics)
Co-Supervisor

Dr Laura Cowen (Department of Mathematics and Statistics)
Co-Supervisor

Dr David Giles (Department of Economics)
Departmental Member

Dr Nilanjana Roy (Department of Economics)
Departmental Member

Obtaining variances for the plug-in estimator of the Gini coefficient for inequality has preoccupied researchers for decades with the proposed analytic formulae often being regarded as being too cumbersome to apply, as well as usually based on the assumption of an *iid* structure. We examine several variance estimation techniques for a Gini coefficient estimator obtained from a complex survey, a sampling design often used to obtain sample data in inequality studies. In the first part of the dissertation, we prove that Bhattacharya's (2007) asymptotic variance estimator when data arise from a complex survey is equivalent to an asymptotic variance estimator derived by Binder and Kovačević (1995) nearly twenty years earlier. In addition, to aid applied researchers, we also show how auxiliary regressions can be used to generate the plug-in Gini estimator and its asymptotic variance, irrespective of the sampling design.

In the second part of the dissertation, using Monte Carlo (MC) simulations with 36 data generating processes under the beta, lognormal, chi-square, and the Pareto distributional assumptions with sample data obtained under various complex survey designs, we explore two finite sample properties of the Gini coefficient estimator: bias of the estimator and empirical coverage probabilities of interval estimators for the Gini coefficient. We find high sensitivity to the number of strata and the underlying distribution of the population data. We compare the performance of two standard normal (SN) approximation interval estimators using the asymptotic variance estimators of Binder and Kovačević (1995) and Bhattacharya (2007), another SN approximation interval estimator using a traditional bootstrap variance estimator, and a standard MC bootstrap percentile interval estimator under a complex survey design. With few exceptions, namely with small samples and/or highly skewed distributions of the underlying population data where the bootstrap methods work relatively better, the SN approximation interval estimators using asymptotic variances perform quite well.

Finally, health data on the body mass index and hemoglobin levels for Bangladeshi women and children, respectively, are used as illustrations. Inequality analysis of these two important indicators provides a better understanding about the health status of women and children. Our empirical results show that statistical inferences regarding inequality in these well-being variables, measured by the Gini coefficients, based on Binder and Kovačević's and Bhattacharya's asymptotic variance estimators, give equivalent outcomes. Although the bootstrap approach often generates slightly smaller variance estimates in small samples, the hypotheses test results or widths of interval estimates using this method are practically similar to those using the asymptotic variance estimators.

Our results are useful, both theoretically and practically, as the asymptotic variance estimators are simpler and require less time to calculate compared to those generated by bootstrap methods, as often previously advocated by researchers. These findings suggest that applied researchers can often be comfortable in undertaking inferences about the inequality of a well-being variable using the Gini coefficient employing asymptotic variance estimators that are not difficult to calculate, irrespective of whether the sample data are obtained under a complex survey or a simple random sample design.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BDHS	Bangladesh Demographic and Health Survey
BKAM	SN Approximation Confidence Interval Estimator using Binder and Kovačević's (1995) Standard Error
BMI	Body Mass Index
BTAM	SN Approximation Confidence Interval Estimator using Bhattacharya's (2007) Standard Error
CI	Confidence Interval
CP	Coverage Probability
DGP	Data Generating Process
DHS	Demographic and Health Survey
ECP	Empirical Coverage Probability
EDF	Empirical Distribution Function
EE	Estimating Equations
EL	Empirical Likelihood
G	Gini Coefficient
GMM	Generalized Method of Moment
Hb	Hemoglobin
ICC	Intracluster Correlation Coefficient
iid	Independent and Identically Distributed
LA	Lorenz Area
LC	Lorenz Curve
MC	Monte Carlo
MCS	Monte Carlo Sample
NIPORT	National Institute of Population Research and Training
OLS	Ordinary Least Squares
PPS	Probability Proportional to the Size
SBPM	Standard Bootstrap MC Percentile CI Estimator
SBSM	SN Approximation CI using Bootstrap Standard Error
SN	Standard Normal
SRS	Simple Random Sample
SSR	Sum of Squared Residuals
UN	United Nations
UNDP	United Nations Development Program
WHO	World Health Organization
WSPM	Warp-speed MC Percentile CI Estimator

Acknowledgments

This dissertation could not have been completed without the help and support of many people through my Ph.D. journey. I would like to gratefully acknowledge all of them here.

I would like to express my deepest and sincere appreciation to my advisor, Judith A Clarke for her enormous efforts, patient and advice, and not to mention her expert guidance throughout my graduate study years. Working on a Ph.D. dissertation while pursuing a career in the university teaching would have not been possible without her unconditional help and support. I am also grateful to my statistics supervisor Laura Cowen, and the dissertation committee, David Giles, and Nilanjana Roy for their insightful comments and suggestions from time to time. My special thanks go to my wife for supporting me over the years to get the program done.

Dedication

To my family

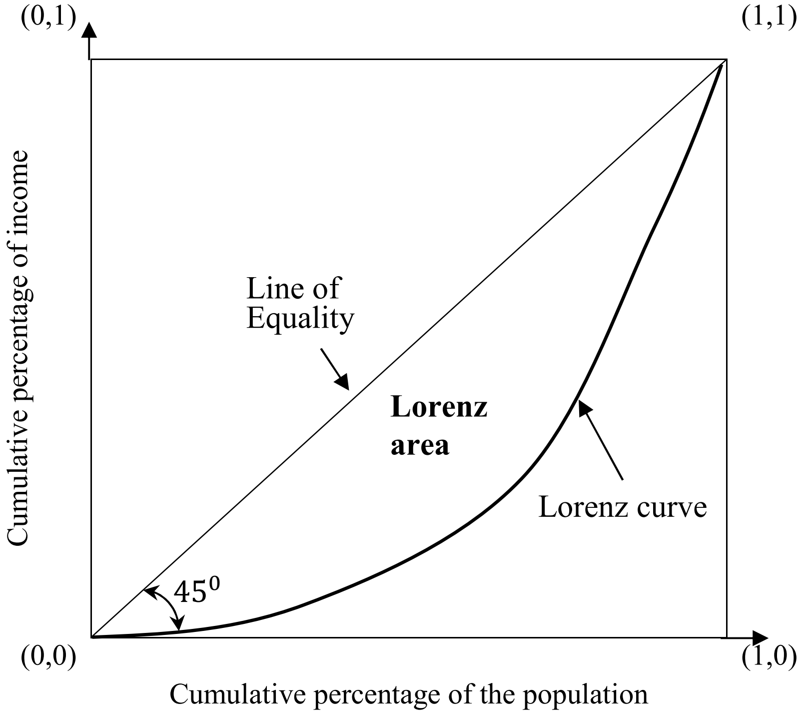
CHAPTER ONE: THE GINI COEFFICIENT

1.1 Introduction

Corrado Gini (1914), an Italian statistician, proposed an inequality index, which he called the *concentration ratio*. Since its inception, this index has attracted a lot of attention and generated an enormous amount of research. Over time, the measure was renamed after Gini as the Gini index or the Gini coefficient (G). Widely used as a measure of income (and wealth) inequality, G has more recently been applied to other measures of *well-being*, e.g., consumption, education, health; see for instance, Slater et al. (2009), Thomas et al. (2001), López et al. (1998). When developing the index, Gini linked his measure to the Lorenz curve or the Lorenz area.

The Lorenz curve (LC) (Lorenz, 1905) graphically illustrates the distribution of the variable of interest, showing the cumulative share of the variable against its recipient share. For instance, to illustrate for income, the LC shows the percentage of the total income received by the bottom x percent of the population against the percentage of the population when arranged in a non-decreasing order (from the poorest to the richest). An illustration of a LC is shown in Figure 1, where the horizontal axis shows the cumulative percentage of the population and the vertical axis shows the cumulative percentage of the income received by x percent of the population. For an equal distribution of income, the LC is the 45° diagonal line, known as the line of equality. The gap between the line of equality and the LC, known as the Lorenz area (LA), forms the basis of inequality.

Figure 1 The Lorenz Curve



The Gini coefficient is defined as twice the LA or the ratio of the LA to the triangular area below the line of equality. The axes are percentiles between 0 and 1 and the value of the area of the triangle is ½. The Gini coefficient indicates the degree of the inequality with larger values showing a greater level of inequality. When the LC coincides with the diagonal line, the line of equality, the Lorenz area and G are zero – perfect equality. Whereas, for a LC running from (0,0) to (1,1) via (1,0), the Gini coefficient is 1 – a perfectly unequal distribution. Hence, the Gini coefficient is bounded by 0 and 1.

Some mathematical expressions for G in terms of the LA and the LC are derived below. There are many other ways to formulate G. Some of the commonly used G expressions are provided in Appendix A.

The LC corresponding to a random variable $Y \in [0, \infty)$ (e.g., income or wealth or some other well-being variable) with cumulative distribution function $F(y)$ and finite nonzero mean is given by: $L(p) = \frac{1}{\mu}\int_0^{F^{-1}(p)} ydF(y)$, where $\mu = \int_0^\infty ydF(y)$, $F^{-1}(p) = \inf\{y \mid F(y) \geq p\}$ is the pth quantile or fractile of the distribution function for $0 \leq p \leq 1$, which is also denoted by ξ_p and $p = F(\xi_p) = \int_0^{\xi_p} dF(y)$. The LC is defined by the first moment distribution function, $F_1(\xi_p)$.[1] On the 45°-line, the line of equality, $p = L(p)$. If $p > L(p)$, the distribution is unequal. The LA is the total area between the line of equality and the LC, given by

$$LA = \int_0^1 \big(p - L(p)\big)dp = \int_0^\infty \big(F(\xi_p) - F_1(\xi_p)\big)dF(\xi_p). \tag{1.1}$$

The Gini coefficient, G, is

$$G = 2 \times LA = 2\int_0^1 \big(p - L(p)\big)dp. \tag{1.2}$$

A commonly used expression for G, derived from (1.2) is one minus twice the integral of the LC with respect to p with a limit of [0,1]:

$$G = 1 - 2\int_0^1 L(p)dp \tag{1.3}$$

The Gini coefficient is a sufficiently simple summary measure of the inequality of a distribution, and its visual description is elegant. While the measure has many desirable properties, such as population size independence, mean independence, symmetry, and Pigou-Dalton transfer sensitivity, it has some limitations as well.

The Gini coefficient does not allow for negative values of the variable. If the variable of interest is negative, the distribution function takes on the value zero. If negative values are incorporated into the equation and the mean is negative, G becomes a negative

---

[1] Nygård and Sandström (1981, p. 132), for instance, define the jth moment distribution function, $F_j(\xi_p)$, as $F_j(\xi_p) = \mu_j^{-1}\int_0^{\xi_p} y^j dF_0(y)$, where μ_j is the jth moment about zero, which is assumed to exist and be nonzero.

number; going below its lower boundary of zero. The Gini coefficient may also exceed its upper limit of 1 with large negative values (e.g., Scott and Litchfield, 1994).

If half the population of an economy has no well-being (e.g., income) and the other half shares total available well-being equally, G is ½. In another example, where the well-being variable is equally distributed except for one household with half the total well-being, G is also ½. Although, according to the Gini measure both of these economies have similar income inequality, the distributions are quite different. Plotting the distributions would highlight such differences, whereas just examining the Gini coefficient would not differentiate in this way.

The Gini coefficient meets the first four[2] of five basic axiomatic principles that an inequality measure should meet (see, e.g., Yitzhaki and Schechtman, 2013, p.4). It does not easily meet the fifth principle, decomposability, which is used to show the sources of inequality. The principle requires the overall inequality to be able to be decomposed into components of within-group inequality and between-group inequality when the population is divided into groups of interest (e.g., rural; urban). When the inequality among subgroups of the population increases, the overall inequality is expected to increase. Bourguignon (1979) discusses additive decomposability, whereby the total inequality of a population is

---

[2] The first four principles (with an income inequality illustration) are: i) *Anonymity* or symmetry: independent of the income earners' qualities other than income, i.e., it does not matter who is earning the income; ii) *Population Independence:* invariant to replications of the population. In other words, the size of the population of a country is inconsequential; iii) *Scale Independence* or the income-zero-homogeneity: unaffected by uniform proportional changes, i.e., when every individual's income in the population is multiplied by the same scalar, inequality remains unchanged; iv) *Transfer Principle* or the Pigou-Dalton transfer principle: inequality in income should decrease (or at least not increase) in response to a transfer of income from a rich person to a poor person. For details, see, e.g., Bourguignon (1979).

expressed “as the sum of a weighted average of the inequality within subgroups of the population and of the inequality existing between them” (p. 902).[3]

This decomposability is met by some other inequality measures, e.g., Theil’s index (1967) and Atkinson’s measure (1970). However, some authors including Jędrzejczak (2008), Dikhanov (2005, 1996), Morales and Costa (1998), Lerman and Yitzhaki (1984), Shorrocks (1983), Pyatt et al. (1980), show decomposability of G may be met under certain conditions,[4] specifically if the sub-groups of the population are non-overlapping in the variable of interest (Litchfield, 1999).

In spite of some limitations associated with G that perhaps motivated a number of alternative inequality measures,[5] the use of G in inequality analysis remains popular. Frequent attempts to decompose G highlight its importance in applied work. Bourguignon (1979) points out that the lack of decomposability for an inequality measure does not mean the measure lacks usefulness, as it may have other relevant features.

1.2 Applications of the Gini coefficient

There is a large number of studies on the application of G in inequality analysis. In income inequality analysis, in addition to measuring inequality of a distribution, G can be used to compare income distributions across different population groups, e.g., urban and

---

[3] Algebraically, the total inequality of the distribution of a well-being variable, y, $I = \sum_{q=1}^{Q} \gamma_q I_q + I(\bar{y}_1, \dots, \bar{y}_Q)$, $\sum_{q=1}^{Q} \gamma_q = 1$, where $\sum_{q=1}^{Q} \gamma_q I_q$ is the weighted sum of the inequality values calculated for population groups Q, $I(\bar{y}_1, \dots, \bar{y}_Q)$ is the contribution arising from differences between group means, $I_q$is the inequality index calculated within the qth group, and γ_q is a weighting function. For example, if $y = (2,4,3,1)$ is grouped into $y_1 = (2,4)$ and $y_2 = (3,1)$, the additive decomposability requires $I(y) = \gamma_1 I(2,4) + \gamma_2 (3,1) + I(3,3,2,2)$; see, e.g., Foster and Shneyerov (1999).

[4] For a list of decomposition attempts made by several authors, see Nygård and Sandström (1981, p. 314-326)

[5] For instance, the Theil’s index (1967) and Atkinson’s measure (1970); for a list of inequality measures, see Cowell (1977, p. 72-73) and Nygård and Sandström (1981, p. 406-407).

rural, as well as across countries. The Gini coefficient is also used as an indicator by organizations such as the United Nations (UN), the World Bank and the Central Intelligent Agency (CIA), to rank countries based on income inequality. The United Nations Development Programme (UNDP) reports estimated income Gini coefficients periodically for most countries. In its 2014 Human Development Report (p. 168-171), the UNDP published estimated income Gini coefficients for 137 out of the 187 member countries around the globe.

For instance, Mitra and Yemtsov (2006) and Milanovic (2005, 1999) investigate income inequality in the transition economies of Eastern Europe and the former Soviet Union using Gini coefficients. Milanovic (2009, 2008, 2006, 2002), a World Bank researcher, has undertaken an extensive policy research using G on regional and global income inequality. Dikhanov (2005), who presents several expressions for G and attempts to decompose the measure, estimates a projected G for the global income distribution for the period 2000-2015 based on 1990-2002 trends in economic growth and UN population projections for 2015. For Korean non-agricultural household incomes, Nho (2006) estimates Gini coefficients for the years 1999 and 2000. He also estimates the measure to compare income inequalities across various provinces of the country.

Apart from the applications of G from a policy perspective, there are a number of theoretical papers that examine G that also include applications; for example, Davidson (2009), Bhattacharya (2007), Modarres and Gastwirth (2006), Giles (2004), and Binder and Kovačević (1995). Aside from income inequality analysis, G can also be applied to other well-being variables, such as educational attainment, women's body mass index, education level, to name but a few; see, e.g., Araar et al. (2009), Slater et al. (2009), Contoyannis and Wildman (2007), Thomas et al. (2001), López et al. (1998), Mass and Creil (1982).

However, despite using samples from the underlying population, most of these applied studies merely report the estimated Gini coefficient without indicating the sampling error, or undertaking hypothesis tests to make statistical inferences about the measure. However, reporting sampling errors and undertaking inference are equally important as the inequality analysis itself. In addition, many of these studies often apply various estimation techniques and use different types of sample data to estimate G seldom discussing these features. This provides at least three neglected features that motivate this research. We briefly elaborate below, with more details in the subsequent subsection.

First: most applied studies that estimate G implicitly assume that they have an independent and identically distributed (*iid*) sample, without explicitly discussing the sampling techniques used to obtain the sample data, and the implications this may have on estimation. However, in practice, survey data, especially large-scale cross-section data on household behaviour, are rarely *iid*. Our work considers estimation of G assuming non-*iid* sample data.

Second: an estimator for G $(\hat{G})$ is a statistic with a sampling distribution. Whenever we report a statistic, we should report an indicator of sampling error, such as a standard error. However, in practice, reporting the standard error of $\hat{G}$ is rare. As stated by Yitzhak (1991), "Although it has been in use for almost 80 years, the standard error of the estimator is seldom reported" (p. 235). According to Karoly (1992) "Despite the existence of methodologies for estimating the variances of many inequality measures (e.g., Sandström et al., 1988; Gastwirth, 1972; Glasser, 1962; Wold, 1935), many researchers do not report standard errors or discuss sampling variability" (p. 108). We show that there is likely no reason to avoid reporting the standard error of the commonly employed plug-in estimator of G.

Third: recognizing the importance of the standard error, some studies (e.g., Luus et al., 2012; Nho, 2006; Moran, 2005a) estimate it using resampling techniques (e.g., the jackknife and the bootstrap) asserting that the traditional delta or linearization method of variance estimation for G is computationally burdensome. We demonstrate that this is indeed not the case. We examine each of these three issues in the following section.

1.3 Sampling Design, Standard Error and Estimation Technique

An appropriate sample is crucial to understanding the features of a population. The amount of information gained from a sample depends on two factors: the size of the sample and the amount of variation in the data (see, e.g., Scheaffer et al., 2006, p. 7). The sample size is often influenced by budgetary issues, but the latter factor is mostly controlled by the sample selection method or sampling design.

The sample selection technique is an important aspect in estimation. There are many sampling designs and their effects on estimation of a parameter are discussed in standard statistics textbooks; e.g., Wolter (2007), Cochran (1977), Raj (1968). A sample can be obtained applying one or more sampling techniques and features, e.g., stratification, clustering, etc. (see Wolter, 2007, p. 11-16, for discussion on various sampling designs, associated estimators and their variances for a population total).

A sampling design is said to be simple random sampling if all individual units in the population have the same chance of being selected into the sample. The subsequent sample is referred to as a simple random sample (SRS). This sampling technique serves two key purposes: in comparing the relative efficiency of other sampling methods, as it sets a baseline; and in advanced sampling methods (e.g., stratified multi-stages sampling) it is sometimes applied to select final sample elements or primary sampling units to ensure randomness in the data set (see, e.g., Lehtonen and Pahkinen, 1995, p. 21). There are two approaches employed under such a method when the size of the population is finite[6]: simple random sampling with replacement and simple random sampling without replacement. However, as discussed in the next chapter, the difference between these approaches does not concern our research.

We assume that elements in a SRS are *iid*, resulting in an *iid* sample. A random sample of size n on a random variable Y is a set of *iid* random values $\{y_i\}_{i=1,\dots,n}$ drawn from the same population; i.e., each of them has the same distribution as Y. That said, not all simple random samples need be *iid* samples. Qin et al. (2010) define the *iid* sample as the SRS when the sampling fraction is negligible, whereas, Lehtonen and Pahkinen (1995, p. 9) assume samples obtained using SRS with replacement approximate *iid* samples. As this work does not need to distinguish between with and without replacement under simple random sampling, occasionally we refer to an *iid* sample as a SRS.

The *iid* sample is a very specific and narrow form of sample. Should there be any reason for which the similarity of the distributions of sampled elements in the sample and that in the population break down, the subsequent sample will no longer be *iid*. There are many situations when this can happen. For instance, many sampling designs produce non-*iid* samples, e.g., unequal probability of selection, probability proportional to size, double

---

[6] In the infinite population case, samples can be selected under with- or without replacement techniques. For a finite sample drawn from an infinite population, both methods usually lead to similar conclusions. As the population size is undetermined, sampling a *random sample* from an infinite population is often regarded as sampling with replacement (see, e.g., Kozak et al., 2008, p. 111-113).

sampling. Broadly, a complex survey sample, which may involve one or more combinations of several sampling techniques and features, leads to a non-*iid* sample. For instance, stratification, clustering, unequal probability of selection, multistage sampling, double sampling, multiple frames, estimation features such as large observations or outliers, adjustments for nonresponse and undercoverage, poststratification, etc. (e.g., Wolter, 2007, p. 2), fall into this class. Given the nature of the sampling techniques and features, a complex survey design automatically violates the main feature of the *iid* sample – sampled observations having the same distribution as their distribution in the population.

Despite this, it is convenient to apply an estimation technique assuming an *iid* sample data, and, as stated previously, most of the applied studies discussed earlier assume an *iid* sample when estimating G. However, in practice, sample data used to estimate inequality measures rarely maintains the *iid* assumption, especially survey data that may contain one or more of the above mentioned sampling designs or features.

In particular, many nationally representative surveys data used in applications in economics and statistics, e.g., the Canadian Survey of Consumer Finance (SCF), Canadian Labor Survey Force, the Indian National Sample Survey (NSS), the Demographic and Health Surveys (DHS) and National Health Interview Survey, are collected using multi-stage or complex survey designs. For example, the Bangladesh DHS 2011 survey is a two-stage stratified sample of households. Before sampling, a total of 20 sampling strata were created. In the first stage, 600 clusters (primary sampling units) were selected with probability proportional to the cluster size, and with independent selection from each stratum. In the second stage, a fixed number – 30 households per cluster - were selected with an equal probability systematic selection. With this design, the survey selected 18,000 residential households for interviews.

Stratification means that the original population is divided into homogeneous subgroups before sampling; e.g., households divided into rural, urban or country regions. The selection of the sample is completed independently within each group. The strata are mutually exclusive (i.e., every element in the population must be assigned to only one stratum), and are collectively exhaustive (i.e., no population element is excluded). Typically, stratification breaks down the *identical* part of an *iid* assumption because of the dissimilarity between observations across strata. Stratification normally reduces the variability of statistics over repeated samples; i.e., it increases the precision of estimators.

On the other hand, the *independent* part of the *iid* assumption, is usually violated with clustering. Clustering is a sampling technique whereby the population is divided into several groups, commonly known as primary sampling units. Often, these clusters of the population contain elements that are contiguous, e.g., villages or metropolitan areas or cities. Therefore, observations are likely to be correlated. Usually the survey design leads to only sampling from a subset of clusters, so that although clustering reduces survey costs and facilitates fieldwork, it results in correlation between observations within clusters and hence normally reduces the precision of estimators.

It is important that an estimation technique takes the sampling design properly into account, as this can result in markedly different estimates from those obtained under an *iid* assumption, especially that of the variance of estimators, including of inequality indices. For instance, Bhattacharya (2007, 2005) rigorously discusses the importance of sampling design on estimation, especially on variance estimation. He also derives a variance formula for $\widehat{G}$ under the complex survey design, with the variance estimator being disaggregated into three parts: simple random sampling variance, cluster effect on variance and stratum effect on variance. If the cluster effect is not fully offset by the stratum effect, variance estimates using a SRS and a complex design sample will be different. We elaborate this research in the next chapter.

Although the importance of the standard error of an estimator is well understood, as stated above, estimating this statistic for a $\widehat{G}$ is often avoided by applied researchers. We assume that this is because of perceived difficulties. For instance, as the usual estimator of G is a nonlinear statistic that cannot be represented by functions of moments alone, its variance estimator computation is reputed to be complicated by standard techniques (e.g., the delta or linearization methods).

Despite the lack of attention in applied studies, the importance of the standard error for $\widehat{G}$ has been well documented by statisticians and econometricians in theoretical research. For example, Shao (1994), Schechtman and Yitzhaki (1987), Gastwirth and Gail (1985), Nygård and Sandström (1981), Sendler (1979), Mehran (1976), Glasser (1962), and Hoeffding (1948) provide formula to calculate the variance of a Gini coefficient estimator using U-statistics and L-statistics when data are from a simple random sample. Qin et al. (2010), Davidson (2009), Modarres and Gastwirth (2006), Giles (2004), Karagiannis and Kovačević (2000), Ogwang (2000), Shao (1994), Yitzhaki (1991), Nygård and Sandström (1989), and Sandström et al. (1988, 1985) use resampling techniques under an *iid* assumption; Bhattacharya (2007) uses a generalized method of moment approach while Binder and Kovačević (1995) use estimating equations techniques to provide variance estimators under a complex survey sampling design. Davidson (2009) also provides an analytical variance formula for $\widehat{G}$ under the *iid* framework.

As this brief discussion highlights, in contrast to many studies that have considered variance estimation under simple random sampling, there are only a small number of studies available on standard error estimation for $\widehat{G}$ with complex survey sampling design, and applied researchers have not readily adopted such variance estimators, perhaps believing that the complicated mathematical expressions are burdensome to code with standard computer software, which we show is indeed not the case.

Variance estimation for $\widehat{G}$ under a complex sampling design depends on the sampling plans at the different stages, so that it is extremely difficult to obtain an exact estimator. There are two approaches commonly used to approximate the variance: Taylor series linearization and resampling techniques. Both the estimating equations and the generalized method of moment approaches to obtain variance estimators are based on a Taylor series linearization. We show algebraically that both methods yield the same result for the variance estimator of $\widehat{G}$ with a complex survey sample. In addition, we use Monte Carlo (MC) simulation experiments to compare these asymptotic variance estimators with those that would be obtained using commonly employed bootstrap techniques.

1.4 Contributions of this study

This dissertation contributes to both the theory and applications of G in inequality analysis of a well-being variable. We outline these contributions below.

In Chapter 2*,* we provide a theoretical framework that demonstrates that the variance estimators proposed by Binder and Kovačević (1995; based on estimating equations theory) and by Bhattacharya (2007; based on a generalized method of moments approach) are asymptotically equivalent for the plug-in estimator of G with a complex survey sample. This finding is useful for applied researchers, because the variance formula of Binder and Kovačević is easier to code. We also show mathematically how Davidson's (2009) variance estimator for $\widehat{G}$ obtained from an iid sample is a special case of Bhattacharya's (2007) variance estimator, as well as Binder and Kovačević's (1995) variance estimator, from a complex survey sample. In addition, we provide a straightforward auxiliary regression technique to calculate the plug-in estimator for G and its asymptotic variance estimator, regardless of the sampling design, that reduces the computational burden substantially.

In Chapter 3, we use MC simulations to examine the finite sample properties of our studied estimators. Thirty-six data-generating processes (DGPs), with different combinations of strata, clusters, observations, and intracluster correlation coefficients under four probability distributions, are used to examine two properties of an estimator for the Gini coefficient: the bias of $\widehat{G}$ and the empirical coverage probability (ECP) of the nominal 95% confidence interval (CI) estimator, with the complex survey sample. Simulation results show that the distribution of data, the number of strata, and the relatedness of households within a clusters are important population features, in addition to the sampling design, that affect the accuracy and performance of Gini coefficient estimators.

To further ascertain how the two linearization variance estimators, proposed by Bhattacharya (2007) and Binder and Kovačević (1995), perform in finite samples, our MC simulations examine the performance of standard normal (SN) approximation CI estimators. We find that their ECPs are highly comparable. In addition, we compare these with a SN approximation interval estimator using a standard bootstrap variance estimator and a bootstrap MC percentile interval estimator. Although for small samples (with few sampled clusters), the bootstrap SN approximation interval estimators often have somewhat higher ECPs, with more clusters in the sample, the performance of this method is usually no better than the two SN approximation interval estimators using analytical variance estimators. More often the three SN approximation estimators provide similar ECPs for samples with more clusters. The bootstrap MC percentile interval estimators work well both for small samples and heavy-tailed distributions of data. However, as estimating interval estimators using bootstrap techniques that account for the complex survey design is more time consuming, the gains are often not significant enough when compared to those using asymptotic variance estimators.

Finally, in Chapter 4, we consider applications of G in inequality analysis for two well-being variables: women's body mass index (BMI) and children's hemoglobin level (Hb) using the 2004, 2007, and 2011 Bangladesh Demographic and Health Survey data. In addition to estimating G and making various inferences about inequality for our two well-being variables, we use descriptive statistics of the variables to discussion health status of women and children in Bangladesh.

CHAPTER TWO: VARIANCE ESTIMATION WITH A COMPLEX SURVEY SAMPLE

2.1 Introduction

Since the inception of the Gini coefficient (G), many scholars have extensively searched for a convenient way to estimate the variance or the standard error of the estimator of G. Most of these studies are based on the simple random sample or *iid* assumption of the sample data. Ways to account for a non-simple random sample or a non-*iid* sample have received little attention, despite the prevalence of the use of such data. No matter what type of sample data are considered for estimation, as G itself is a nonlinear function of the sample data, the proposed standard error formulae in the literature are typically complicated mathematical expressions and considered difficult to code in practice. Consequently, most empirical studies report $\hat{G}$ without a standard error. Nevertheless, the large theoretical literature on variance estimation for $\hat{G}$ reveals its importance and offers scope for further research for finding a computationally convenient formula.

In this chapter, we investigate several formulae for the variance of the common plug-in estimator of G with complex survey data. Objectives include: showing that two broad, unlinked, existing methods for estimating the variance for $\hat{G}$ yield the same asymptotic estimator, and proposing a convenient and relatively straightforward estimation technique, via use of some auxiliary regressions, to obtain a variance estimator with a complex survey sample. In Section 2.2, we give a brief literature review on variance estimation for G with *iid* samples and non-*iid* samples. In Section 2.3, we provide a general discussion on three estimation methods that arise with our work: the estimating equations approach, the generalized method of moment method, and the use of regressions for estimation of population parameters. Estimation of G with both *iid* and non-*iid* samples is detailed in Section 2.4. To obtain the asymptotic variance estimator for $\hat{G}$, an approximation for $\left(G - \hat{G}\right)$ is needed, which is derived in Section 2.5. In Section 2.6, we show that the variance formulae derived by Binder and Kovačević (1995) and Bhattacharya (2007) are asymptotically equivalent. In Section 2.7, we propose the regression technique as a straightforward method to obtain the variance estimator for $\hat{G}$ with a complex survey; we also detail how this approach can be used with a SRS or under an *iid* assumption.[7]

2.2 Literature Review

Although there are some early studies that provide variance formulae for an estimator of the Gini coefficient (e.g., Kendall and Stuart, 1977, p. 240-42; Glasser, 1962; Hoeffding, 1948), the literature remained relatively dormant until the revival of inequality measurement by Atkinson (1970). Subsequently, interest in variance estimation increased significantly. Researchers have taken various approaches with one goal being to make the formulae accessible to applied researchers. But, most of this theoretical work assumes an *iid* sample. Our work, in contrast, falls into the small, but growing, literature that theoretically allows for complex sampling designs that lead to non-*iid* samples. In the following subsection, we provide brief details on the theoretical research that examines variance estimation associated with an estimator of G with *iid* samples, followed by a

---

[7] The theoretical results in this chapter, along with a brief empirical application taken from Chapter 4, have now been published in the paper Hoque and Clarke (2015).

discussion on the relevant literature that allows for complex survey sampling, which results in a non-*iid* sample.

2.2.1 Studies that assume an *iid* sample

The literature on variance estimation for an estimator of G assuming an *iid* sample is relatively large. A survey of the early literature on this issue that focused mainly on Lorenz dominance and G estimation with various forms of the variance formulae is given in Nygård and Sandström (1981). The recent developments in this area are also noteworthy. We summarize some of the studies here.

In the early research, deriving a sample variance formula for the Gini's mean difference expression was popular. As detailed in Appendix A, the sample Gini coefficient can be written as $\hat{G} = \frac{\hat{\Delta}}{2\hat{\mu}}$, where, $\hat{\Delta} = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| y_i - y_j \right|$ is the Gini's mean difference with repetition; and without repetition $\hat{\Delta} = n^{-1}(n-1)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|$. Niar (1936) was among the first to derive the sample variance for $\hat{\Delta}$ under a simple random sampling design, followed by many others, e.g., Lomnicki (1952). A number of subsequent studies, under an *iid* sample, propose variance formulae for the estimator of the mean difference form of G using various statistical approaches, e.g., Yitzhaki (1991), Schechtman and Yitzhaki (1987), Gastwirth and Gail (1985), Glasser (1962), and Hoeffding (1948) derive the asymptotic variance formula applying U-statistics. In contrast, Shao (1994), Nygård and Sandström et al. (1988), and others use the Gini's mean difference form to derive the variance formula for $\hat{G}$ based on L-statistic theory.

Other notable earlier studies on variance estimation for $\hat{G}$ assuming an *iid* sample include Beach and Davidson (1983), and Sendler (1979). Sendler provides a distribution-free variance formula for $\hat{G}$ that depends on the Lorenz curve ordinates. Beach and Davidson estimate the covariance for the interpolated G of K ordinates of the Lorenz curve corresponding to K percentiles and show the joint asymptotic normality of these estimators of the K ordinates.

The theoretical contribution of the above studies to variance estimation for $\hat{G}$ is widely accepted. However, despite this remarkably large theoretical literature, applied work using this theory is rare. Indeed, applying these formulae is extremely difficult. Recently, a number of studies have examined ways to obtain a variance formula based on the order statistics of the assumed *iid* observations that reduces the computational burden; e.g., Davidson (2009), Modarres and Gastwirth (2006), Giles (2004), and Ogwang (2000).

These authors propose regression approaches[8] to provide the estimator of G. Modarres and Gastwirth (2006) and Ogwang (2000) propose resampling techniques (e.g., jackknife and bootstrap methods), to estimate the standard error of $\hat{G}$. While Davidson (2009) provides algebraic formula for an analytical variance formula, Giles (2004) suggests that a simple regression approach is sufficient to estimate the standard error of $\hat{G}$, which eliminates the computational burden significantly. Ogwang presents a regression model based on the covariance approach of the Gini coefficient's estimator developed by Shalit (1985), Lerman and Yitzhaki (1984), and Anand (1983). He then derives an algorithm to compute the standard error of $\hat{G}$ using the jackknife method and proposes that the regression G can be reported with the jackknife standard error.

Giles (2004) examines Ogwang's resampling approach, showing that an extension of the regression framework can be used to construct an appropriate standard error for $\hat{G}$. He uses consumption data for 133 countries from the Penn World Table (Summers and

---

[8] Given the ease of use of auxiliary regression approaches, we detail this method explicitly in Section 2.6.

Heston, 1995) to estimate the standard error for $\hat{G}$ by ordinary least squares (OLS) and weighted least squares (WLS) regression methods and the jackknife method proposed by Ogwang (2000) to make a comparison among them.

However, Davidson (2009) and Modarres and Gastwirth (2006) point out that the regression technique proposed by Giles (2004) can be improved on by accounting for the correlation introduced in the error terms when the *iid* data are ordered. Modarres and Gastwirth show that ignoring the correlation in the error terms in the regression technique can result in overestimating the standard error. They recommend a more complex mathematical variance formula for $\hat{G}$. In particular, they suggest use of Hoeffding's (1948) approach to obtain an asymptotic variance along with resampling methods if desired.

Supporting Modarres and Gastwirth, Davidson (2009) presents an asymptotically correct standard error estimation technique for $\hat{G}$ with an *iid* sample, based on a Taylor series approximation. Davidson's method of estimating the variance for $\hat{G}$ based on the asymptotic approximation for $(\hat{G} - G)$ with an *iid* sample is readily generated. In Section 2.5, we show how Davidson's variance estimator fits in with those proposed by Bhattacharya (2007) and Binder and Kovačević (1995).

Using MC simulations, Davidson (2009) finds that the quality of the variance approximation is "very good" if the tail of the underlying distribution is not too heavy. This suggests that for applied work using income or expenditure data from developing countries, which are normally heavily skewed (e.g., see Langel and Tillé, 2013), the variance estimator may not work well for the lower end of the distribution. In addition, Davidson also considers use of both jackknife and bootstrap methods. Re-examining the consumption data used by Giles (2004), under an iid assumption, Davidson finds that his analytic asymptotic variance estimator compares well with the resampling estimates.

For computational ease, Davidson (2009), Giles (2004) and Ogwang (2000) show that an auxiliary regression can be estimated to readily obtain $\widehat{G}$ with iid data. We extend their results to a complex survey, along with showing that another auxiliary regression can be used, if needed, to obtain an asymptotically valid variance estimator.

Recently, Qin et al. (2010) also propose CI estimators for G using normal and bootstrap approximations and empirical likelihood based methods for iid samples, and allowing for stratified samples as well. Their variance formulae for $\widehat{G}$ use the U-statistics theory. A simulation study is undertaken to examine their methods with five types of CI estimators for G: the normal approximation interval, the bootstrap percentile interval, the bootstrap-t interval, the empirical likelihood (EL) interval based on the scaled χ^2 approximation, and the EL ratio interval using a bootstrap calibrated method. In contrast to Giorgi et al. (2006), where the bootstrap-t confidence interval for the generalized G is preferred over the normal approximation, Qin et al. suggest a preference for the bootstrap-calibrated EL ratio confidence interval over the other four intervals based on speed of convergence in probabilities. See also Peng (2011).

2.2.2 Studies that assume a non-*iid* sample

Large-scale survey samples, as mentioned in Section 1.3 of Chapter 1, are usually collected using complex survey designs that do not satisfy the iid assumption. The separate layers of a complex survey design (e.g., clustering and stratification) are discussed extensively in standard sampling textbooks (e.g., Cochran, 1977; Raj, 1968), and variance estimation for a multistage complex survey that simultaneously incorporates all possible design features, sampling weights and techniques has also appeared in statistics and econometrics references/textbooks. For instance, Wolter (2007), Deaton (1997), Lehtonen and Pahkinen (1995) and Skinner et al. (1989) discuss estimation of parameters and sampling variance techniques using complex survey samples. Yet the number of studies on statistical inference on inequality measures assuming a complex survey sample is quite limited. Below we review some studies that consider complex survey samples or non-*iid* samples in determining variance estimators for inequality statistics.

Kovačević and Binder (1997), Binder and Kovačević (1995), Binder and Patak (1994), and Binder (1991) use the theory of estimating equations (EE), first proposed by Godambe (1976, 1960), and Godambe and Thompson (1984, 1978), for variance estimation for complicated parameter estimators using non-*iid* sample data. They argue that estimating some inequality measures and their standard errors using EE is more convenient and applicable under different types of sample data. Binder and Kovačević (1995), in particular, use the EE approach to provide variances estimators for $\hat{G}$, Lorenz curve and the Low Income measure when sample data are from a complex survey. They point out that when the method is applied with an *iid* sample, the subsequent variance estimator for $\hat{G}$ is equivalent to those obtained by Sendler (1979) and Glasser (1962).

Biewen and Jenkins (2006) propose variance estimators of Generalized Entropy (GE) and Atkinson inequality indices under the non-*iid* sampling framework that can be calculated easily with available software. They adopt the linearization method for variance estimation under the complex survey design, based on a Taylor series approximation, drawing on ideas from Woodruff (1971). Further research on these inequality measures with a complex survey are conducted by Clarke and Roy (2012), who examine inference using Wald statistics and consider decomposition the inequality statistics.

Qin et al. (2010) and Yitzhaki (1991) adopt alternative methods to calculate variances for $\hat{G}$ using stratified random samples only, as an extension to the iid sample analyses in both studies. Yitzhaki proposes the use of jackknife tools method and Qin et al. formulate normal approximation confidence intervals and EL ratio confidence interval for G based on U-statistics. Qin et al. also use a bootstrap-calibrated EL ratio confidence interval constructed by drawing independent bootstrap samples from each of the strata with simple random sampling with replacement.

Bhattacharya (2005) adopts a generalized method of moment (GMM) approach for asymptotic inference with complex survey data for some parameter vector of interest, usually at the individual or household level. In Bhattacharya (2007), he derives the influence functions for a generalized methods of moments estimator of G, which are linear functionals of the influence functions for the Lorenz share, the quantile and the mean of the variable of interest. The asymptotic normality of the Lorenz process, $\varphi(p) = \hat{L}(p) - L(p),$ implies normality of $\hat{G}$ (Bhattacharya, 2007). Several authors, including Berger (2008), Colwell and Victoria-Feser (2003), and Deville (1999), use influence functions for deriving the asymptotic variance of the Gini coefficient, and consider application to survey data. Langel and Tillé (2013) provide a comprehensive summary of a number of approaches by different authors to variance estimation for G under both iid and survey samples, in addition to expressing their concerns regarding the missing linkages between much of the literature outcomes that are similar; our work falls into this area of creating links between previous studies. This study also criticizes Bhattacharya (2007), in particular, for not acknowledging previous works using similar approaches (e.g., influence functions) and results on variance estimation for the Gini coefficient.

In generating estimators for variances, Bhattacharya (2007, 2005) breaks down the variance into the individual effects of stratification and clustering from that from simple random sampling. As explained in the previous chapter, stratification deflates the magnitude of the sample variance of the estimator of interest, while clustering inflates; his variance formula provides a visible sense of how the variance estimator can be overestimated or underestimated if these sampling features are ignored. For instance, when units within clusters are more homogeneous, e.g., households with high level of incomes living in similar areas, the cluster effect can be larger than the stratum effect. In such a case, if we ignore the sampling design, i.e., we assume simple random sampling was used, the sample variance of the estimator of interest will likely be underestimated, possibly leading to narrower confidence intervals or hypothesis tests with inflated type I error, when in fact, they are not (see, e.g., Kreuter and Valliant, 2007). However, the asymptotic expression for the estimated variance of $\widehat{G}$ (Bhattacharya, 2007, p. 684) is not user friendly, in terms of coding, for applied researchers. We elaborate extensively on Bhattacharya's approach in subsection 2.6.

2.3 Estimation Techniques

In this section, we briefly review three estimation techniques – estimating equations, GMM and a regression approach – to estimate G and its variance with *iid* and non-*iid* samples. In sample variance estimation, Binder and Kovačević (1995), Kovačević and Binder (1997) use an EE approach, while Bhattacharya (2007, 2005) uses a GMM approach under a non-*iid* sample framework. Davidson (2009), Modarres and Gastwirth (2006), Giles (2004) and Ogwang (2000) use regression methods to provide variance estimates under *iid* sampling, with estimators obtained from asymptotic principles.

It is reasonable to ask whether these asymptotic methods are providing the same variance estimators when the statistic of interest is the same. We provide mathematical evidence that the subsequent asymptotic variance formulae from the EE and GMM methods produce equivalent results under the complex survey design. We also extend the regression approach to asymptotic variance estimation for $\hat{G}$ as in Davidson (2009), for the complex survey case.

2.3.1 The Estimating Equations Theory

The estimating equations (EE) theory was first proposed by Godambe (1960) and Godambe and Thompson (1978, 1984) to optimally estimate an unknown population parameter of interest $\theta_0 \in \Omega$. For estimation, let Y be an observed value of y such that $\Upsilon = \{y\}$, an abstract sample space and F is a distribution on y such that $\mathcal{F} = \{F\}$, a class of distributions, which can be described by

$$F(y) = \begin{cases} \Pr\{Y \leq y\} & \text{for infinite populations} \\ N^{-1}\displaystyle\sum_{i=1}^{N} I\{Y_i \leq y\} & \text{for finite populations of size } N \end{cases}.$$

Let $g = g\{y, \theta_0(F)\}$ be a real function on $\Upsilon \times \Omega$ which is continuous and differentiable with respect to θ_0. Any function $g \in \mathcal{G}$ is called a regular estimating function if it satisfies certain conditions provided in Godambe (1960, p. 1208). The parameter θ_0 can be estimated as a solution to the equation $E_Y[g\{y, \theta(F)\}] = g(Y, \theta) = 0$, where E_Y is the expectation under F. Equivalently, the solution is obtained from

$$\int_{-\infty}^{\infty} g(y,\theta_0)\, dF(y) = 0. \tag{2.1}$$

For an observed sample $Y = y$ with θ an arbitrary value of θ_0, an estimating equation for θ_0 may be represented by

$$g(Y,\theta) = 0, \tag{2.2}$$

where $g(Y,\theta) = E_Y[g\{y,\theta(F)\}]$ and which is satisfied by $\theta_0 = \theta_g(Y)$, where $\theta_g(Y)$ is an estimator of θ_0. We denote the estimator by $\hat{\theta}$, hence, $\hat{\theta} = \theta_g(Y)$. For instance, for an *iid* sample of size n, $\{y\}_{i=1}^{n}$, the estimator $\hat{\theta}$ is obtained by solving the estimating equation: $n^{-1}\sum_{i=1}^{n} g(y_i,\theta) = 0$.

An estimating function $g^* \in \mathcal{G}$ is said to be optimal if $\lambda(g^*,F) \geq \lambda(g,F)$, where λ is the efficiency in estimating θ_0 through the equation $g(y,\theta_0) = 0$, and $\lambda(g,F) = [\int g^2(y,\theta_0)\, dF(y)\,]^{-1}\left[\int \frac{\partial g(y,\theta_0)}{\partial \theta_0} dF(y)\right]^2$. The optimal estimating equation is given by

$$g^*(Y,\theta) = 0. \tag{2.3}$$

For illustration, to estimate the population mean and variance, $\theta_0 = [\mu_0, \sigma_0^2]'$, with an *iid* sample of size n, under certain regulatory conditions, Godambe and Thompson (1978) show that the optimal estimating equations are:

$$\begin{aligned} g^*(\bullet) &= n^{-1}\sum_{i=1}^{n}(y_i - \mu) = 0, \quad \text{and} \\ g^*(\bullet) &= \sum_{i=1}^{n}\left[(y_i - \mu)^2 - \left(\frac{n-1}{n}\right)\sigma^2\right] = 0. \end{aligned} \tag{2.4}$$

Solving the equations, the estimator for the mean and variance are given by $\hat{\mu} = \bar{Y} = n^{-1}\sum_{i=1}^{n} y_i$ and $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n-1}$ respectively.

For infinite populations with a continuous and differentiable probability density function $p(y, \theta)$, the parameter θ_0 can be estimated from the optimal estimating function $g^*(Y, \theta)$, defined as $g^*(Y, \theta) = \frac{\partial \log p\,(y,\theta)}{\partial \theta}$, and an estimator of θ_0 is the solution to $\int g^*(Y, \theta)\, dF(y) = 0$ (e.g., see Kovačević and Binder, 1997). For a finite population, parameters may also be estimated by this approach under the assumption that the finite population is a sample drawn from an underlying infinite population. For example, for an *iid* sample, using the estimating equation technique, $\hat{\theta}$ is the solution to

$$\int_{-\infty}^{\infty} g^*(y, \theta) d\hat{F}(y) = 0, \tag{2.5}$$

where $\hat{F}$ is the sample (empirical) distribution function. For instance, if θ_0 is the population mean then the optimal estimator is given by $\hat{\theta} = \bar{Y} = n^{-1} \sum_{i=1}^{n} y_i$. More examples are discussed in Binder and Patak (1994).

Godambe and Thompson (1978) show that estimators obtained using optimal estimating equations given by expression (2.5) are consistent, and assert that under certain non-restrictive conditions, the estimator $\hat{\theta}_n$, based on a sample of size n, obtained from such an estimating equation, has the property that $n^{1/2}(\hat{\theta}_n - \theta_0)$ is asymptotically normally distributed with mean 0 and variance $E\left[g_n \left\{E\left(\frac{\partial g_n}{\partial \theta}\right)\right\}^{-1}\right]^2$.

2.3.2 The Generalized Method of Moments (GMM) Theory

GMM is an overriding principle for estimating parameters of both linear and nonlinear models that require a certain number of moment conditions to derive estimators. These moment conditions are functions of the model's parameters and the data. When the probability distribution is unknown or incompletely specified, or the system is over-identified (more moment equations than parameters, i.e., $l > k$, where l is the number of equations and k is the number of parameters), the GMM approach provides computationally convenient estimators (e.g., Hall, 2005, p. 2). Under certain assumptions, GMM estimators are consistent and asymptotically normal.

The theory proceeds with defining some moment conditions. We illustrate here with moment equations from an underlying population distribution. Let Y be a continuous random variable and θ_0 be a vector of unknown parameters, with probability density function $p(y; \theta_0)$ and distribution function F. For a positive integer j, which can be up to $2k$, where k is the number of parameters in the model, the jth moment of the distribution F is given by

$$E_Y\left(Y_j\right)' = \int_{-\infty}^{\infty} y^j \, dF(y), \tag{2.6}$$

and the jth moment about the mean or central moment of F is given by

$$E_Y\left(Y_j\right) = \int_{-\infty}^{\infty} [y - E(Y)]^j \, dF(y). \tag{2.7}$$

Then the corresponding population moment condition is expressed by

$$E_Y[m\{y, \theta_0(F)\}] = \int_{-\infty}^{\infty} m(y, \theta_0) \, dF(y) = 0\,, \tag{2.8}$$

where $m(\cdot)$ is a vector of functions containing the model's variables and parameters; the so-called moment equations.

The system in expression (2.8) is said to be identified if there is a unique solution, such that $E_Y[m(y, \theta)] = 0$ iff $\theta = \theta_0$. Also, when the distribution in the population is known, expression (2.8) can usually be solved to provide an estimator of θ_0. For instance, the parameter vector $\theta_0 = (\mu_0, \sigma_0^2)'$ satisfies the population moment conditions in expression (2.8), with, $m(y, \theta_0) = \begin{bmatrix} y - \mu_0 \\ y^2 - (\sigma_0^2 + \mu_0^2) \end{bmatrix}$.

For a given *iid* sample of size n, $\{y_i\}_{i=1}^{n}$, the analogous sample moment conditions are solved to obtain method of moment estimators. The sample moment equations can be written as $\int_{-\infty}^{\infty} \bar{m}_j(y_i, \hat{\theta}_j)\, d\hat{F}(y) = 0$ or $\bar{m}_j = \frac{1}{n}\sum_{i=1}^{n} m_j(y_i)$. Subsequently, the corresponding sample jth row/uncentered moment and jth moment about the centre condition are given by $\bar{m}_j' = \frac{1}{n}\sum_{i=1}^{n} y_i^j$ and $\bar{\bar{m}}_j = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{m}_1')^j$, respectively. For instance, moment estimators, $\hat{\mu}_{MM}$ and $\hat{\sigma}_{MM}^2$ of the population mean and variance respectively, can be obtained by solving first and second sample moment conditions: $\hat{\mu}_{MM} = \bar{m}_1' = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\hat{\sigma}_{MM}^2 = \bar{m}_2' - \bar{m}_1'^{\,2} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_{MM})^2$. The moment estimator $\hat{\mu}_{MM}$ is unbiased and consistent for the population mean regardless of the distribution of the population, whereas $\hat{\sigma}_{MM}^2$ is a consistent, but biased, estimator of the population variance.

The century old method of moments approach to parameter estimation can be used in situations that are more complicated, leading to the theory of generalized method of moments. For instance, for a population with $\mu > 0$, there are two unbiased and consistent estimators of the population mean: $n^{-1}\sum_{i=1}^{n} y_i = \bar{Y}$ and $[n^{-1}\sum_{i=1}^{n}(y_i - \hat{\mu}_{MM})^2]/3 = \hat{\sigma}_{MM}^2/3$. Although choice of selection of one of them is made based on the smallest variance, it is undetermined which one will have better precision in advance. GMM theory assists in such a circumstance. For a large sample, GMM combines the corresponding sample moment conditions in a way to obtain an asymptotically optimal estimator of μ (e.g., Wooldridge, 2001).

GMM involves choosing parameter estimators to minimize a quadratic form, the distance from the sample moment condition to zero. Let the sample analog of expression (2.8) be

$$\frac{1}{n}\sum_{i=1}^{n} m_i(y_i,\theta) = \tilde{m}(Y,\theta)\,. \tag{2.9}$$

Under appropriate Laws of Large Numbers, $\tilde{m}(Y,\theta)$ converges to $E_Y[m\{y,\theta_0(F)\}]$. The GMM estimator is the value that minimizes the distance

$$M_n(\theta) = \tilde{m}(Y,\theta)'W_n\,\tilde{m}(Y,\theta)\,, \tag{2.10}$$

where W_n is a positive semi-definite weight matrix and $p\lim_{n\to\infty} W_n = W_0$, where W_0 is a positive definite matrix of constants. The restriction on the weighting matrix ensures that $M_n(\hat{\theta}) = 0$ for $\tilde{m}(Y,\hat{\theta}) = 0$, for some estimator $\hat{\theta}$. The GMM estimator is given by

$$\hat{\theta}_{GMM} = \arg\min_{\theta\in\Omega}\,\{\tilde{m}(Y,\theta)'W_n\,\tilde{m}(Y,\theta)\}\,. \tag{2.11}$$

Under certain conditions, $\hat{\theta}_{GMM}$ is a consistent estimator of θ_0; i.e., $p\lim_{n\to\infty}\hat{\theta}_{GMM} = \theta_0$, and asymptotically normal (see, e.g., Hall, 2005, p. 50-69). Applications of GMM are widespread[9] due to its ability to offer asymptotically consistent and normally distributed estimators with a few auxiliary assumptions about the distribution of the population. Bhattacharya (2007, 2005) adopts such a framework to generate estimators of parameters obtained from complex survey samples. We elaborate below.

2.3.3 Regression Estimation Theory

Estimators we discuss extensively in this chapter are arrived at using the principles outlined in the previous two subsections: estimating equations and GMM equations. It turns out that these estimators, for our case, are readily obtainable using auxiliary

---

[9] For instance, see the diverse cases cited in Wooldridge (2001).

regressions whether the sample data is from an *iid* sample or from a complex survey. Consequently, here in this subsection, we provide a brief outline of some basic regression theory, when the sample is assumed to be *iid* followed by a discussion on extending the method to allow for complex survey samples. To estimate a set of unknown population parameters using the standard regression approach, we consider the classical linear regression model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \ , \tag{2.12}$$

where $\boldsymbol{y}$ is an $(n \times 1)$ vector on the regressand, $\boldsymbol{X}$ is an $(n \times K)$ matrix of regressors, $\boldsymbol{\beta}$ is a $(K \times 1)$ vector of parameters and $\boldsymbol{\varepsilon}$ is an $(n \times 1)$ vector of errors. The disturbance/error term in the stochastic model is unobservable.

The most commonly used estimation principle is the Ordinary Least Squares (OLS) method. Under the full set of classical regression model assumptions (e.g., Greene, 2008, p. 44), with a sample of *iid* observations $\{(y_i, x_{i1}, x_{i2}, \dots, x_{ik}), i = 1,2, \dots, n\}$, the OLS method uses the data most efficiently to estimate the model's parameters. The estimator of $\boldsymbol{\beta}$ is obtained by minimizing the sum of the squared residuals[10] associated with some arbitrary estimator $\widehat{\boldsymbol{\beta}}$:

$$\min_{\{\widehat{\boldsymbol{\beta}}\}}(\boldsymbol{e}'\boldsymbol{e}) = \min_{\{\widehat{\boldsymbol{\beta}}\}}[(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})],$$

where $\boldsymbol{e} = [e_1, \dots, e_n]'$. Solving this minimization problem, the OLS estimator is:

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \tag{2.13}$$

which is a linear and unbiased estimator of $\boldsymbol{\beta}$ regardless of the distribution of the error $\boldsymbol{\varepsilon}$.

---

[10] A residual e_i is the difference between the actual value (y_i) and the fitted value $(\hat{y}_i)$ of the dependent variable, for some arbitrary estimator of β, $\widehat{\beta}$, i.e., $e_i = y_i - \hat{y}_i$.

The covariance matrix of $\boldsymbol{b}$ is given by $Var[\boldsymbol{b}|\boldsymbol{X}] = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$, where σ^2 is the finite population variance of the error term. A natural estimator of σ^2 is given by $\hat{\sigma}^2 = n^{-1}(\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}})$, where $\hat{\boldsymbol{e}}$ is the OLS residual vector , $\hat{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}$. The estimator $\hat{\sigma}^2$ is biased, with the bias inversely related to the sample size. Under classical assumptions, an unbiased estimator of σ^2 is given by: $\hat{\sigma}^2 = \frac{\hat{e}'\hat{e}}{n-k}$, where $(n-k)$ is the degrees of freedom – the number of independent sources of information in the "n" residuals. Hence, using $\hat{\sigma}^2$, a semiparametric estimator of the covariance of $\boldsymbol{b}$ is

$$\widehat{Var[\boldsymbol{b}|\boldsymbol{X}]} = \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{2.14}$$

The OLS estimator $\boldsymbol{b}$ is a consistent estimator of $\boldsymbol{\beta}$ in the model. As $\hat{\sigma}^2$ is a consistent estimator of σ^2, the estimated covariance of the estimator $\boldsymbol{b}$ is also consistent.

A detailed discussion on various ways of fitting regression models to estimate population parameters and their respective variance estimators that allow for various sampling designs can be found in Wolter (2007), Cameron and Trivedi (2005, p. 813-858) and Skinner et al. (1989, p. 231-260). For instance, Wolter (2007, p. 249-253) shows that the regression estimator for a probability sample, s, can be represented by

$$\boldsymbol{b_s} = (\boldsymbol{X}_s'\boldsymbol{W}_s\boldsymbol{X}_s)^{-\mathbf{1}}\boldsymbol{X}_s'\boldsymbol{W}_s\boldsymbol{Y}_s, \tag{2.15}$$

where $\boldsymbol{Y}_s = \left(y_1, \dots, y_{n(s)}\right)'$, $\boldsymbol{X}_s = \begin{pmatrix} \boldsymbol{X_1} \\ \vdots \\ \boldsymbol{X_{n(s)}} \end{pmatrix}$, $\boldsymbol{W}_s = \text{diag}\left(w_1, \dots, w_{n(s)}\right)$ and $n(s)$ is the size of the probability sample s, an appropriate weight matrix, such that, for instance, w_i is the reciprocal of the probability of inclusion, and takes other survey features, e.g., nonresponse, stratification, clustering, into account. An estimator of the covariance matrix of the sample-based estimated regression coefficient, $\boldsymbol{b_s}$, is given by

$$\widehat{Var(\boldsymbol{b_s})} = (\boldsymbol{X}_s'\boldsymbol{W}_s\boldsymbol{X}_s)^{-1}\widehat{\boldsymbol{T}}(\boldsymbol{X}_s'\boldsymbol{W}_s\boldsymbol{X}_s)^{-1}, \tag{2.16}$$

where $\widehat{\boldsymbol{T}}$ is an estimator of the $(K \times K)$ covariance matrix $\boldsymbol{T}$ of the vector of estimated totals $\boldsymbol{X}_s'\boldsymbol{W}_s\boldsymbol{E}_s$, where $\boldsymbol{E}_s = (\hat{e}_1, \dots, \hat{e}_{n(s)})'$, is the vector of LS residuals, that account for the sampling designs.

We detail some auxiliary regressions in Section 2.7 that enable us to estimate G using $\hat{G}$, and its variance estimator in a straightforward way. This can be carried out through commonly used statistical software packages that may or may not account for the survey design. As we are concerned with generating estimates only, not with any intrinsic properties of regression models, we do not detail any such properties here.

2.4 Estimating the Gini Coefficient

The population Gini coefficient is

$$G = 1 - 2\int_0^1 L(p)dp, \tag{2.17}$$

where $L(p) = \frac{1}{\mu}\int_0^{\xi_p} ydF(y)$ is the Lorenz curve for the pth percentile of the population and $p = F(\xi_p) = \int_0^{\xi_p} dF(y)$, with ξ_p being the pth quantile function for $0 \leq p \leq 1$, $\xi_p = F^{-1}(p) = \inf\{y \mid F(y) \geq p\}$, and $\mu = \int_0^{\infty} ydF(y)$ is the population mean of the well-being random variable $Y \in [0, \infty)$ with cumulative distribution function $F(y)$.[11] As shown in Appendix A, G can equivalently be expressed as:

$$G = \frac{2}{\mu}\int_0^{\infty} yF(y)dF(y) - 1. \tag{2.18}$$

---

[11] As defined in subsection 2.3.1, the CDF of a random variable Y is defined as $F(y) = \Pr(Y \leq y)$, $0 \leq F(y) \leq 1$ for all y. The funcion F is non-decreasing and is twice continuously differentiable with density $f(y) = F'(y)$.

2.4.1 Estimating the Gini Coefficient with an *iid* Sample

To estimate G expressed in (2.18) with an *iid* sample, we first define the empirical distribution function (EDF) of the random variable y. For an *iid* sequence $\{Y_i\}_{i=1}^{N}$, the corresponding EDF, $\hat{F}_n$, for any ordered[12] sample of size $n, \{y_{(i)}: i = 1, \dots, n\}$, is represented as

$$\hat{F}_n(y_{(i)}) = \frac{1}{n}\sum_{j=1}^{n} I\{y_j \leq y_{(i)}\}\,, \qquad\qquad y_j \in [0, \infty)$$

where $I(\bullet)$ is an indicator function satisfying $I\{y_j \leq y_{(i)}\} = \begin{cases} 1 & \text{if} \quad y_j \leq y_{(i)} \\ 0 & \text{if} \quad y_j > y_{(i)} \end{cases}$. With the sample mean $\hat{\mu} = n^{-1}\sum_{i=1}^{n} y_i = n^{-1}\sum_{i=1}^{n} y_{(i)}$, an estimator for G is given by

$$\hat{G} = 2(\hat{\mu}n)^{-1}\sum_{i=1}^{n} y_{(i)}\,\hat{F}_n(y_{(i)}) - 1\,. \tag{2.19}$$

We have the following results (e.g., see Serfling, 1980, p.55-67): $\hat{F}_n(y_{(i)})$ is an unbiased estimator of $F(y)$: $E\{\hat{F}_n(y_{(i)})\} = F(y)$. Then, as shown by Serfling, using the Glivenko-Cantelli theorem, as the Kolmogorov-Smirnov distance converges to 0 with probability 1, $\sup_{y_{(i)}}\left|\hat{F}_n(y_{(i)}) - F(y)\right| \xrightarrow{asym} 0$.

Hence, $\hat{F}_n(y_{(i)})$ is a consistent estimator of $F(y)$ for each fixed $y_{(i)}$, so that $\hat{F}_n(y_{(i)}) \xrightarrow{p} F(y)$. For estimation purposes, the EDF of the random sample when the y's are expressed as order statistics is given by $\hat{F}_n(y_{(i)}) = \frac{i}{n}$. Plugging this into the formula for the sample G in expression (2.19), we obtain

$$\hat{G} = \frac{2}{\hat{\mu}n^2}\sum_{i=1}^{n} i y_{(i)} - 1. \tag{2.20}$$

---

[12] The sequence $\{y\}_{(i)=1}^{n}$ is $y_{(1)} \leq \cdots \leq y_{(n)}$ and denotes the order statistics of $y_1, \dots y_n$.

Davidson (2009) evaluates the EDF at its points of discontinuity by the average of the lower and upper limits: $\hat{F}_n\left(y_{(i)}\right) = \left(\frac{i}{n} + \frac{i-1}{n}\right)/2 = n^{-1}\left(i - \frac{1}{2}\right)$, which leads to $\hat{G}$ given by

$$\hat{G} = \frac{2}{\hat{\mu}n^2}\sum_{i=1}^{n} y_{(i)}\left(i - \frac{1}{2}\right) - 1. \tag{2.21}$$

However, to estimate G with a non-*iid* sample, especially with a complex survey sample, it is essential to incorporate the sampling design and appropriate sampling weights in the estimator. In the following subsection (2.4.2), we obtain a plug-in estimator for G with a complex survey sample. Prior to that, we provide a brief discussion on sampling designs and sampling weights, as they pertain to a complex survey sample.

2.4.2 Estimating the Gini Coefficient with a Complex Survey Sample

2.4.2.1 Sampling Designs and Sampling Weights

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, within each stratum there are N_h clusters or primary sampling units, 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} households, so that the total number of households in stratum h is $M_h = \sum_{c=1}^{N_h} M_{hc}$ and in the population is $\mathcal{M} = \sum_{h=1}^{H}\sum_{c=1}^{N_h} M_{hc}$. For example, suppose the population is stratified by location of residence – rural and urban. The clusters are then all villages in the rural area and all blocks in the urban area. Normally, in a sampling design, a sample of clusters is selected with probability proportional to the cluster size (PPS) (i.e., the number of households in a cluster) and sampling is independent from stratum to stratum. In the second stage, a fixed number of households from each cluster are selected with replacement[13] using simple random sampling (SRS). A further stage of sampling may be undertaken when not all individuals from the sampled households are interviewed. For instance, in the Bangladesh Demographic and Heath Surveys (BDHS) 2007, individuals for interview are chosen from a second stage design based on further two-stage stratification. The first stratification is by gender: male and female; the second stratification is whether a female is ever-married and aged 10-49 and slept in the selected household the night before the survey is conducted. For the male survey, similar criteria are applied but for an age group of 15-54. However, such further sampling stages do not complicate our analysis, as the nonparametric variance estimator we consider is calculated from the quantities formed from the M_{hc} households (e.g., Skinner et al., 1989, p. 47-48).

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

---

[13] Sampling with and without replacement is important when the sample size is large relative to the population. Most studies assume sampling with replacement in order to ignore the finite population correction factors (e.g., Deaton, 1997, p. 44). When the sample is a small fraction of the population, sampling with and without replacement may not be a concern. Moreover, when applying an asymptotic theory with the number of clusters going to infinity, sampling with or without replacement has no effect on the asymptotic results (e.g., Bhattacharya, 2005, p. 148).

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.

Now suppose a sample of N clusters is drawn from the population and within each stratum the sampling is performed independently. Let n_h be the number of sampled clusters from the hth stratum, i.e., the total number of clusters in the sample is $N = \sum_{h=1}^{H} n_h$. From cluster c within stratum h we suppose a fixed number of households, m, are selected, i.e., the total number of households in the sample is $M = m\sum_{h=1}^{H} n_h = mN$. If s_{hci} is the number of members in the ith household in the cth cluster in the hth stratum, then the total number of individuals or elements in the finite sample is $n_0 = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} s_{hci}$.

As the complex survey design may involve different sampling techniques at different stages and post survey adjustments to balance unexpected nonresponse by some of the households or missing variable values, an appropriate weight needs to be attached to the household. Without appropriate weighting, obtained estimators may not have desirable statistical properties such as unbiasedness or consistency (e.g., Lehtonen and Pahkinen, 1995, p. 2; Deaton, 1997, p. 15). The weight is constructed as follows: within stratum h, the probability that the cth cluster is selected in the first stage is $\pi_{hc} = M_{hc}/\sum_{c=1}^{N_h} M_{hc}$, probability proportional to size. The probability that the ith household in the cth cluster is selected in the second stage, given that cth cluster in the hth stratum is selected in the first stage, is $\pi_{hci} = 1/M_{hc}$, equal probability. Then, let $\Pi_{hc} = (n_h \pi_{hc})^{-1}$ be the number of population clusters in stratum h represented by the cth cluster, and let $\Pi_{hci} = (m\pi_{hci})^{-1}$ be the number of cth cluster households in stratum h represented by the ith household in the cth cluster in the hth stratum (e.g., see Deaton, 1997, p. 52-55). Now the weight attached to a member of the ith household in the cth cluster in the hth stratum is given by

$$\begin{aligned} W_{hci} &= (\Pi_{hc}\Pi_{hci})s_{hci} \\ &= \left(n_h \times \frac{M_{hc}}{\sum_{c=1}^{N_h} M_{hc}}\right)^{-1} \left(m \times \frac{1}{M_{hc}}\right)^{-1} s_{hci} = \left(\frac{m\, n_h}{\sum_{c=1}^{N_h} M_{hc}}\right)^{-1} s_{hci} \\ &= \frac{\sum_{c=1}^{N_h} M_{hc}}{m\, n_h}\, s_{hci}\,. \end{aligned} \tag{2.22}$$

Although this is similar to the probability in a SRS, a complex survey design is not the same as a SRS as the computation of the estimators involves various-stage designs (e.g., Deaton, 1997, p. 55) and post survey adjustments. The sum of the weights might be considered as an estimator of the total number of households $\mathcal{M}$, given by $\widehat{\mathcal{M}} = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} W_{hci}$. Typically with many surveys, the weights are normalized to ensure that $\widehat{\mathcal{M}} = M$. In addition, sampling weights are often normalized to sum to unity, using

$$w_{hci} = \frac{W_{hci}}{\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} W_{hci}}\,. \tag{2.23}$$

To illustrate the use of weights in estimation with a complex survey, consider, for example, estimating a finite population total τ_0 of a well-being variable y. The finite population total, τ_0, is the sum of all individual values of the random variable Y_{hci} for elements i in cluster c in stratum h, given by $\tau_0 = \sum_{h=1}^{H}\sum_{c=1}^{N_h}\sum_{i=1}^{M_{hc}} Y_{hci}$, which can be estimated by summing appropriately weighted values over all the units in the sample. An estimator for the total is $\hat{\tau} = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} W_{hci} y_{hci}$.

When individual elements from a cluster are first considered as $y_{hc} = \sum_{i=1}^{m} W_{hci} y_{hci}$, the above estimator can also be expressed as $\hat{\tau} = \sum_{h=1}^{H} \sum_{c=1}^{n_h} y_{hc}$. Another example of use of a complex sampling design is in estimating the population mean, $\mu_0 = \mathcal{M}^{-1} \sum_{h=1}^{H} \sum_{c=1}^{N_h} \sum_{i=1}^{M_{hc}} Y_{hci}$, an estimator of which is given by

$$\hat{\mu} = \frac{\sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} W_{hci} y_{hci}}{\sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{k} W_{hci}} = \sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci} y_{hci} \, .$$

To estimate G using such ideas, we need to obtain the EDF under a complex survey design. In the following subsection, we define this and provide a discussion of the consistency of the estimator. We then provide the plug-in estimator for G.

2.4.2.2 EDF, Consistency, and the Plug-in Estimator for the Gini Coefficient with a Complex Survey

For a finite population which is divided into H strata with $\{\mathcal{U}[hci]: h = 1, \ldots, H; c = 1, \ldots, N_h; i = 1, \ldots, M_{hc}\}$, the distribution function for the random variable Y is given by

$$F(y) = \mathcal{M}^{-1} \sum_{h=1}^{H} \sum_{c=1}^{N_h} \sum_{i=1}^{M_{hc}} I\{Y_{hci} \leq y\},$$

where $\mathcal{M}$ is the size of the population in terms of households, with $\mathcal{M} = \sum_{h=1}^{H} \sum_{c=1}^{N_h} M_{hc}$. A plug in estimator of $F(y)$, the empirical distribution function, for a complex sample of size M, $\{u[hci]: h = 1, \ldots, H; c = 1, \ldots, n_h; i = 1, \ldots, m\}$, (see, e.g., Bhattacharya, 2007) is

$$\hat{F}(y) = \widehat{\mathcal{M}}^{-1} \sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} W_{hci} I\{y_{hci} \leq y\}$$

$$= \frac{\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} W_{hci} I\{y_{hci} \leq y\}}{\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} W_{hci}}, \tag{2.24}$$

and with the normalized weight as defined in expression (2.23), the EDF for the complex survey sample is

$$\hat{F}(y) = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci} I\{y_{hci} \leq y\}. \tag{2.25}$$

The question of consistency of $\hat{F}(y)$ with a complex survey is interesting as the asymptotic consistency of an estimator can be determined either by the number of the strata going to infinity or by the number of the clusters going to infinity, depending on the nature of the survey and the analysis under study. For example, Kovačević and Binder (1997), Binder and Kovačević (1995), Rao and Wu (1985) and Krewski and Rao (1981) consider the consistency of estimators as the number of strata H tends to infinity assuming that the number of clusters in each of the stratum is fixed. On the other hand, Dippo (1984) and Bhattacharya (2007, 2005) examine consistency with respect to the number of clusters within the sample, assuming that the number of strata is constant. Specifically, they let the number of clusters within each stratum, n_h, and, hence, the total number of clusters in the sample, N, go to infinity. With this asymptotic analysis, the ratio $a_h = \frac{n_h}{N}$ is assumed constant. This seems more reasonable when there are only a few strata in the sample.

As applications in Chapter 4 using the DHS survey data, typically have few strata and many clusters, we follow Bhattacharya's (2007, 2005) approach. We consider the consistency of estimators[14] allowing for the number of clusters within each stratum to

---

[14] As discussed, for an *iid* sample, consistency refers to a limiting property of an estimator, often regarded as a minimal requirement for an estimator. With an *iid* sample, the property states that as the sample size n increases, the sequence of estimators $\{\hat{\theta}_n\}$ approaches the true parameter θ in probability, i.e., plim $\hat{\theta}_n = \theta$. By consistency here, we mean the weak or simple consistency of an estimator. For a detailed discussion of consistency in estimation, see e.g., Serfling (1980, p. 47-57).

increase, while keeping the number of strata fixed. Hence, the consistency of the estimator $\hat{F}(y)$ is examined with regard to the total number of clusters, $N = \sum_{h=1}^{H} n_h$, and for a large N, we have that $\hat{F}(y) \xrightarrow{p} F(y)$ (e.g., Francisco and Fuller, 1986).

Given this, a plug-in estimator for G with data from a complex survey (e.g., Binder and Kovačević, 1995) is

$$\hat{G} = \frac{2}{\hat{\mu}} \sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci}\, y_{hci} \hat{F}(y_{hci}) - 1, \tag{2.26}$$

where $\hat{\mu} = \sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci}\, y_{hci}$ and $\hat{F}(y_{hci}) = \sum_{r=1}^{H} \sum_{s=1}^{n_r} \sum_{t=1}^{m} w_{rst} I\{y_{rst} \leq y_{hci}\}$.

An equivalent estimator for G with the ordered form of the complex survey sample is given by

$$\hat{G} = \frac{2}{\hat{\mu}} \sum_{d=1}^{M} w_d y_{(d)} \hat{F}\left(y_{(d)}\right) - 1, \tag{2.27}$$

where $y_{(d)}$ is the dth order statistic in the full sample and w_d is its associated sampling weight, and the EDF is $\hat{F}\left(y_{(d)}\right) = \sum_{l=1}^{M} w_l \hat{F}\left(y_l \leq y_{(d)}\right)$. 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.

Binder and Kovačević (1995) obtain this estimator for the Gini coefficient solving the following sample estimating equation under a complex survey design

$$\sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci} \left\{2\hat{F}(y_{hci}) y_{hci} - y_{hci} - \hat{G} y_{hci}\right\} = 0. \tag{2.28}$$

It is straightforward to show that the solution to (2.28) for $\hat{G}$ yields the same estimator as given in expression (2.26). It can also be shown that when the Gini coefficient in expression (2.17) is estimated with a complex survey, it yields the same estimator as in expression (2.26) (see Appendix B.1 for a derivation).

2.5 An approximation for $(\hat{G} - G)$

When obtaining an asymptotic variance estimator for $\hat{G}$ via a traditional linearization method with a complex survey, design features, and weights make the estimator more complicated compared to an estimator based on an *iid* assumption. To estimate the variance for $\hat{G}$, it is convenient to obtain an approximation for $(\hat{G} - G)$ first.

There are several ways to obtain an approximation for $(\hat{G} - G)$. Binder and Kovačević (1995) derive a general approximation form based on the EE method, and for $(\hat{G} - G)$ they use the Gini coefficient expression given in Nygård and Sandström (1981), $G_J = \frac{1}{\mu_Y}\int_0^\infty J[F(y)]y\, dF(y)$, where $J[F(y)] = [2F(y) - 1]$. Davidson (2009) derives an asymptotic approximation for $\hat{G}$ using the conventional method for an *iid* sample. On the other hand, Bhattacharya (2007) presents an approximation for $(\hat{G} - G)$ via an approximation for the Lorenz curve, $(\hat{L} - L)$. Although Binder and Kovačević's method to obtain the approximation is based on a general version of the Gini coefficient and is easily adaptable to our plug-in estimator, for completeness we present the derivation for the approximation for $(\hat{G} - G)$ directly in line with Davidson (2009). Our derivation is a simplification of that provided by Binder and Kovačević (1995).

The Gini coefficient, $G = \frac{2}{\mu}\int_0^\infty yF(y)dF(y) - 1$, can be written as

$$\left(\frac{G+1}{2}\right)\mu = \int_0^\infty yF(y)dF(y), \tag{2.29}$$

with estimator:

$$\left(\frac{\hat{G}+1}{2}\right)\hat{\mu} = \int_0^\infty y\hat{F}(y)d\hat{F}(y), \tag{2.30}$$

where $\hat{\mu} = \int_0^\infty yd\hat{F}(y)$ and $\hat{F}(y)$ is the empirical distribution function for any sample y.

Now the right hand side of (2.30) can be further expanded as

$$\int_0^\infty yF(y)d\hat{F}(y) - \int_0^\infty yF(y)dF(y) + \int_0^\infty y\hat{F}(y)dF(y) + \mathbb{R}\,,$$

where $\mathbb{R} = \int_0^\infty y\big[\hat{F}(y) - F(y)\big]d\big[\hat{F}(y) - F(y)\big]$. The reminder term $\mathbb{R}$ is asymptotically negligible, under the condition that $\hat{F}(y) - F(y) \xrightarrow{p} 0$, which is typically guaranteed.

Ignoring the reminder term $\mathbb{R}$ and rearranging the expression in (2.30), we obtain

$$\hat{G} \approx \frac{2}{\hat{\mu}}\left[\int_0^\infty yF(y)d\hat{F}(y) - \int_0^\infty yF(y)dF(y) + \int_0^\infty y\hat{F}(y)dF(y)\right] - 1. \tag{2.31}$$

Using the expression in (2.31), an approximation for $\big(\hat{G} - G\big)$ is

$$\begin{aligned}\hat{G} - G &\approx \frac{2}{\hat{\mu}}\left[\int_0^\infty yF(y)d\hat{F}(y) - \int_0^\infty yF(y)dF(y) + \int_0^\infty y\hat{F}(y)dF(y)\right] - \frac{2}{\mu}\int_0^\infty yF(y)dF(y) \\ &= \frac{2}{\hat{\mu}\mu}\Bigg[\mu\int_0^\infty yF(y)d\hat{F}(y) - \mu\int_0^\infty yF(y)dF(y) + \mu\int_0^\infty y\hat{F}(y)dF(y) \\ &\qquad - \hat{\mu}\int_0^\infty yF(y)dF(y)\Bigg].\end{aligned}$$

Using expression (2.29) and $\hat{\mu} = \int_0^\infty yd\hat{F}(y)$, we can write

$$\begin{aligned}\hat{G} - G \approx \frac{2}{\hat{\mu}\mu}\Bigg[&\mu\int_0^\infty yF(y)d\hat{F}(y) - \mu\left(\frac{G+1}{2}\right)\mu + \mu\int_0^\infty y\hat{F}(y)dF(y) \\ &- \int_0^\infty yd\hat{F}(y)\left(\frac{G+1}{2}\right)\mu\Bigg].\end{aligned}$$

Now, as $\hat{\mu} \xrightarrow{p} \mu$, we can replace the denominator by μ^2, leading to

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

$$\frac{2}{\mu}\left[\int_0^\infty yF(y)d\hat{F}(y) - \left(\frac{G+1}{2}\right)\left(\mu + \int_0^\infty y\, d\hat{F}(y)\right) + \int_0^\infty y\hat{F}(y)dF(y)\right]. \quad (2.32)$$

The term $\int_0^\infty y\hat{F}(y)dF(y)$ in expression (2.32) can be written as

$$\int_0^\infty y \int_0^y d\hat{F}(x)\, dF(y) = \int_0^\infty \int_x^\infty ydF(y)\, d\hat{F}(x)$$

$$= \int_0^\infty \left[\int_y^\infty xdF(x)\right] d\hat{F}(y),$$

and, as μ is a parameter, we can write

$$\mu = \mu \int_0^\infty d\hat{F}(y) = \int_0^\infty \mu\, d\hat{F}(y),$$

so that the expression (2.32) is now

$$\hat{G} - G \approx \frac{2}{\mu}\int_0^\infty \left[yF(y) - \left(\frac{G+1}{2}\right)(\mu + y) + \int_y^\infty xdF(x)\right] d\hat{F}(y). \quad (2.33)$$

For an *iid* sample of size n, the expression (2.33) is given by

$$\hat{G} - G \approx \frac{2}{\mu}\frac{1}{n}\sum_{i=1}^{n}\left[yF(y) - \left(\frac{G+1}{2}\right)(\mu + y) + \int_y^\infty xdF(x)\right], \quad (2.34)$$

which is equivalent to the approximation given in Davidson (2009, p. 32). For a complex survey sample, the approximation for $(\hat{G} - G)$ in (2.33) can be written as

$$\hat{G} - G \approx \frac{2}{\mu}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\left[yF(y) - \left(\frac{G+1}{2}\right)(\mu + y) + \int_y^\infty xdF(x)\right]. \quad (2.35)$$

When estimated with a complex survey sample, we obtain

$$\hat{G} - G \approx \frac{2}{\hat{\mu}}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\,\hat{\Omega}_{hci}\,, \quad (2.36)$$

where the estimator $\hat{\Omega}_{hci}$ is

$$\hat{\Omega}_{hci} = y_{hci}\hat{F}(y_{hci}) - \left(\frac{\hat{G}+1}{2}\right)(\hat{\mu} + y_{hci}) \;+ \sum_{r=1}^{H}\sum_{s=1}^{n_r}\sum_{t=1}^{m} w_{rst}\, y_{rst} I\{y_{rst} \geq y_{hci}\}.$$

This approximation for $(\hat{G} - G)$ is equivalent to that obtained by Binder and Kovačević (1995, p. 140-143) with the complex survey, given by,

$$\hat{G} - G \approx \sum_{h=1}^{H}\sum_{c=1}^{n_h} u_{hc}^{*}\ , \tag{2.37}$$

$$\text{where } u_{hc}^{*} = \sum_{i=1}^{m} w_{hci}\left\{\frac{2}{\hat{\mu}}\left[y_{hci}\left(\hat{F}(y_{hci}) - \frac{(\hat{G}+1)}{2}\right) + B(y_{hci}) - \frac{\hat{\mu}}{2}(\hat{G}+1)\right]\right\}, \tag{2.38}$$

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

We now turn to the work of Bhattacharya (2007), who frames estimation as a GMM problem, using an approximation for the Lorenz share at a fixed percentile p. After rearranging an "influence function" for $\widehat{G}$ (Bhattacharya, 2007, p. 685), Bhattacharya's approximation for $(\widehat{G} - G)$ is

$$\hat{G} - G \approx -2\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\,\widehat{\Phi}_{hci}, \tag{2.39}$$

where

$$\widehat{\Phi}_{hci} = \sum_{d=1}^{M} w_d \left[\begin{array}{c} \frac{1}{\hat{\mu}}\left(y_{hci}I\{y_{hci} \leq y_{(d)}\} - \hat{\alpha}(y_{(d)})\right) \\ +\frac{1}{\hat{\mu}}\left(y_{(d)}\left(\hat{F}(y_{(d)}) - I\{y_{hci} \leq y_{(d)}\}\right)\right) - \frac{\hat{\alpha}(y_{(d)})}{\hat{\mu}^2}(y_{hci} - \hat{\mu}) \end{array}\right], \tag{2.40}$$

and $y_{(d)}$ is the dth order statistic in the full sample and w_d its associated sampling weight, $\hat{F}(y_{(d)}) = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\, I\{y_{hci} \leq y_{(d)}\}$ and $\hat{\alpha}(y_{(d)}) =$

$\sum_{r=1}^{H}\sum_{s=1}^{n_r}\sum_{t=1}^{m} w_{rst}y_{rst}I\{y_{rst} \le y_{(d)}\}$, where $\Phi_{hci} = \int_0^1 \frac{1}{\mu}\Big[y_{hci}I\{y_{hci} \le z(p)\} - \alpha(p) +$

$z(p)(p - I\{y_{hci} \le z(p)\}) - \frac{\alpha(p)}{\mu}(y_{hci} - \mu)\Big]\,\mathrm{d}p,$

$\alpha(p) = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}y_{hci}I\{y_{hci} \le z(p)\}$, and $z(p) = F^{-1}(p)$.

Although Bhattacharya's (2007) presented approximation for $(\hat{G} - G)$ comes via a different approach, it can be shown that this is equivalent to the approximation derived much earlier by Binder and Kovačević (1995). We show the mathematical equivalency of the two approximations, given in (2.37) and (2.39), below. To do that, it is sufficient to show that $-2\sum_{i=1}^{m} w_{hci}\widehat{\Phi}_{hci}$ and u_{hc}^{*} are equivalent using expressions (2.38) and (2.40). To proceed, we start with rearranging expression (2.40):

$$\widehat{\Phi}_{hci} = \frac{1}{\hat{\mu}}\left[\begin{array}{c} \sum_{d=1}^{M} w_d y_{hci} I\{y_{hci} \le y_{(d)}\} - \sum_{d=1}^{M} w_d y_{(d)} I\{y_{hci} \le y_{(d)}\} \\ + \sum_{d=1}^{M} w_d y_{(d)}\,\hat{F}(y_{(d)}) - \sum_{d=1}^{M} w_d \left(\frac{\hat{\alpha}(y_{(d)})}{\hat{\mu}}\right) y_{hci} \end{array}\right] \tag{2.41}$$

Individual components in the expression are manipulated to obtain Binder and Kovačević's expression. The first term

$$\begin{aligned} \sum_{d=1}^{M} w_d y_{hci} I\{y_{hci} \le y_{(d)}\} &= y_{hci}\sum_{d=1}^{M} w_d I\{y_{(d)} \ge y_{hci}\} \\ &= y_{hci}\left[1 - \sum_{d=1}^{M} w_d I\{y_{(d)} \le y_{hci}\}\right] \\ &= y_{hci} - y_{hci}\hat{F}(y_{hci}). \end{aligned}$$

The second term,

$$\sum_{d=1}^{M} w_d y_{(d)} I\{y_{hci} \le y_{(d)}\} = \sum_{d=1}^{M} w_d y_{(d)} I\{y_{(d)} \ge y_{hci}\} = B(y_{hci}).$$

The third term, $\sum_{d=1}^{M} w_d y_{(d)} \hat{F}\big(y_{(d)}\big) = \frac{\hat{\mu}}{2}\big(\hat{G}+1\big)$.

Finally, the fourth term

$$\sum_{d=1}^{M} w_d \left(\frac{\hat{\alpha}\big(y_{(d)}\big)}{\hat{\mu}}\right) y_{hci} = \frac{y_{hci}}{\hat{\mu}}\left[\sum_{d=1}^{M} w_d \left(\sum_{r=1}^{H}\sum_{s=1}^{n_r}\sum_{t=1}^{m} w_{rst} y_{rst} I\big\{y_{rst} \le y_{(d)}\big\}\right)\right]$$

$$= \frac{y_{hci}}{\hat{\mu}}\left[\sum_{r=1}^{H}\sum_{s=1}^{n_r}\sum_{t=1}^{m} w_{rst} y_{rst} \left(\sum_{d=1}^{M} w_d\, I\big\{y_{(d)} \ge y_{rst}\big\}\right)\right]$$

$$= \frac{y_{hci}}{\hat{\mu}}\left[\sum_{r=1}^{H}\sum_{s=1}^{n_r}\sum_{t=1}^{m} w_{rst} y_{rst} \left(1 - \hat{F}(y_{rst})\right)\right]$$

$$= y_{hci} - \frac{y_{hci}}{\hat{\mu}}\left[\sum_{r=1}^{H}\sum_{s=1}^{n_r}\sum_{t=1}^{m} w_{rst} y_{rst}\ \hat{F}(y_{rst})\right]$$

$$= y_{hci} - \frac{y_{hci}}{\hat{\mu}}\left[\frac{(\hat{G}+1)}{2}\hat{\mu}\right] = y_{hci} - \frac{(\hat{G}+1)}{2} y_{hci}\,.$$

Using these results, we have

$$\hat{\Phi}_{hci} = \frac{1}{\hat{\mu}}\left[y_{hci} - y_{hci}\hat{F}(y_{hci}) - B(y_{hci}) + \frac{\hat{\mu}}{2}\big(\hat{G}+1\big) - y_{hci} + \frac{(\hat{G}+1)}{2} y_{hci}\right]$$

$$= \frac{1}{\hat{\mu}}\left[-y_{hci}\hat{F}(y_{hci}) + \frac{(\hat{G}+1)}{2} y_{hci} - B(y_{hci}) + \frac{\hat{\mu}}{2}\big(\hat{G}+1\big)\right],$$

and

$$-2\sum_{i=1}^{m} w_{hci}\hat{\Phi}_{hci} = \frac{2}{\hat{\mu}}\sum_{i=1}^{m} w_{hci}\left[y_{hci}\hat{F}(y_{hci}) - \frac{(\hat{G}+1)}{2} y_{hci} + B(y_{hci}) - \frac{\hat{\mu}}{2}\big(\hat{G}+1\big)\right]$$

$$= u_{hc}^{*}\,. \qquad (2.42)$$

Therefore, the approximation for $\left(\hat{G}-G\right)$ in (2.37) derived by Binder and Kovačević (1995) and the approximation for $\left(\hat{G}-G\right)$ in (2.39) presented by Bhattacharya (2007) are algebraically equivalent, despite having originated from different premises.

It can also be shown that Binder and Kovačević's (1995) approximation expression in (2.37) for the case of a randomly drawn *iid* sample is equivalent to that obtained by Davidson (2009). In the *iid* case, expression (2.37) becomes

$$\hat{G}-G \approx \frac{1}{M}\sum_{j=1}^{M}\frac{2}{\hat{\mu}}\left[y_j\left(\hat{F}(y_j)-\left(\frac{\hat{G}+1}{2}\right)\right)+B(y_j)-\frac{\hat{\mu}}{2}(\hat{G}+1)\right] \tag{2.43}$$

where $B(y_j)=\frac{1}{M}\sum_{i=1}^{M}y_i I(y_i\geq y_j)$. Using Davidson's notation, $B(y_j)=\hat{\mu}-\widehat{m}(y_j)$, $\frac{\hat{I}}{\hat{\mu}}=\frac{(\hat{G}+1)}{2}$ 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

$$\hat{G}-G \approx \frac{1}{M}\sum_{j=1}^{M}\frac{2}{\hat{\mu}}\left[-\frac{\hat{I}}{\hat{\mu}}(y_j-\hat{\mu})+y_j\hat{F}(y_j)-\widehat{m}(y_j)-(2\hat{I}-\hat{\mu})\right], \tag{2.44}$$

which is Davidson's (p. 32) estimated approximation. Note that we can equivalently write (2.42) as[15]

$$\hat{G}-G \approx \frac{1}{M}\sum_{j=1}^{M}u_j^*, \tag{2.45}$$

where $u_j^*=\frac{2}{\hat{\mu}}\left[y_j\left(\hat{F}(y_j)-\frac{\hat{G}+1}{2}\right)+B(y_j)-\frac{\hat{\mu}}{2}(\hat{G}+1)\right]=\frac{(\hat{Z}_i-\bar{Z})}{\hat{\mu}}$, where with order statistics, as shown by Davidson, if $y_j=y_{(i)}$, $\hat{Z}_i=-(\hat{G}+1)y_{(i)}+\frac{2i-1}{n}y_{(i)}-\frac{2}{n}\sum_{j=1}^{i}y_{(j)}$ and $\bar{Z}=n^{-1}\sum_{i=1}^{n}\hat{Z}_i$. Given these approximations, we now turn to variance estimators.

---

[15] A proof is provided in Appendix B2.

2.6 Variance Estimation

As the Gini coefficient estimator is a complex nonlinear statistic, it is usual to examine its asymptotic distribution, e.g., Davidson (2009), Bhattacharya (2007), Binder and Kovačević (1995), Nygård and Sandström (1989, 1983), , Gastwirth and Gail (1985), Beach and Davidson (1983), Glasser (1962), Hoeffding (1948). However, in most of these studies the formula for the asymptotic variance for $\hat{G}$ is so complicated that it is easy to understand why applied researchers refrain from attempting to code the variance estimator. Typically, those few applied studies that consider reporting a variance do so using resampling methods (e.g., Nho, 2006 and studies cited therein). In this section, we consider an analytic, asymptotic, variance that can be readily obtained in practice.

To begin, to illustrate techniques, we first examine a simple illustration of variance estimation for the sample mean; see, e.g., Skinner et al. (1989, p.46-47). The variance of the estimator of the population mean, $\hat{\mu}$, with a complex survey design is given by

$$Var(\hat{\mu}) = Var\left(\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}y_{hci}\right). \qquad (2.46)$$

As strata are independent but not identical, we estimate this variance by adding up all strata variances. So,

$$Var(\hat{\mu}) = \sum_{h=1}^{H} Var\left(\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}y_{hci}\right).$$

In each stratum, as clusters, which are identical but not independent, are assumed to be selected via PPS from each stratum, we have that the cluster variances within a stratum are identical, and, hence,

$$Var(\hat{\mu}) = \sum_{h=1}^{H} n_h Var\left(\sum_{i=1}^{m} w_{hci} y_{hci}\right). \tag{2.47}$$

So, it turns out that with a complex survey, the variance of an estimator, e.g., $\hat{\mu}$, is obtained using the variance of a cluster total. Following, for instance, Skinner et al. (1989, p. 47), an estimator of $Var(\hat{\mu})$ is given by

$$\widehat{Var(\hat{\mu})} = \sum_{h=1}^{H} \frac{n_h}{n_h - 1} \sum_{c=1}^{n_h} \left(\sum_{i=1}^{m} w_{hci} y_{hci} - n_h^{-1} \sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci} y_{hci}\right)^2. \tag{2.48}$$

Adopting this notion for asymptotic behaviour of the variance estimator, $\frac{n_h}{n_h - 1} \to 1$ as $n_h \to \infty$, we also have that

$$\widehat{Var(\hat{\mu})} = \sum_{h=1}^{H} \sum_{c=1}^{n_h} \left(\sum_{i=1}^{m} w_{hci} y_{hci} - n_h^{-1} \sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci} y_{hci}\right)^2. \tag{2.49}$$

Using the same approach, an estimator for the variance of $\hat{G}$, using the approximation in (2.36), is given by

$$\widehat{Var(\hat{G})} = \frac{4}{\hat{\mu}^2} \sum_{h=1}^{H} \sum_{c=1}^{n_h} \left(\sum_{i=1}^{m} w_{hci}\, \hat{\Omega}_{hci} - n_h^{-1} \sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci} \hat{\Omega}_{hci}\right)^2, \tag{2.50}$$

where $\hat{\Omega}_{hci} = y_{hci} \hat{F}(y_{hci}) - \left(\frac{\hat{G}+1}{2}\right)(\hat{\mu} + y_{hci}) \;\; + \sum_{r=1}^{H} \sum_{s=1}^{n_r} \sum_{t=1}^{m} w_{rst}\, y_{rst} I\{y_{rst} \geq y_{hci}\}$.

Using the total in (2.38), Binder and Kovačević's (1995) estimator for the variance of $\hat{G}$ is given by

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

where $\bar{u}_h^* = n_h^{-1} \sum_{c=1}^{n_h} \sum_{i=1}^{m} u_{hc}^*$. Using $\frac{n_h}{n_h - 1} \to 1$ as $n_h \to \infty$, we write the asymptotic form of the formula as

$$\widehat{Var_{BK}}(\hat{G}) = \sum_{h=1}^{H}\sum_{c=1}^{n_h}(u_{hc}^{*} - \bar{u}_{h}^{*})^2 \ . \tag{2.51}$$

For an *iid* sample, a natural estimator for the variance of $\hat{G}$ from expression (2.45) is given by

$$\widehat{Var}(\hat{G}) = \frac{1}{M^2}\sum_{j=1}^{M}\left(u_j^{*} - \bar{u}^{*}\right)^2 = \frac{1}{M^2}\sum_{j=1}^{M} u_j^{*2}, \tag{2.52}$$

where $\bar{u}^{*} = \frac{1}{M}\sum_{j=1}^{M} u_j^{*} = 0.$[16] In terms of Davidson's (2009, p. 32) notation, this is equivalent to $\widehat{Var}(\hat{G}) = \frac{1}{(M\hat{\mu})^2}\sum_{j=1}^{M}\left(\hat{Z}_j - \bar{Z}\right)^2$, where $\hat{Z}_j = -(\hat{G}+1)y_j + 2\left(\hat{F}(y_j)y_j - \widehat{m}(y_j)\right)$ and $\bar{Z} = \hat{\mu}(\hat{G}-1)$. That is, as expected, Davidson's proposed variance estimator is a special case of that provided by Binder and Kovačević (1995).

On the other hand, Bhattacharya (2007), proposes the following estimator for the variance of $\hat{G}$:

$$\widehat{Var_B}(\hat{G}) = 4\left[\begin{array}{c}\displaystyle\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}^2\hat{\Phi}_{hci}^2 + \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m}\sum_{i'\neq i}^{m} w_{hci}w_{hci'}\ \hat{\Phi}_{hci}\,\hat{\Phi}_{hci'} \\ \displaystyle -\sum_{h=1}^{H}\frac{1}{n_h}\left(\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\,\hat{\Phi}_{hci}\right)^2\end{array}\right] \tag{2.53}$$

where $\hat{\Phi}_{hci} = \int_0^1 \tilde{\Phi}_{hci}(p)\mathrm{d}p$, and

$\tilde{\Phi}_{hci}(p) = \frac{1}{\hat{\mu}}\left[y_{hci}I\{y_{hci} \le \hat{z}(p)\} - \hat{\alpha}(p) + \hat{z}(p)(p - I\{y_{hci} \le \hat{z}(p)\}) - \frac{\hat{\alpha}(p)}{\hat{\mu}}(y_{hci} - \hat{\mu})\right]$,

where $\hat{\alpha}(p) = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}y_{hci}I\{y_{hci} \le \hat{z}(p)\}$, p is the sample percentile, $p = \hat{F}(\hat{z}(p))$ and $\hat{z}(p)$ is the sample quantile, $\hat{z}(p) = \hat{F}^{-1}(p)$, and, with a finite sample,

---

[16] This also holds for the complex survey; i.e., $\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}u_{hci}^{*} = 0$ (see the proof in Appendix B.4).

$\hat{z}(p) = \inf\{y_{hci} \in s: \hat{F}(y_{hci}) \geq p\}$. Also, $\frac{\hat{\alpha}(p)}{\hat{\mu}}$ is the estimator of the Lorenz curve, $L(p)$.

As given in (2.41), the estimator

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

and using that $-2 \sum_{i=}^{k} \hat{\Phi}_{hci} = u_{hc}^{*}$, and that approximations for $(\hat{G} - G)$ proposed by Binder and Kovačević and Bhattacharya are equivalent, we now show that the variance estimator proposed by Binder and Kovačević in (2.51) and Bhattacharya in (2.53) for $\hat{G}$ are equivalent.

To begin, as it illustrates an approach, we first show in the following subsection that the variance estimators proposed by Binder and Kovačević (1995) and Bhattacharya (2007) for the sample mean, via Theorem 1, are the same.

2.6.1 Unifying the approaches to estimate the variance of $\hat{G}$ proposed by Binder and Kovačević (1995) and Bhattacharya (2007)

As stated, we first show that the variance estimators proposed by Binder and Kovačević (1995) and Bhattacharya (2007) for the sample mean are the same.

Specifically, we have

Theorem 1. The variance estimator for the sample mean with a complex survey sample obtained by Binder and Kovačević (1995, p. 142) using an estimating equations method:

$$\widehat{Var(\hat{\mu})} = \sum_{h=1}^{H} \frac{n_h}{n_h - 1} \sum_{c=1}^{n_h} \left(u_{hc}^{*} - n_h^{-1} \sum_{c=1}^{n_h} u_{hc}^{*} \right)^2, \tag{2.54}$$

where $u_{hc}^{*} = \sum_{i=1}^{m} w_{hci}(y_{hci} - \hat{\mu})$, and, the variance estimator obtained by Bhattacharya (2005, p. 154), using a generalized methods of moments approach[17]

$$\widehat{Var(\hat{\mu})} =$$

$$\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}^{2}(y_{hci} - \hat{\mu})^2 + \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m}\sum_{j\neq i}^{m} w_{hci}^{2}\,(y_{hci} - \hat{\mu})(y_{hcj} - \hat{\mu})$$

$$-\sum_{h=1}^{H}\frac{1}{n_h}\left(\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}(y_{hci} - \hat{\mu})\right)^{2}, \qquad (2.55)$$

asymptotically equivalent, as $n_h \to \infty$, in the sense that the difference between the two statistics converges to zero for every draw from the sample space.

Proof: The EE for the population mean, μ, is given by

$$\int_{0}^{\infty} (y - \mu)\, dF(y) = 0,$$

and the corresponding complex survey sample estimating equation for the mean is

$$\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}(y_{hci} - \hat{\mu}) = 0.$$

Solving the sample EE, Binder and Kovačević (1995, p. 142) obtain the estimator $\hat{\mu} = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}y_{hci}$. Forming the total, $u_{hc}^{*} = \sum_{i=1}^{m} w_{hci}(y_{hci} - \hat{\mu})$, so that the sample EE reduces to $\sum_{h=1}^{H}\sum_{c=1}^{n_h} u_{hc}^{*} = 0$, they show that a variance estimator for $\hat{\mu}$ as the estimator of mean squared error of $\hat{\mu}$ is:

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

---

[17] In the last term of the following expression, Bhattacharya (p. 154) uses the expanded form of the weight function and calculates the cluster mean, $\hat{\mu}_{hc} = \frac{1}{k}\sum_{i=1}^{m} w_{hci}y_{hci}$, leaving other components in the term. We use the normalized weight and the actual observation values, y_{hci}.

where $\bar{u}_h^* = n_h^{-1} \sum_{c=1}^{n_h} u_{hc}^*$. As asymptotically, $\frac{n_h}{n_h - 1} \to 1$, we can write

$$\widehat{Var(\hat{\mu})} = \sum_{h=1}^{H} \sum_{c=1}^{n_h} (u_{hc}^* - \bar{u}_h^*)^2 \tag{2.56}$$

$$= \sum_{h=1}^{H} \sum_{c=1}^{n_h} [(u_{hc}^*)^2 - 2\, u_{hc}^*\, \bar{u}_h^* + (\bar{u}_h^*)^2]$$

$$= \sum_{h=1}^{H} \left[\sum_{c=1}^{n_h} (u_{hc}^*)^2 - 2n_h \left(n_h^{-1} \sum_{c=1}^{n_h} u_{hc}^* \right) \bar{u}_h^* + \sum_{c=1}^{n_h} (\bar{u}_h^*)^2 \right]$$

$$= \sum_{h=1}^{H} \left[\sum_{c=1}^{n_h} (u_{hc}^*)^2 - 2n_h (\bar{u}_h^*)^2 + n_h (\bar{u}_h^*)^2 \right]$$

$$= \sum_{h=1}^{H} \left[\sum_{c=1}^{n_h} (u_{hc}^*)^2 - n_h (\bar{u}_h^*)^2 \right]$$

$$= \sum_{h=1}^{H} \left[\sum_{c=1}^{n_h} \left(\sum_{i=1}^{m} w_{hci} (y_{hci} - \hat{\mu}) \right)^2 - n_h^{-1} \left(\sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci} (y_{hci} - \hat{\mu}) \right)^2 \right]. \tag{2.57}$$

The first term, $\sum_{c=1}^{n_h} (\sum_{i=1}^{m} w_{hci}(y_{hci} - \hat{\mu}))^2$, in (2.57) can be further decomposed as

$$\sum_{c=1}^{n_h} \big[\big(w_{hc1}(y_{hc1} - \hat{\mu}) + \cdots + w_{hcm}(y_{hcm} - \hat{\mu}) \big)$$

$$\times \big(w_{hc1}(y_{hc1} - \hat{\mu}) + \cdots + w_{hcm}(y_{hcm} - \hat{\mu}) \big) \big]$$

$$= \sum_{c=1}^{n_h} \left[\sum_{i=1}^{m} w_{hci}^2 (y_{hci} - \hat{\mu})^2 + \sum_{i=1}^{m} \sum_{j \neq i}^{m} w_{hci} w_{hcj} (y_{hci} - \hat{\mu}) \left(y_{hcj} - \hat{\mu} \right) \right]$$

$$= \left[\sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci}^2 (y_{hci} - \hat{\mu})^2 + \sum_{c=1}^{n_h} \sum_{i=1}^{m} \sum_{j \neq i}^{m} w_{hci} w_{hcj} (y_{hci} - \hat{\mu}) \left(y_{hcj} - \hat{\mu} \right) \right]. \tag{2.58}$$

The first term in (2.58) is the simple random sample variance with weights, and the second term is the variance component showing the cluster effect when households within a cluster are correlated. If households within a cluster are not correlated, the design based variance and the simple random sample variance with weights are the same (e.g., Deaton, 1997, p. 50-56; Cochran, 1977, p. 209-210).[18] Substituting equation (2.58) in expression (2.57), we obtain,

$$\widehat{Var(\hat{\mu})} = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}^2(y_{hci}-\hat{\mu})^2 + \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m}\sum_{j\neq i}^{m} w_{hci}w_{hcj}(y_{hci}-\hat{\mu})\left(y_{hcj}-\hat{\mu}\right) - \sum_{h=1}^{H} n_h^{-1}\left(\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}(y_{hci}-\hat{\mu})\right)^2, \qquad (2.59)$$

which is provided by Bhattacharya (2005, p. 154) to estimate the variance for the sample mean. As such we conclude that the difference between the two statistics converges to zero for every draw from the sample space. ■

Based on a similar approach to Theorem 1, it is straightforward to show that the asymptotic variance estimator for $\hat{G}$ given by expression (2.51) derived by Binder and Kovačević (1995), based on estimating equations is equivalent to the variance estimator in expression (2.53) derived by Bhattacharya (2007) from a generalized methods of moments approach.

Binder and Kovačević's (1995) expression (2.51) is equivalent in form to the expression (2.56) for the variance of the sample mean, with appropriate definitional changes. Likewise, Bhattacharya's expression (2.53) comprises the same components as his expression (2.58) for the variance of sample mean. Hence, the proof of the equivalence

---

[18] A sample estimate of the intracluster correlation coefficient, $\hat{\rho} \approx \frac{\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{k=m}\sum_{j\neq i}^{m} w_{hci}w_{hcj}(y_{hci}-\hat{\mu})(y_{hcj}-\hat{\mu})}{(m-1)\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}^2(y_{hci}-\hat{\mu})^2}$, is used to measure the household correlation within a cluster. When $\hat{\rho} = 0$, $\sum_{h=1}^{H}\sum_{c=1}^{n_h}(\sum_{i=1}^{m} w_{hci}(y_{hci}-\hat{\mu}))^2 = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}^2(y_{hci}-\hat{\mu})^2$.

of expressions (2.51) and (2.53) follows the same approach as for the proof of the two expressions for the variance of the sample mean. All that is essentially required is to show that u_{hc}^* in Binder and Kovačević's (1995) expression in (2.51) is equal to $-2\sum_{i=1}^{m} w_{hci}\widehat{\Phi}_{hci}$ in Bhattacharya's (2007) expression in (2.53), which we have shown in equation (2.42).[19]

Therefore, the estimators for the variance of $\hat{G}$ in expression (2.51),

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

and in expression (2.53),

$$\widehat{Var_B}(\hat{G}) = 4\left[\begin{array}{c}\displaystyle\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}^2\widehat{\Phi}_{hci}^2 + \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m}\sum_{i'\neq i}^{m} w_{hci}w_{hci'}\ \widehat{\Phi}_{hci}\ \widehat{\Phi}_{hci'} \\ \displaystyle -\sum_{h=1}^{H}\frac{1}{n_h}\left(\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\ \widehat{\Phi}_{hci}\right)^2\end{array}\right],$$

are asymptotically equivalent in the sense that the difference between the two estimators converges to zero for every draw from the sample space. ■

Showing the decomposition of variance formula for an estimator in terms of its components, as provided by Bhattacharya (2007), is extremely helpful in understanding the impacts of sampling features on the variance estimator. As Bhattacharya (2005) explains, the first term in (2.53) is an estimate of the variance without taking the appropriate sampling design into account, e.g., estimation with an *iid* sample with weights. The second term is the cluster effect, which accommodates the intracluster correlation. As it is usual for sampled units within a cluster to be correlated, hence, it increases the magnitude of the

---

[19] A more detailed proof of this equivalence is provided in Appendix B.3.

variance estimate. The third term is the stratum effect. As strata are independent, but not identical, this feature reduces the magnitude of the variance estimate.

Despite the benefits of writing the variance as given in expression (2.53), it is not especially friendly for practitioners to code, whereas the form of expression (2.51) (or (2.50)) is simpler to practically calculate. If a researcher is using a software package designed for surveys,[20] then it is easy to calculate Binder and Kovačević's (1995) asymptotic variance estimator

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

With such software packages, the variance (often called the linearization variance) is easily generated using an appropriate "total" command after forming the series $\sum_{i=1}^{m} w_{hci}\, u_{hci}^{*} = u_{hc}^{*}$, 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):

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

We also show in the section below that using some auxiliary regressions, we can estimate G and its variance, to yield equivalent results. A major advantage of using such a regression trick is that one can estimate a coefficient and its variance with a complex survey using software packages that do not take the survey design into account.

---

[20] Examples include Stata, SPSS, SAS, SUDAAN and the "survey" package developed by Lumley for R (see Lumley, 2010).

2.7 Obtaining estimates using auxiliary regressions

To obtain estimates of the variance for $\widehat{G}$, a researcher could directly code expressions (2.50), (2.51) or (2.53). However, it turns out that an auxiliary regression technique can readily be used to obtain such estimates with *iid* and non-*iid* samples. As discussed earlier, Davidson (2009), Giles (2004) and Ogwang (2000) show that the regression technique can be used to obtain the sample G. We first extend some of these regression approaches to estimate G and its variance with an *iid* sample, and we then show that it is also straightforward with complex survey data. It is important to note that we are using the regression method to obtain estimates that are equivalent to the estimators obtained in the theory above. We are not concerned with any other feature or property of a regression method.

2.7.1 Computing $\widehat{G}$

When data are from an *iid* random sample, it has been shown that $\widehat{G}$ can easily be obtained from an artificial OLS regression: Davidson (2009), Giles (2004) and Ogwang (2000). As discussed by these authors, with unordered data and denoting v as an error term, we estimate the artificial regression:

$$\left(2\widehat{F}(y_d) - 1\right)\sqrt{y_d} = \theta\sqrt{y_d} + v_d, \quad d = 1, \dots, M \tag{2.60}$$

by OLS to yield $\widehat{\theta} = \widehat{G}$, where for a sample *iid* sequence $\{y_d\}_{d=1}^{M}$, $\widehat{F}(y_d) = \frac{d}{M}$. If the data are ordered with $\widehat{F}$ computed using the average of the lower and upper limits, as advocated by Davidson (2009), the artificial regression:

$$\left(\frac{2d}{M} - \frac{1}{M} - 1\right)\sqrt{y_{(d)}} = \theta\sqrt{y_{(d)}} + v_d, \quad d = 1, \dots, M \tag{2.61}$$

estimated by OLS results in $\widehat{\theta} = \widehat{G}$ as defined by expression (2.21).

A similar approach can be adopted with a complex survey. In the following auxiliary regressions with complex survey data, the sampling weights are assumed to be normalized such that $\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci} = 1$. Should a researcher be using a software package not explicitly able to handle the sampling design, with observations, labeled from 1 through M, the OLS estimator of θ in the artificial regression

$$\left(2\hat{F}(y_d) - 1\right)\sqrt{w_d y_d} = \theta\sqrt{w_d y_d} + v_d, \qquad d = 1, \ldots, M \tag{2.62}$$

leads to $\hat{G}$ given in expression (2.26); 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, estimation of the regression

$$\left(2\hat{F}(y_{hci}) - 1\right)\sqrt{y_{hci}} = \theta\sqrt{y_{hci}} + v_{hci}, \tag{2.63}$$

by OLS, having declared appropriate elements of the sampling design, yields $\hat{\theta} = \hat{G}$, where $\{y_{hci}: h = 1, \ldots, H; c = 1, .., n_h; i = 1, \ldots, m\}$.

To obtain $\widehat{Var}(\hat{G})$, we can either simply calculate the appropriate estimator or we can also estimate additional auxiliary regressions. We consider this in the next subsection.

2.7.2 Computing $\widehat{Var}(\hat{G})$

We now consider such an auxiliary regression equation to estimate $Var(\hat{G})$ with a complex survey when computational software does not account for survey designs. In such a situation, one way to obtain Bhattacharya's variance estimator is to estimate the artificial regression:

$$u_{hc}^{*} = \sum_{h=1}^{H} \beta_h D_{hc} + v_{hc}, \tag{2.64}$$

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 $u_{hc}^* = \sum_{i=1}^{m} w_{hci} u_{hci}^*$, which only requires a few lines of code. 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:

$$\left(\frac{n_h}{n_h - 1}\right)^{0.5} u_{hc}^* = \sum_{h=1}^{H} \beta_h D_{hc}^* + v_{hc}\,, \qquad (2.65)$$

where $D_{hc}^* = \sqrt{(n_h/(n_h - 1))}$ if cluster c is in stratum h, 0 otherwise.

At first glance, it may seem counterintuitive that a SSR can be a variance estimator that goes to zero as $n_h \to \infty, \forall\, n_h$.[21] However, each u_{hci}^* includes a weight $w_{hci} = \frac{W_{hi}}{\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} W_{hci}} = \frac{W_{hci}}{M} \to 0$ as $N \to \infty$, so that $SSR \to 0$ under the asymptotic analysis, i.e., $\widehat{Var(\hat{G})}_{N\to\infty} \to 0$. In case of an *iid* sample, to obtain the variance estimator for $\hat{G}$ provided by Davidson (2009, p. 32), $\widehat{Var(\hat{G})} = \frac{1}{n^2\hat{\mu}^2}\sum_{i=1}^{n}\left(\hat{Z}_i - \bar{Z}\right)^2$, we consider the following auxiliary regression

$$\frac{\hat{Z}_i}{\hat{\mu}} = \delta + v_i\ , \qquad (2.66)$$

where v_i is an error term. As the OLS estimator of δ is $\hat{\delta} = \bar{Z}/\hat{\mu}$, the variance of $\hat{\delta}$ is $Var(\hat{\delta}) = n^{-1}\sigma^2$ and using that a consistent estimator for σ^2 is $\hat{\sigma}^2 = \frac{\hat{v}'\hat{v}}{n}$, we have

$$\widehat{Var(\hat{\delta})} = \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{Z}_i}{\hat{\mu}} - \frac{\bar{Z}}{\hat{\mu}}\right)^2\right) = \frac{1}{n^2\hat{\mu}^2}\left(\sum_{i=1}^{n}\left(\hat{Z}_i - \bar{Z}\right)^2\right) = \widehat{Var(\hat{G})}.$$

We could equivalently estimate the auxiliary regression

$$\frac{\hat{Z}_i}{n\hat{\mu}} = \delta^* + v_i^*, \qquad (2.67)$$

---

[21] I am grateful to James MacKinnon, Queen's University, Ontario, for raising this possible initial reaction to the SSR estimator.

by OLS which would result in $\widehat{Var(\hat{G})} = SSR$ from this auxiliary regression, analogous to the estimator for the complex survey case.

2.8 Concluding Remarks

In this chapter, we examine and show that variance estimators for $\hat{G}$ using complex survey samples proposed by Bhattacharya (2007) and Binder and Kovačević (1995) are equivalent. A key advantage of this equivalence result is that the variance formula provided by Binder and Kovačević, along with the approximation for $\hat{G}$, is 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. This is indeed not the case; asymptotic variances can be readily calculated, even for researchers without access to specialized complex survey software, so providing an alternative to resampling methods such as the bootstrap. As an *iid* sample can be regarded as a special case of a complex survey sample, we also link recent work of Davidson (2009) to the earlier research undertaken by Binder and Kovačević, showing that Davidson's derived approximation for $\hat{G}$ and his proposed variance estimator also follow directly from Binder and Kovačević's work.

We provide applied researchers with an easily implementable way, i.e., an auxiliary regression technique, to calculate both a Gini coefficient estimator and an estimator of its associated variance. This straightforward technique dismisses the saga that asymptotic variances for Gini indices, especially with complex survey data, are computationally burdensome to calculate. We now turn to using Monte Carlo experiments to explore the finite sample properties of these asymptotic estimators, along with comparing these properties to some resampling approaches.

CHAPTER THREE: FINITE SAMPLE PROPERTIES OF GINI COEFFICIENT ESTIMATORS USING MONTE CARLO EXPERIMENTS

3.1 Introduction

In Chapter 2, we considered using asymptotic approximations to obtain variances for a plug-in estimator of G obtained from a complex survey sample. To recap, we studied the links between the research of Binder and Kovačević (1995), Bhattacharya (2007), and Davidson (2009). Assuming a complex survey design, Bhattacharya (2007) provides an asymptotic variance estimator for the plug-in estimator of G, based on a generalized method of moments framework, which, we showed, is equivalent to an asymptotic version of the variance estimator provided by Binder and Kovačević (1995) obtained from an estimating equations principle. Davidson (2009) also provides an asymptotic variance for $\hat{G}$, based on an *iid* sample; we showed his variance estimator as a special case of Binder and Kovačević's (1995) variance estimator. To assist applied researchers, we detailed different ways to calculate the variances for both *iid* and complex survey samples. In this chapter, we use Monte Carlo (MC) simulation methods to explore the finite sample properties of the Gini coefficient estimator, along with various proposed variance estimators.

The finite sample properties of the Gini coefficient estimator $(\hat{G})$ are of interest, though their finite sample features are not analytically tractable, and in real data sets the strata in surveys often contain a small number of clusters. In this Chapter, we take up this topic by exploring such issues using MC experiments, a commonly adopted methodology in the face of intractable finite sample properties. We examine: (i) biases associated with $\hat{G}$; and (ii) performance of interval estimators for G through empirical coverage probabilities (ECPs). To assess the sensitivity of the results from the assumed data generating process (DGP), we consider a total of 36 DGPs under four probability distribution functions (beta, lognormal, chi-square, and the Pareto distributions).

The bias of an estimator is the deviation of the average of all possible estimates from the true parameter value. An estimator is unbiased when the expected value of the estimator is equal to the parameter value. Unbiasedness is one of the desirable statistical properties of an estimator, and a large number of studies examine the bias of $\hat{G}$ under both iid and non-iid frameworks. For example, Sudheesh and Dewan (2013), Antal et al. (2011), Davidson (2009), Moran (2005b), Deltas (2003), and Dixon et al. (1987) use iid samples, and Luus et al. (2012), Van Ourti and Clarke (2011), Moskowitz et al. (2008), and Lerman and Yitzhaki (1989) use non-iid samples to evaluate the bias of $\hat{G}$. This study contributes to the broadening of the literature in this area by exploring the bias of $\hat{G}$ using simulation studies when sample data are obtained from a wide range of DGPs under the complex survey design.

The confidence interval (CI) or the interval estimator is a statistical procedure commonly used to make inferences about the population and its parameters based on the sample data. Several studies have used ECPs to evaluate performance of an interval estimator for G using MC experiments with iid samples including Giacomini et al. (2013), Sudheesh and Dewan (2013), Peng (2011), Qin et al. (2010), Davidson (2009), Berger (2008), Gastwirth et al. (2005), and Dixon et al. (1987). Only a small number of studies focus on ECP of CIs for G with sample data obtained from a complex survey design; see, for example, Qin et al. (2010), Berger (2008), Wu and Rao (2006), and Sandström et al. (1988). Our work extends this body of research.

We explore the performance of four interval estimators for G using an extensive range of MC experiments under the complex survey framework. Three of these interval estimators use standard normal (SN) approximations based on asymptotic theory and one is based on the bootstrap MC percentile method. Of the three SN approximation interval estimators, two use the analytical variance estimators examined in Chapter 2, and the other uses a standard bootstrap variance estimator. In addition, to assess the performance of the bootstrap MC percentile interval estimator, we consider a MC percentile interval estimator formed using the warp-speed method proposed by Giacomini et al. (2013).

The asymptotic variance formulae discussed in Chapter 2 for $\hat{G}$ with complex survey design data are computationally less complicated and timesaving. An alternative approach to using asymptotic approximations, however, is to adopt bootstrap resampling methods, originally proposed by Efron (1979), which is commonly embraced by applied researchers to obtain variances. The bootstrap methods may have some optimal properties, but do not require knowledge of the distribution of the statistic of interest (e.g., Dixon et al., 1987). For instance, variances generated from resampling methods may perform better, with samples containing small numbers of clusters in strata, than variances obtained from asymptotic approximations that assume that the number of clusters in strata are large. Given this possibility, our MC experiments explore ECP of the CIs using the standard bootstrap variance, along with interval estimators constructed using asymptotic variance estimators. It would be a useful finding if the ECPs of these three CIs have similar magnitudes, as the asymptotic approximation interval estimators are far easier to calculate than the bootstrap interval estimators.

In any case, exploring finite sample properties of interval estimators obtained using bootstrap methods, for example, interval estimator using standard bootstrap variance or standard bootstrap MC percentile, can be computationally expensive. To assist researchers, Giacomini et al. (2013) propose the warp-speed MC method to estimate the ECP of a bootstrap statistic, which they suggest is close to the coverage using the standard double bootstrap method. To explore whether or not this is the case here, we consider Giacomini et al.'s (2013) warp-speed method to estimate the ECPs of CIs for G when data are obtained from complex surveys. We elaborate further on this approach in subsection 3.2.2.

The key goal of obtaining ECPs of CIs for G using four methods is to ascertain whether any strong recommendations can be made for applied researchers. Typically, methods with high ECPs of interval estimators, across a wide range of DGPs, are recommended to practitioners. Nevertheless, when considering the cost of computation, methods that have reasonably good coverages but are computationally less expensive may be preferred. To this end, our MC experiments consider a wide range of DGPs, aimed to represent cases found in practice. Although not exhaustive, we believe enough cases have been considered for us to make the following claims:

1. The number of strata in the DGP and the underlying distribution of the well-being variable are the two most important population features affecting the bias of $\hat{G}$ and the performance of interval estimators for G. The bias is higher and the ECP of a SN approximation interval estimator is lower for samples obtained from DGPs created with more strata and skewed distributions. For example, the bias of $\hat{G}$ is -0.12% and 13.81% of G and the ECP of the interval estimator for G using Bhattacharya's variance estimator is 93.1% (cf. a nominal 95%) and 53% for a sample with 90 clusters (10% of the population clusters from each stratum) drawn from DGPs containing two strata created under the lognormal and the Pareto distributions, respectively. The Pareto distribution we consider is highly skewed compared to other distributions such as beta and lognormal distributions.

2. The bias of $\hat{G}$ decreases with an increasing value of the intracluster correlation coefficient (ICC) for up to 0.5; after that it deteriorates and the sign of the bias becomes negative. Although the sampling variance of $\hat{G}$ increases with the ICC parameter, the ECP of a SN approximation interval estimator does not always improve for larger values of ICC.
3. Samples with less than 5% of the population clusters produce $\hat{G}$ that are highly biased, and interval estimators that do not perform well. Nevertheless, the performance of an interval estimator for G does not constantly improve in samples with more clusters. For example, ECPs of the SN approximation interval estimators fall with increasing number of clusters in samples drawn from DGPs with five strata generated under the beta and chi-square distributions.
4. The SN approximation CIs using the bootstrap standard errors are comparable to the two SN interval estimators using asymptotic variance estimators, except for some samples with fewer clusters, where the bootstrap method performs better.
5. The bootstrap MC percentile interval estimators usually provide better coverage than the three SN approximation interval estimators for samples with fewer clusters and/or when samples are drawn from DGPs generated under the heavy-tail Pareto distribution. In contrast, for samples with more clusters obtained from the DGP with fewer strata (e.g., three strata) under the Pareto distribution, the SN approximation interval estimators provide much higher empirical coverage probabilities than the bootstrap MC percentile interval estimator.
6. The ECP of the warp-speed MC method is around, often exceeding, the 95% nominal level, regardless of the finite population features and sample size that we examined, such as the ICC value, distribution of data, number of strata, and number of clusters in the sample. The coverage of the method is more often considerably

higher than the bootstrap MC percentile method that it is supposed to approximate. In some cases, however, especially for samples with a large number of clusters, ECPs of the bootstrap MC percentile and warp-speed MC percentile intervals are comparable. Overall, our results suggest that the warp-speed method cannot assess the performance of the bootstrap MC percentile CIs, at least under the complex survey framework that we consider.

Our findings have practical implications for applied researchers in undertaking inferences about the Gini coefficient using complex survey samples. For real life survey data with low ICC values, $\hat{G}$ can be biased. Nevertheless, if the well-being variable does not arise from a highly skewed underlying distribution or the sample contains a high proportion of the population's clusters, the bias of $\hat{G}$ may be negligible. In fact, this is the case with most of the Demographic Health Surveys (DHS). For example, BMI among ever-married women of age 15-49 years from Bangladesh DHS 2011 is comprised of 600 clusters (chosen from 296,718 enumeration areas) from 20 strata, where the ICCs across strata range from 0.02 to 0.18. On the other hand, to obtain a reliable CI for G under survey sampling designs, researchers may adopt one of the asymptotic variances we examined to avoid burdensome computations involved with bootstrap resampling techniques. Even though the CIs using bootstrap methods perform slightly better in small cluster samples than the asymptotic variance based CIs, when computational costs and time are accounted for, the latter methods may be preferred over the former methods.

The remainder of this chapter is organized as follows. Section 3.2 provides details of how we approach estimating bias of $\hat{G}$ and ECPs of our nominal 95% interval estimators. In Section 3.3, we detail the MC simulation experiment designs, while Section 3.4 reports on the simulation results. We conclude in Section 3.5 with final remarks and recommendations from our MC experiments for applied researchers.

3.2 Estimating Empirical Bias and Coverage Probabilities of Interval Estimators

To recap, the plug-in estimator for G for a well-being variable y presented in Chapter 2 from a stratified two-stage cluster sampling design is

$$\hat{G} = \frac{2}{\hat{\mu}} \sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci}\, y_{hci} \hat{F}(y_{hci}) - 1, \tag{3.1}$$

with two asymptotically equivalent plug-in variance expressions provided by Bhattacharya (2007, 2005) and Binder and Kovačević (1995) being

$$\widetilde{Var_{BT}}(\hat{G}) = \sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci}^2\, \hat{\Psi}_{hci}^2 + \sum_{h=1}^{H} \sum_{c=1}^{n_h} \left(\sum_{i=1}^{m} \sum_{j \neq i}^{m} w_{hci}\, w_{hcj}\, \hat{\Psi}_{hci} \hat{\Psi}_{hcj} \right) - \sum_{h=1}^{H} \frac{1}{n_h} \left(\sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci}\, \hat{\Psi}_{hci} \right)^2, \tag{3.2}$$

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

$$\widehat{Var_{BK}}(\hat{G}) = \sum_{h=1}^{H} \sum_{c=1}^{n_h} (u_{hc}^* - \bar{u}_h^*)^2, \tag{3.3}$$

where $u_{hc}^* = \sum_{i=1}^{m} w_{hci} u_{hci}^*$, $\bar{u}_h^* = (1/n_h) \sum_{c=1}^{n_h} u_{hc}^*$, $u_{hci}^* = \frac{2}{\hat{\mu}} \left[y_{hci} \left(\hat{F}(y_{hci}) - \frac{(\hat{G}+1)}{2} \right) + B(y_{hci}) - \frac{\hat{\mu}}{2} (\hat{G} + 1) \right]$ and $B(y_{hci}) = \sum_{r=1}^{H} \sum_{s=1}^{n_r} \sum_{t=1}^{m} w_{rst} y_{rst} I(y_{rst} \ge y_{hci})$, respectively.

Let G be the true Gini coefficient calculated from a finite population $\{\mathcal{U}[hci]: h = 1, \dots, H; c = 1, \dots, N_h; i = 1, \dots, M_{hc}\}$, denoted by DGP. We consider the MC method in the context of sampling from the DGP following the complex survey design to estimate G, using $\hat{G}$, and an estimator for the variance of $\hat{G}$. A stratified two-stage cluster sample $S \subset \mathcal{U}$ of size M households (ultimate units) is drawn with the sample being $\{S[hci]: h =$

$1, \dots, H; c = 1, \dots, n_h; i = 1, \dots, m\}$; subscripts h, c and i refer to stratum, cluster within a stratum, and household within a cluster, respectively. First, from each stratum h, n_h clusters are drawn via probability proportional to size with the total number of sampled clusters being $N = \sum_{h=1}^{H} n_h$. In the second stage, from cluster c, within stratum h, we suppose that a fixed m households are selected, using simple random sampling. We have: $m < M_{hc}$, $n_h < N_h$ and the total number of sampled households is $m \sum_{h=1}^{H} n_h = mN = M$.

Under the described sampling framework, the weight assigned to each household is $W_{hci} = \frac{M_{hc} N_h}{n_h m}$, which is normalized so that $\sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} W_{hci} = M$. In addition, weights are standardized as $w_{hci} = \frac{W_{hci}}{\sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} W_{hci}}$, such that $\sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci} = 1$. Given the sampling design, let $\hat{G}_k$ be the estimated G from the MC sample k, $k = 1, \dots, K$, where K denotes the number of MC samples.

3.2.1 Empirical Bias of $\hat{G}$

The bias of an estimator is the deviation of its expected value of the sampling distribution of the estimator from the corresponding parameter value. So, the bias of $\hat{G}$ is $\text{bias}(\hat{G}) = E(\hat{G}) - G$. We can estimate this bias from a simulation experiment as:

$$\widehat{\text{bias}}(\hat{G}) = \widehat{E(\hat{G})} - G. \tag{3.4}$$

where $\widehat{E(\hat{G})} = \frac{1}{K} \sum_{k=1}^{K} \hat{G}_k$ is a MC estimator of $E(\hat{G})$ and $\hat{G}_k$ is the kth MC sample estimator of G. The estimator $\hat{G}$ is said to be unbiased when $\text{bias}(\hat{G}) = 0$, with positive bias corresponding to $\text{bias}(\hat{G}) > 0$, and the bias is downward or negative when $\text{bias}(\hat{G}) < 0$. We use the empirical MC bias, $\widehat{\text{bias}}(\hat{G})$, as our estimator of $\text{bias}(\hat{G})$.

An unbiased estimator is often preferred to a biased estimator, ceteris paribus. However, when other properties are considered, for example, variance of the estimator, biased estimators may lead to better performance, for instance, smaller mean squared error. We estimate the bias of $\hat{G}$ as a percentage of G from complex survey samples using MC simulations to examine factors that affect the bias. Our experiments on bias estimation can help applied researchers to choose an appropriate sampling design and size of the sample.

3.2.2 Empirical Coverage Probability for 95% Confidence Interval for G

In statistical inference, the accuracy and performance of a CI is typically assessed through its coverage probability. A CI estimator is reliable to undertake inference about the population parameter, when the estimated CP of the estimator is close to the nominally specified confidence level. The ECP is estimated as the fraction of the replications for which the true value of a parameter lies within a pre-set confidence level. For example, an ECP of 0.90 for a 95% nominal CI means that the interval includes the true parameter value 90% of the times in the sample data. When the ECP is lower than the nominal confidence level, but increases in repeated sampling with larger sample sizes, the accuracy of the interval estimator is said to be increasing.

Specifically, the CI for G is such that the coverage probability $CP(G) = P(G \in CI) \approx 1 - \alpha$, where $1 - \alpha$ is the nominal level of confidence. We provide ECPs of three SN approximations, bootstrap MC percentile, and warp-speed MC percentile CIs for G from $K = $1,000 MC samples, for a commonly adopted 95% confidence level ($\alpha = $0.05). Construction of each CI and the corresponding ECP are given below.

1. ***SN approximation confidence interval (CI) using $\widetilde{Var}_{BT}(\hat{G})$ (BTAM)***

   The steps involved in constructing ECPs for this interval estimator are:

   a. Calculate Bhattacharya's (2007) variance estimator using the expression in (3.2) for each MC replication k; we denote this variance estimator as $\widetilde{Var}_{BT}(\hat{G}_k)$.

   b. Form the equal-tail 95% normal approximation CI for G using $\widetilde{Var}_{BT}(\hat{G}_k)$ for each k, $k = 1, \ldots, K$, as

   $$\widehat{CI}_{BT,k}(0.95) = \left[\hat{G}_k - z_{0.975}\sqrt{\widetilde{Var}_{BT}(\hat{G}_k)}\,, \hat{G}_k - z_{0.025}\sqrt{\widetilde{Var}_{BT}(\hat{G}_k)}\right], \tag{3.5}$$

   where $z_{0.025}$ and $z_{0.975}$ are the empirical values from the standard normal distribution for 0.025 and 0.975 probabilities, respectively. The ECP is the proportion of the intervals that covers the true G, over the K replications. Let $A_{BT,k}$ be an indicator function that is 1 when $G \in \widehat{CI}_{BT,k}(0.95)$, 0 otherwise.

   c. Thus, the BTAM ECP is $\frac{1}{K}\sum_{k=1}^{K} A_{BT,k}$.

2. ***SN approximation CI using $\widehat{Var}_{BK}(\hat{G})$ (BKAM)***

   a. Calculate Binder and Kovačević's (1995) variance estimator using the expression in (3.3) for each k, $\widehat{Var}_{BK}(\hat{G}_k)$.

   b. For each k, form the equal-tailed 95% normal approximation CI for G using $\widehat{Var}_{BK}(\hat{G}_k)$ as:

   $$\widehat{CI}_{BK,k}(0.95) = \left[\hat{G}_k - z_{0.975}\sqrt{\widehat{Var}_{BK}(\hat{G}_k)}\,, \hat{G}_k - z_{0.025}\sqrt{\widehat{Var}_{BK}(\hat{G}_k)}\right]. \tag{3.6}$$

   c. Let $A_{BK,k}$ be an indicator function that is 1 when $G \in \widehat{CI}_{BK,k}(0.95)$, 0 otherwise.

   d. The BKAM ECP is $\frac{1}{K}\sum_{k=1}^{K} A_{BK,k}$.

3. ***SN approximation CI using standard bootstrap standard error (SBSM)***

Forming CIs using bootstrap standard errors is a common practice in applied research when deriving analytical (finite sample or asymptotic) variances is deemed cumbersome. In addition, using such bootstrap standard errors may be preferred in finite samples to asymptotic standard errors. The standard bootstrap method involves subsampling from the sample and using the subsamples to estimate the variance of $\hat{G}$. An extensive elaboration of different bootstrap methods for constructing reliable interval estimates with *iid* samples is given in, for example, DiCiccio and Efron (1996). For the complex survey design, we construct bootstrap CIs for G following the bootstrap technique detailed in Wolter (2007, p. 211-12)

a. From each MC sample k, draw B bootstrap samples of size $n_h^* = n_h - 1$ from each stratum h, $h = 1, \dots, H$,[22] with replacement, and independently estimate the Gini coefficient, $\hat{G}_{b,k}^*, b = 1, \dots, B$.
b. For each MC replication, calculate the mean of $\hat{G}_{b,k}^*$, $\bar{\hat{G}}_k^* = \frac{1}{B}\sum_{b=1}^{B} \hat{G}_{b,k}^*$.
c. Calculate the variance estimator for the Gini coefficient estimator in each kth MC sample:

$$Var_{BOOT,k}\left(\hat{G}_b^*\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{G}_{b,k}^* - \bar{\hat{G}}_k^*\right)^2 \tag{3.7}$$

d. For each k, over $k = 1, \dots, K$, the equal-tailed 95% normal approximation CI for G using the bootstrap standard error is

$$\widehat{CI}_{BOOT,k}(0.95) = \left[\hat{G}_k - z_{0.975}\sqrt{Var_{BOOT,k}\left(\hat{G}_b^*\right)}, \hat{G}_k - z_{0.025}\sqrt{Var_{BOOT,k}\left(\hat{G}_b^*\right)}\right]. \tag{3.8}$$

e. Let $A_{BOOT,k}$ be an indicator function that is 1 when $G \in \widehat{CI}_{BOOT,k}(0.95)$, 0 otherwise.
f. The empirical coverage probability of SBSM is then formed by $\frac{1}{K}\sum_{k=1}^{K} A_{BOOT,k}$.

---

[22] When $n_h^* = n_h - 1$, the bootstrap variance estimator of an estimator with a complex survey sample is unbiased; otherwise, the variance estimator could be substantially biased for $n_h^* = n_h$, especially when n_h is small (see, e.g., Wolter, 2007, p. 211-212).

4. *Standard bootstrap MC percentile CI (SBPM)*

Now, we form the CI for G and calculate empirical coverage considering the bootstrap MC percentile method akin to the "standard" bootstrap MC percentile outlined in Giacomini et al. (2013) following the steps below:

a. Consider the B bootstrap Gini estimates referred to from the previous bootstrap resampling, for each k, $\hat{G}^*_{b,k}, b = 1, \ldots, B$.
b. Form $\Delta_{b,k} = \hat{G}^*_{b,k} - \hat{G}_k, b = 1, \ldots, B$, for each k. Calculate $\hat{q}_{B,k}(\alpha)$, where $\hat{q}_{B,k}(\alpha)$ is the α-quantile of the empirical distribution of $\Delta_{b,k}, b = 1, \ldots, B$.
c. For each k, the equal-tailed 95% CI for G is:

$$\widehat{CI}_{BP,k}(0.95) = \left[\hat{G}_k - \hat{q}_{B,k}(0.975), \hat{G}_k - \hat{q}_{B,k}(0.025)\right] \tag{3.9}$$

d. Let $A_{BP,k}$ be an indicator function that is 1 when $G \in \widehat{CI}_{BP,k}(0.95)$, 0 otherwise.
e. The ECP is then formed by $\frac{1}{K}\sum_{k=1}^{K} A_{BP,k}$.

5. *Warp-speed MC percentile CI method (WSPM)*

Assessing the performance of bootstrap estimators and CIs using the double-bootstrap method can be computationally expensive, especially when the statistic of interest is nonlinear and complex (see Giacomini et al., 2013, and additional citations therein). Suggesting that taking just one bootstrap sample from each MC replication is often sufficient to provide a useful approximation to the statistic of interest, Giacomini et al. (2013) propose a warp-speed method that involves bootstrap estimators. The main motivation for this method is to reduce the computational burden of assessing the MC performance of bootstrap methods by speeding up the MC experiment.

We estimate CP using this warp-speed MC method for comparison with the results we obtain from the other bootstrap approaches. Although the CPs of the warp-speed method are not practically useful to practitioners, we include this in our work to indicate whether or not such a MC approach does indeed provide a good approximation to the bootstrap methods we consider. If the warp-speed coverage probabilities are similar to those obtained under the standard bootstrap MC percentile method (SBPM), then this further supports (in a completely different framework) the results of Giacomini et al. (2013). On the other hand, the presence of substantial differences between the two ECPs suggests that the warp-speed MC method is not a good indicator of the accuracy of the bootstrap CI formed under the SBPM method. For our complex survey framework, we adopt the following steps:

a. Let G be the population Gini coefficient for a DGP, and $\widehat{G}_k$ be the estimated Gini coefficient from the MC samples, $k = 1, \ldots, K$, using the sampling technique discussed previously.
b. Draw a *single* bootstrap resample from each MC replication and obtain the Gini coefficient estimator. We consider the first bootstrap resample and let $\widetilde{G}_{b=1,k}$ be the estimated Gini coefficient from the bootstrap resamples, $k = 1, \ldots, K$.
c. Form the difference series over K, $\Delta_k = \widetilde{G}_{b=1,k} - \widehat{G}_k$, for $k = 1, \ldots, K$.
d. Calculate $\hat{q}_K(\alpha)$, where $\hat{q}_K(\alpha)$ is the α −quantile of the empirical distribution of $\Delta_k, k = 1, \ldots, K$.
e. Construct K equal-tailed 95% CIs:

$$\widehat{CI}_{WP,k}(0.95) = \left[\widehat{G}_k - \hat{q}_K(0.975), \widehat{G}_k - \hat{q}_K(0.025)\right] \tag{3.10}$$

f. Let $A_{WP,k}$ be an indicator function that is 1 when $G \in \widehat{CI}_{WP,k}(0.95)$, 0 otherwise.
g. The ECP using the WSPM method is then formed by $\frac{1}{K}\sum_{k=1}^{K} A_{WP,k}$.

To evaluate the performance of an interval estimation method, the empirical coverage of tail probabilities are sometimes also reported in simulation studies; see, for example, Arasan and Adam (2014), Antal and Tillé (2014), Qin et al. (2010). These are useful because the lower and upper tail probabilities give us an idea of how skewed is the empirical distribution of a statistic. We report tail empirical coverage probabilities (rates), in addition to ECPs for the five interval estimators for G to assess their performance.

The empirical lower (upper) tail probability is calculated as the number of times the lower (upper) endpoint of the 95% CI for G is more (less) than the true parameter value divided by the total number of samples. For the SN approximation methods, the lower and upper tail CPs are calculated as:

Lower tail coverage probability (L) in %:

$$L = \frac{100}{K}\sum_{k=1}^{K} I\left[\left(\hat{G}_k - z_{0.975}\sqrt{v_k(\hat{G})}\right) > G\right], \text{ and}$$

upper tail coverage probability (U) in %: (3.11)

$$U = \frac{100}{K}\sum_{k=1}^{K} I\left[\left(\hat{G}_k - z_{0.025}\sqrt{v_k(\hat{G})}\right) < G\right],$$

where K is the number of MC replications, and $I[\,.\,]$=1 if the argument is true and 0 otherwise; $v_k(\hat{G})$ is estimated as the variance for $\hat{G}$ using Bhattacharya (2007), Binder and Kovačević (1995) and bootstrap as appropriate methods for $k = 1, \dots, K$.

For the bootstrap MC percentile method, lower and upper tail coverage probabilities are calculated as:

Lower tail coverage probability (L) in %:

$$L = \frac{100}{K}\sum_{k=1}^{K} I\left[\left(\hat{G}_k - \hat{q}_k(0.975)\right) > G\right], \text{ and}$$

upper tail coverage probability (U) in %: (3.12)

$$U = \frac{100}{K}\sum_{k=1}^{K} I\left[\left(\hat{G}_k - \hat{q}_k(0.025)\right) < G\right].$$

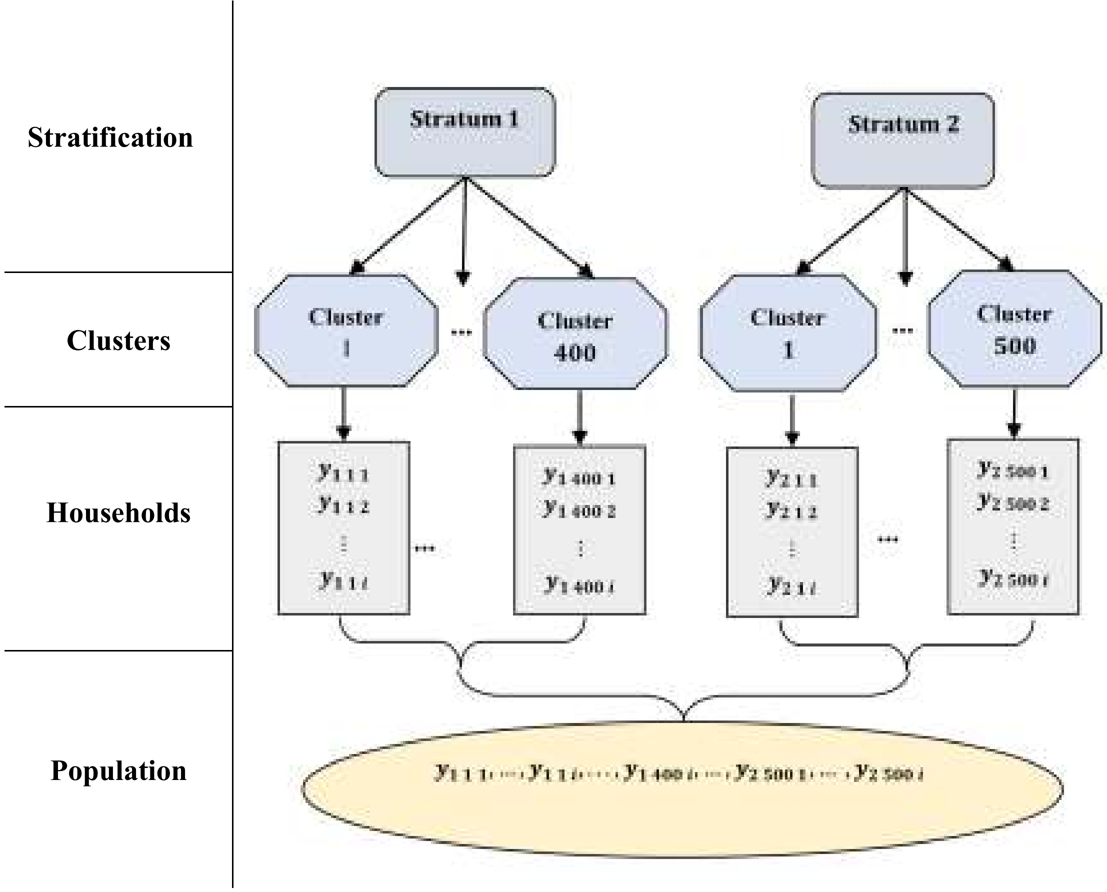
In the warp-speed method case, lower and upper tail coverage probabilities are calculated using the formulae in expression (3.12) by replacing $\hat{q}_K$ for $\hat{q}_k$, as shown in expression (3.10).

3.3 Monte Carlo Experiment Designs

The Monte Carlo experiment is a controlled experiment, commonly applied in statistics for examining sampling distributions and comparing the properties of estimators and test statistics. An appropriate design of a MC experiment is vital in achieving the desired outcomes from computer-based simulation experiments, where the selection of the DGP is considered to be the most important part of the design (see, e.g., Paxton et al., 2001). It is desirable to select DGPs that are consistent with theoretical models and assumptions and reflect real-world applications of interest. The DGP must be fully known – the forms of equations, numerical values of the parameters, and actual values of the random numbers (see, e.g., Davidson and Mackinnon, 2004, p. 19; Hendy, 1984, p. 945).

We generate stratified two-stage cluster finite populations following the survey design described in Chapter 2. To recap, 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 within stratum h, and in cluster c within stratum h, there are M_{hc} households, leading to $\mathcal{M}_h = \sum_{c=1}^{N_h} M_{hc}$ households in stratum h. Hence, the finite population contains a total of $\mathcal{N} = \sum_{h=1}^{H} N_h$ clusters and $\mathcal{M} = \sum_{h=1}^{H} \sum_{c=1}^{N_h} M_{hc}$ households. Basic principles of constructing our chosen DGPs is provided in Figure 2, with an example of two strata. The population is formed for each stratum with the desired number of clusters and households, and then combined across the strata.

Figure 2 Principles of a complex survey design with two strata



We explore properties for 36 DGPs using a random effects framework. The underlying well-being variable y in each of the DGPs is formed using the following random effects model:

$$y_{hci} = \mu_0 + \lambda_{hc} + z_{hci}, \qquad h = 1, \dots, H; c = 1, \dots, N_h; \; i = 1, \dots, M_{hc}\,, \qquad (3.13)$$

where μ_0 is a constant chosen to ensure that all observations are positive, λ_{hc} and z_{hci} are independently distributed random effects – cluster effect and household idiosyncratic effect, respectively. The cluster effect variable λ_{hc} incorporates the between cluster variation in the population and the variation among households within cluster c in stratum h is captured by the random effect variable, z_{hci}.

For ease of demonstration, we assume that the random effects for clusters are normally distributed as $\lambda_{hc} \sim N(0, \gamma_h^2)$. Although we recognize the potential impact of this choice on our results, we leave such an exploration for future research. For household effects, we study the effect of four different probability distributions: beta, lognormal, chi-square, and the Pareto, with varying parameter values, so that $z_{hci} \sim (\mu_{hc}, \sigma_h^2)$. Our choice of these distributions for z_{hci} are motivated by the fact that they are skewed, some are heavy-tailed distributions, and seem to approximately be a good fit to DHS well-being variables, such as women's body mass index and hemoglobin level in children.

This framework implies that $E(y_{hci}) = \mu_h$, allowing for the population mean of y to vary across strata, and $Var(y_{hci}) = \psi_h^2 = \gamma_h^2 + \sigma_h^2$. The first term of the variance of the well-being variable, $Var(\lambda_{hc}) = \gamma_h^2$, is determined by the intracluster correlation coefficient (ICC), ρ_h, which is allowed to take on different values, such that $\gamma_h^2 = \psi_h^2 \rho_h$.

The ICC in stratum h, ρ_h, a key parameter in our experiments, measures the relatedness of the clustered data within a stratum h, and is the ratio of the between-cluster variance to the total variance (between and within clusters) (e.g., Skinner et al., 1989, p. 37):

$$\rho_h = \frac{\gamma_h^2}{\gamma_h^2 + \sigma_h^2}. \qquad 0 \leq \rho_h \leq 1, h = 1, \ldots, H. \tag{3.14}$$

Clearly, $\rho_h = 0$, when $\gamma_h^2 = 0$; i.e., there is no cluster effect within the stratum. An increasing ρ_h corresponds to the case that units within the same cluster in the hth stratum are more similar than units selected from the stratum population at random. When $\rho_h \to 1$ as $\sigma_h^2 \to 0$, we have the theoretical case that y is identical for all households in a cluster in stratum h. In this case, the effective number of observations in the stratum is equal to the number of clusters. For real-world complex surveys, $0 < \rho_h < 1$, with similarities among households in clusters leading to variances for sample statistics being higher than would be the case under a simple random sample. For a particular DGP, when parameters of the distribution for z_{hci} are specified, the variance of the distribution for the cluster effect variable, γ_h^2, is calculated using expression (3.14), which is $\gamma_h^2 = \frac{\rho_h \sigma_h^2}{(1-\rho_h)}$.

We generate two categories of finite populations: DGPs with two strata and five strata. In the category with two strata, there are seven DGPs under each probability distribution for z_{hci}, totalling 28 DGPs. For the first six DGPs from each of the distributions, we assume that both strata have the same ICC value, and the seventh DGP contains varying ICC values across strata. In the second group of DGPs with five strata, two DGPs from each distribution are created, one with a fixed ICC and the other with varying ICC values across strata, a total of eight DGPs. In addition, an extra DGP with three strata and a constant ICC value across strata is created for further investigation of the heavy-tail Pareto distribution. However, we drop a DGP under the chi-square distribution for ICC equals 0.9, which produces some negative values of the well-being variable, resulting in a total of 36 DGPs for our explorations.

Our elements of observation are households rather than individuals, which stands in contrast to Bhattacharya (2007) and our framework of Chapter 2, and indeed, with our empirical applications in Chapter 4. Nevertheless, as explained in Chapter 2, whether the well-being variable is measured at the individual or household level does not matter, as long as the relevant sampling weight is used (for example, see, Skinner et al., 1989, p. 47-48, for further discussion).

We now detail each of our fundamental DGPs, motivating the choices of underlying distributions. As we choose the distribution for household effects in the data generating process, which is a stochastic component in the model in expression (3.13), we name DGPs according to the distribution of z_{hci}. In addition, a brief outline on sampling techniques for MC and bootstrapping is presented at the end of this section.

In our study, DGPs from the beta, lognormal, chi-square, and Pareto distributions are defined by DGP1, DGP2, DGP3, and DGP4, respectively. DGPs from each of the distributions are denoted as A, B, C, and so on. For example, 9 DGPs from the beta distribution are DGP1A – DGP1I, and 10 DGPs from the Pareto distributions are DGP4A – DGP4J. Each DGP ends with a letter to represent population features, e.g., the A DGPs have similar population features – number of strata, clusters, households, and ICC values. For the first seven DGPs from each distribution (A – G), each contains two strata, $H = 2$, and the number of clusters in stratum 1 and 2 are N_1 =400 N_2 =500, respectively. The number of households within a cluster is allowed to randomly vary from 80 to 200. For DGPs A – F from each distribution, the ICCs in both strata are assumed to be the same with values of 0.0, 0.1, 0.3, 0.5, 0.7, and 0.9, respectively. For the G DGPs, ICC values are different across strata – ρ_1=0.05 and ρ_2=0.10.

The two H and I DGPs from each of the distributions, are generated with more strata, $H = 5$, so that we may ascertain the influence of this population feature on the sampling properties. The number of clusters in each of the strata are: N_1 =100, N_2 =150, N_3 =200, N_4 =250, and N_5 =300. In each cluster within h, the number of households is chosen randomly to be between 30 and 100. For the H DGPs, the ICC value across strata is assumed to be fixed at 0.1, whereas for the I DGPs, each stratum is allowed to have a different ICC value: ρ_1=0.05, ρ_2=0.2, ρ_3=0.10, ρ_4=0.3, and ρ_5=0.15.

The last DGP from the Pareto distribution, DGP4J, is generated with three strata, $H = 3$, a fixed ICC value of 0.1 across strata, and the numbers of clusters in the strata being N_1 =200, N_2 =250, and N_3 =300. Intending to create a larger population, the number of households are randomly generated to be 80-200 in each cluster within a stratum. We now provide details about each data generating process.

DGP1: ***DGP from the Beta distribution for*** $\boldsymbol{z_{hci}}$

The beta distribution is widely used in empirical and simulation studies (e.g., Van Ourti and Clarke, 2011; Deltas, 2003; Paolino, 2001; Mebane, 2000; Brehm and Gates, 1993) due to its large variation in the shape of the density function, including choices of parameters that enable the distribution to be left- and right-skewed, symmetric and bimodal, and to model the behaviour of random variables limited to intervals of finite length. Given that our studied BDHS well-being variables, BMI and Hb, are skewed, sampling properties of $\hat{G}$ from data generated using the beta distribution may be helpful with well-being variables such as those we examine from DHS data.

The beta distribution, defined on the interval [0, 1], is parameterized by two shape parameters. This unit interval constraint does not produce any loss of generality to our inequality measure, the Gini coefficient, which does not depend on the scale of the distribution (see, e.g., Deltas, 2003). In addition, variance and kurtosis of the distribution can vary independently from the skewness (see, e.g., Van Ourti and Clarke, 2011).

The household effect variable, z, of the random effects model in (3.13), has the beta distribution such that $z \sim Beta(\alpha, \beta)$, where α and β are shape parameters, and $\alpha, \beta > 0$. The probability density function of z is given by $f(z) = \frac{z^{\alpha-1}(1-z)^{\beta-1}}{B(\alpha,\beta)}$, where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta function, and the cumulative distribution function is $F(z) = I(z;\ \alpha, \beta)$. Using the moments for a beta random variable, $E(z) = \frac{\alpha}{(\alpha+\beta)}$, $Var(z) = \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ and $Skewness(z) = \frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}$. To match the right-skewness of many real-world data of interest, we assume $\alpha < \beta$ (see, Figure C.1 in Appendix C for an illustration).

Nine DGPs are generated from the beta distribution with different shape parameters and ICC values. For the first seven DGPs, DGP1A-DGP1G, we assume households in strata 1 and 2 are distributed with $z_{1ci} \sim Beta(2, 5)$ and $z_{2ci} \sim Beta(3,6)$. For DGPs with five strata, DGP1H – DGP1I, households within a stratum are generated following $z_{1ci} \sim Beta(2, 4)$, $z_{2ci} \sim Beta(2.5,4.5)$, $z_{3ci} \sim Beta(3,5)$, $z_{4ci} \sim Beta(4,6)$, and $z_{5ci} \sim Beta(5,7)$.

For a specific ICC value (ρ_h), the variance of the cluster effect random variable (γ_h^2) in stratum h is calculated using $\gamma_h^2 = \frac{\rho_h \sigma_h^2}{(1-\rho_h)}$ formula. We allow the ICC to vary to generate different DGPs. The calculated values of γ_h^2 for DGPs under the beta distribution are presented in Table 1. The constant of the random effect model for DGPs from the beta distribution is kept fixed at $\mu_0 = 2$.

Table 1 Intracluster correlation coefficients, and variances of the cluster effect variable for DGPs from the beta distribution, DGP1A-DGP1I.

DGP	ICC, ρ_h	γ_1^2	γ_2^2	γ_3^2	γ_4^2	γ_5^2
DGP1A	ρ_1=ρ_2=0.0	0.000	0.000	-	-	-
DGP1B	ρ_1=ρ_2=0.1	0.003	0.003	-	-	-
DGP1C	ρ_1=ρ_2=0.3	0.011	0.010	-	-	-
DGP1D	ρ_1=ρ_2=0.5	0.026	0.022	-	-	-
DGP1E	ρ_1=ρ_2=0.7	0.060	0.052	-	-	-
DGP1F	ρ_1=ρ_2=0.9	0.230	0.200	-	-	-
DGP1G	ρ_1=0.05, ρ_2=0.1	0.001	0.003	-	-	-
DGP1H	$\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5$ =0.10	0.004	0.032	0.003	0.003	0.002
DGP1I	ρ_1 =0.05, ρ_2 =0.20, ρ_3 =0.10, ρ_4 =0.30, ρ_5 =0.15	0.002	0.007	0.003	0.012	0.003

DGP2: ***DGP from the Lognormal (LN) distribution for z_{hci}***

The lognormal distribution is positively skewed, and has been commonly used in MC simulations to undertake inference about G; for example, see Van Ourti and Clarke

(2011), Qin et al. (2010), Davidson (2009), Berger (2008), Gastwirth et al. (2005), Deltas (2003), and Dixon et al. (1987). Many suggest that it might reasonably approximate income distributions. We consider this distribution to generate 9 DGPs to explore the sampling properties of $\widehat{G}$ under a complex survey design.

The household effect random variable, z, has a lognormal distribution with mean μ and the shape parameter or standard deviation σ: $z \sim LN(\mu, \sigma^2)$. The density function of z is $f(z) = \frac{1}{z\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln z-\mu)^2}{2\sigma^2}\right\}$, and the CDF is $F(z) = \Phi\left(\frac{\ln(z)}{\sigma}\right) \quad z \geq 0;\ \sigma > 0$, where $\Phi(.)$ is the CDF of a normal distribution. Using the moments conditions, $E(z) = \exp\left\{\mu + \frac{\sigma^2}{2}\right\}$, $Var(z) = \left(e^{\sigma^2} - 1\right)\exp\{2\mu + \sigma^2\}$, $\text{Skewness}(z) = \left(e^{\sigma^2} + 2\right)\sqrt{e^{\sigma^2} - 1}$ and $\text{Excess kurtosis}(z) = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 6$. The variance, skewness and kurtosis of the lognormal distribution increase with σ (see Figure C.1 in Appendix C).

For DGPs with two strata, DGP2A-DGP2G, distributions of z_{hci} in strata 1 and 2 are assumed to be $z_{1ci} \sim LN(1.119, 0.0602^2)$ and $z_{2ci} \sim LN(1.5, 0.4^2)$, respectively. For DGP2H – DGP2I, populations with five strata, distributions of z_{hci} follow $z_{1ci} \sim LN(1.119, 0.602^2)$, $z_{2ci} \sim LN(1.5, 0.4^2)$, $z_{3ci} \sim LN(1.7, 0.35^2)$, $z_{4ci} \sim LN(1.9, 0.25^2)$ and $z_{5ci} \sim LN(2.0, 0.5^2)$. These choices of parameters were made to produce reasonable right skewed shapes for the underlying distributions. Variances of the cluster effect variable for different values of the ICC parameters, $\gamma_h^2 = \frac{\rho_h \sigma_h^2}{(1-\rho_h)}$, are reported in Table 2. The constant in the random effect model in (3.13) is assumed to be $\mu_0 = 20$.

Table 2 Intracluster correlation coefficients, and variances of the cluster effect variable for DGPs from the lognormal distribution, DGP2A-DGP2I

DGP	ICC, ρ_h	γ_1^2	γ_2^2	γ_3^2	γ_4^2	γ_5^2
DGP2A	ρ_1=ρ_2=0.0	0.000	0.000	-	-	-
DGP2B	ρ_1=ρ_2=0.1	0.654	0.454	-	-	-
DGP2C	ρ_1=ρ_2=0.3	3.051	1.753	-	-	-
DGP2D	ρ_1=ρ_2=0.5	7.199	9.543	-	-	-
DGP2E	ρ_1=ρ_2=0.7	16.611	9.543	-	-	-
DGP2F	ρ_1=ρ_2=0.9	64.071	36.810	-	-	-
DGP2G	ρ_1=0.05, ρ_2=0.1	0.375	0.454	-	-	-
DGP2H	$\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = 0.1$	0.791	0.454	0.490	0.034	2.213
DGP2I	ρ_1 =0.05, ρ_2 =0.20, ρ_3 =0.10, ρ_4 =0.30, ρ_5 =0.15	0.375	1.023	0.460	1.315	3.515

DGP3: ***DGP from the Chi-square*** (χ^2) ***distribution for*** $\boldsymbol{z_{hci}}$

This DGP assumes that the household effect variable, z, has a chi-square (χ_v^2) distribution with v degrees of freedom, $z \sim \chi^2(v)$. Following Qin et al. (2010), we generate complex survey data from a $\chi^2(v)$ distribution for different degrees of freedom under the realization that distributions of some well-being variables are slightly skewed, so that a $\chi^2(v)$ may describe the data well.

The probability density and CDF of a $\chi^2(v)$ random variable are $f(z) = \frac{2^{-v/2}}{\Gamma(v/2)} z^{(v/2-1)} e^{-z/2}$ and $F(z) = \gamma\left(\frac{v}{2}, \frac{z}{2}\right)/\Gamma(v/2)$ for $z \geq 0$, respectively, where Γ and γ are the gamma and the incomplete gamma functions, respectively. Using the moments for a $\chi^2(v)$ random variable, $E(z) = v$, $Var(z) = 2v$, $\text{skewness}(z) = \sqrt{8/v}$ and excess kurtosis $(z) = 12/v$.

For DGP3A-DGP3G, we assume that $z_{1ci} \sim \chi^2(5)$ and $z_{2ci} \sim \chi^2(10)$, and for DGP3H – DGP3I, $z_{1ci} \sim \chi^2(5), z_{2ci} \sim \chi^2(10)$, $z_{3ci} \sim \chi^2(15)$, $z_{4ci} \sim \chi^2(18)$, and $z_{5ci} \sim \chi^2(22)$. For various ICC values, calculated variances of the cluster effect variable are reported in Table 3. The constant in the random effect model is assumed to be $\mu_0 = 20$ across DGPs.

Table 3 Intracluster correlation coefficients, and variances of the cluster effect variable for DGPs from the Chi-square distribution, DGP3A – DGP3I.

DGP	ICC, ρ_h	γ_1^2	γ_2^2	γ_3^2	γ_4^2	γ_5^2
DGP3A	ρ_1=ρ_2=0.0	0.000	0.000	-	-	-
DGP3B	ρ_1=ρ_2=0.1	1.111	2.222	-	-	-
DGP3C	ρ_1=ρ_2=0.3	4.286	8.571	-	-	-
DGP3D	ρ_1=ρ_2=0.5	10.000	20.000	-	-	-
DGP3E	ρ_1=ρ_2=0.7	23.333	46.667	-	-	-
DGP3F	ρ_1=ρ_2=0.9	90.000	180.00	-	-	-
DGP3G	ρ_1=0.05, ρ_2=0.1	0.526	2.222	-	-	-
DGP3H	$\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5$ =0.1	1.111	2.222	3.333	4.000	4.888
DGP3I	ρ_1 =0.05, ρ_2 =0.20, ρ_3 =0.10, ρ_4 =0.30, ρ_5 =0.15	0.526	5.000	3.333	15.429	7.765

Note: DGP3F is dropped from further examination as some observations are negative.

DGP4: DGP *from the Pareto distribution for* z_{hci}

For the continuous household effect random variable z, we now consider the use of the Pareto distribution, a heavily right-skewed distribution. The Pareto distribution is often used to model income data (see, e.g., Eckerstoreer et al., 2014, and citations therein; Langel and Tillé, 2013; Omar et al., 2012; Davidson, 2009; Gastwirth et al., 2005) and adopted to explore the consequences of data from a heavy-tailed distribution on sampling properties of a statistic of interest.

The Pareto distribution with shape and scale parameters, α and β, respectively, is $z \sim Pareto(\alpha, \beta)$, where $\alpha \, \epsilon (0, \infty)$ and $\beta > 0$. The scale parameter β can be assumed to equal 1 without any loss of generality (e.g., Omar et al., 2012), reducing the expression to $z \sim Pareto(\alpha)$. Then, the probability density function and CDF of z are given by $f(z) = \frac{\alpha}{z^{\alpha+1}}$, $z \geq 1$ and $F(z) = 1 - (1/z)^{\alpha}$, $z \geq 1$, respectively. Using the moments for a Pareto random variable, where we have $E(z) = \begin{cases} \infty & \text{for } \alpha \leq 1 \\ \frac{\alpha}{(\alpha-1)} & \text{for } \alpha > 1 \end{cases}$, variance$(z) = \sigma^2 =$

$\begin{cases} \infty & \text{for } \alpha\epsilon(1,2] \\ \frac{\alpha}{(\alpha-1)^2(\alpha-2)} & \text{for } \alpha > 2 \end{cases}$, skewness$(z) = \frac{2(1+\alpha)}{(\alpha-3)}\sqrt{\frac{\alpha-2}{\alpha}}$ for $\alpha > 3$ and excess kurtosis= $6(\alpha^3 + \alpha^2 - 6\alpha - 2)/(\alpha(\alpha - 3)(\alpha - 4))$ for $\alpha > 4$.

We generate 10 DGPs assuming z_{hci} has the Pareto distribution with various values of α across strata. For DGP4A-DGP4G, we consider α is 2 and 5 in strata 1 and 2, respectively, so that $z_{1ci} \sim Pareto(2)$ and $z_{2ci} \sim Pareto(5)$. In DGPs with five strata, DGP4H-DGP4I, the household effect variable is distributed with α equaling values of 3, 4, 5, 6, and 7. For DGP4J, $z_{1ci} \sim Pareto(2)$, $z_{2ci} \sim Pareto(3)$ and $z_{3ci} \sim Pareto(5)$. Variances of the cluster effect variable in stratum h, γ_h^2, for those DGPs are reported in Table 4. The constant of the random effects model is kept fixed at μ_0 =200 across DGPs.

Table 4 Intracluster correlation coefficients and variances of the cluster effect variable for DGPs the Pareto distribution, DGP4A-DGPJ.

DGP	ICC, ρ_h	γ_1^2	γ_2^2	γ_3^2	γ_4^2	γ_5^2
DGP4A	ρ_1=ρ_2=0.0	0.000	0.000	-	-	-
DGP4B	ρ_1=ρ_2=0.1	0.012	0.012	-	-	-
DGP4C	ρ_1=ρ_2=0.3	0.045	0.045	-	-	-
DGP4D	ρ_1=ρ_2=0.5	0.104	0.104	-	-	-
DGP4E	ρ_1=ρ_2=0.7	0.243	0.243	-	-	-
DGP4F	ρ_1=ρ_2=0.9	0.938	0.938	-	-	-
DGP4G	ρ_1=0.05, ρ_2=0.1	0.045	0.045	-	-	-
DGP4H	ρ_1=ρ_2=ρ_3=ρ_4=ρ_5=0.1	0.083	0.025	0.012	0.007	0.004
DGP4I	ρ_1=0.05, ρ_2=0.20, ρ_3=0.10, ρ_4=0.30, ρ_5=0.15	0.188	0.039	0.012	0.003	0.021
DGP4J	ρ_1=ρ_2=ρ_3=0.1	0.083	0.083	0.012	-	-

We see that when the ICC increases, households within the same cluster tend to be more similar than when they are randomly selected from the population, the variance of the cluster effect variable within h, γ_h^2, increases (see Tables 1- 4). When γ_h^2 increases, the variance of the well-being variable, y_{hci}, in stratum h increases, but the skewness of the distribution decreases – the distribution moves towards a balanced distribution. We see, under the beta, lognormal, and chi-square distributions, that when the ICC value increases in a particular DGP, for other features given, the cluster effect distribution dominates the household effect distribution, and the stratum and population distributions move towards the center (see the illustration in Figure C.2 under the lognormal distribution in Appendix C). Nevertheless, this is not the case under the Pareto distribution for z_{hci}. Although the variance of y_{hci} increases with the larger values of ICC, the cluster effect shows no sign of dominating the household effect. The population distribution remains extremely right-skewed, like the Pareto distribution (see Figure C.3 in Appendix C) when the ICC is equal to 0.5 or 0.9.

Using our two sets of finite populations (DGPs with two strata and five strata), we conduct two types of simulations. In the first, we consider DGPs with two strata (first six DGPs under each distribution for z_{hci}, A-F), referring to it as *the base case* simulations. The main objective of these experiments is to examine the impacts of the ICC value and the distribution of the well-being variable on the bias of $\hat{G}$ and on the performance of interval estimators for G. In the second type, we use two DGPs with two strata (B and G) and two DGPs with five strata (H and I) from each distribution to explore the behaviour of empirical bias and coverage probabilities when the size of a complex survey sample (number of clusters in the sample) increases.

For the base case experiments, MC samples are selected by applying the stratified two-stage cluster sampling method. In the first stage, 10% of the population clusters are selected with probability proportional to size, i.e., 40 and 50 clusters from strata 1 and 2 ($n_1 = 40$ and $n_2 = 50$), respectively. Then, in the second stage, a fixed 30 households (ultimate units) ($m = 30$) from each cluster are drawn by applying the simple random sampling technique, resulting in a complex survey sample of size 90 clusters or 2,700 households. Bootstrap samples from a MC sample are drawn, as indicated previously, by taking one less cluster of the clusters in the MC sample, $n_h^* = n_h - 1$, though the number of households from each cluster remains the same, $m = 30$. The size of a bootstrap sample is 88 clusters or 2,640 households.

For the second type of simulation experiments, we draw four samples from each DGP examined. Samples are denoted by MCS1-4. When samples are drawn from DGPs with two strata, the number of clusters in MCS1 is 25 ($n_1 = 10, n_2 = 15$), in MCS2 the number is 50 ($n_1 = 25, n_2 = 25$), in MCS3 the number is 90 ($n_1 = 40, n_2 = 50$), and in MCS4 the number is 165 ($n_1 = 75, n_2 = 90$). For DGPs with five strata, 31 clusters are in MCS1 ($n_1 = 3, n_2 = 5$, n_3=6, n_4=8, n_5=9), 63 clusters are in MCS2 ($n_1 = 7, n_2 = 9$, n_3=12, n_4=15, n_5=20), 124 clusters are in MCS3 ($n_1 = 12, n_2$ =20, n_3=24, n_4=32, n_5=36), and 214 clusters are in MCS4 ($n_1 = 25, n_2 = 35$, n_3=45, n_4=64, n_5=70). Four DGPs from each distributional assumption are studied for this category of simulations: DGP-B, DGP-G, DGP-H and DGP-I, and the number of households drawn from each cluster remains fixed at 30 when a DGP contains two strata, but it is fixed at 20 when five strata are in a DGP.

In our simulations, we examine the finite sampling properties of $\hat{G}$ using 1,000 MC samples from a DGP and 99 bootstrap resamples. We limit our attention to 99 bootstrap resamples for ease of computation. Besides, the prior research suggests that often 50 to 200 bootstrap samples are sufficient to achieve desired properties of statistics (see, e.g., Bruch et al., 2011; Wolter, 2007, p. 195). Estimations and simulations are undertaken using STATA 9.2 (StataCorp, 2007), STATA 11 (StataCorp, 2011) and EViews 8 (Quantitative Micro Software, 2013). See Appendix E for some sample EViews and STATA program codes.

3.4 Simulation Results

In this section, we discuss results from our MC simulations on the finite sample properties of Gini coefficient estimators with a wide range of complex survey samples. In subsection 3.4.1, we present and discuss results on the bias. The performance of the 95% CI estimators for G are examined in subsection 3.4.2.

3.4.1 Bias of the Gini Coefficient Estimator

As previously discussed, we perform two sets of MC experiments to evaluate the accuracy of $\hat{G}$. In the first group of experiments, *the base case*, we examine biases of $\hat{G}$ estimated under a fixed sampling design when samples are obtained from DGPs with two strata and the ICC values increase. In the second category, collection of experiments, we study the impact of increasing number of sampled clusters on the bias of $\hat{G}$.

The population Gini coefficients and estimated bias of the estimator $\hat{G}$, as a percentage of G, for the base case experiments for 1,000 MC samples are reported in Table 5. We see that the population values of G, under the different distributional assumptions ranges from 0.002 to 0.177, which seem to be relatively smaller in magnitude compared with the values that are commonly observed when income is the well-being variable (e.g., estimated values that are often 0.30 to 0.50 under an iid assumption). For our simulations, data are generated to be consistent with values observed for our well-being variables illustrated in Chapter 4. One reason the values are smaller is because of the ranges of the well-being variables being examined. For instance, we observe Hb levels that range from 3.1 to 14.7, unlike an income variable that may range from zero to millions of dollars for example. Most of the Gini coefficients for our MC cases are consistent with those for the well-being variables BMI and Hb from the BDHS samples. For instance, the estimated

Gini coefficients range from 0.088 to 0.10 and 0.04 to 0.063 for BMI and Hb variables, respectively.

In the case of the Pareto distribution, some of our population Gini coefficients are near the lower boundary of zero of the parameter space. This may cause concern about the applicability of the limiting normal distribution for our Gini coefficient estimator, considered in Chapter 2. However, under some conditions (see, e.g., Andrews, 1999; Ketz, 2014), it may be possible to determine the asymptotic distribution of an estimator when the true parameter is near or at the boundary of the parameter space. We leave this question for future exploration.

Table 5 Estimated biases of $\hat{G}$, as % of G, in samples with 90 clusters (n_1=40 and n_2=50) for 1,000 MC samples, K=1,000.

DGP	ICC $\rho_1 = \rho_2$	G	$\hat{E}(\hat{G})$	Bias$(\hat{G})$, % of G
		Beta distribution for z_{hci}		
DGP1A	0.0	0.0381	0.0384	0.78
DGP1B	0.1	0.0404	0.0408	0.99
DGP1C	0.3	0.0461	0.0464	0.65
DGP1D	0.5	0.0548	0.0550	0.37
DGP1E	0.7	0.0708	0.0707	-0.14
DGP1F	0.9	0.1226	0.1221	-1.22
		Lognormal distribution for z_{hci}		
DGP2A	0.0	0.0498	0.0505	1.41
DGP2B	0.1	0.0542	0.0546	0.74
DGP2C	0.3	0.0636	0.0640	0.63
DGP2D	0.5	0.0872	0.0871	-0.12
DGP2E	0.7	0.1013	0.1011	-0.20
DGP2F	0.9	0.1773	0.1751	-1.24
		Chi-square distribution for z_{hci}		
DGP3A	0.0	0.0920	0.0923	0.33
DGP3B	0.1	0.0976	0.0982	0.61
DGP3C	0.3	0.1091	0.1097	0.55
DGP3D	0.5	0.1262	0.1264	0.16
DGP3E	0.7	0.1578	0.1569	-0.57
		Pareto distribution for z_{hci}		
DGP4A	0.0	0.0020	0.0025	17.22
DGP4B	0.1	0.0021	0.0024	16.42
DGP4C	0.3	0.0022	0.0025	15.38
DGP4D	0.5	0.0024	0.0027	13.81
DGP4E	0.7	0.0027	0.0030	11.91
DGP4F	0.9	0.0038	0.0041	7.87

In complex survey samples, estimators are assumed to be unbiased or approximately unbiased (see, Skinner et al., 1989, p. 28, 238-242) when sampling weights are accounted for appropriately. While this is true for typical estimators (e.g., population total or mean), it may not be valid for nonlinear and complex statistics, such as $\widehat{G}$ (see, e.g., Langel and Tillé, 2011); this finding is borne out by our simulation results. We see that the overall estimated bias of $\widehat{G}$ is much smaller, less than 1.5% of G, in base cases, but varies across distributions (beta, lognormal, and chi-square) for samples with 10% of the population clusters. In the case of the Pareto distribution, the biases are not negligible, ranging between 7-18%.

We observe that the bias of $\widehat{G}$ falls when the ICC value in a DGP increases. For ICC values from zero to approximately one-half, the bias is positive, whereas for larger ICCs it is negative for the beta, lognormal, and chi-square distributions. These features highlight that the sampling distribution of $\widehat{G}$ depends on the underlying distribution of data (see, e.g., Gastwirth et al., 2005). For an increasing ICC, households within clusters are likely to be more homogeneous and the distribution of the well-being variable tends to be more symmetrical. We see that the magnitudes of the bias under these distributions are somewhat related to the skewness of the population data (see, Figure C.2 in Appendix C).

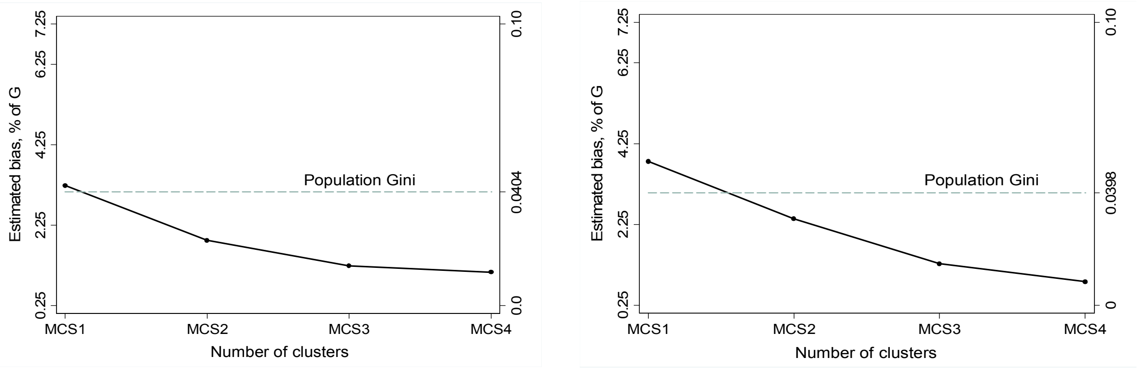
The Gini coefficient is well known to be sensitive to extreme values in a sample drawn from a heavy-tailed distribution (see, e.g., Langel and Tillé, 2011; Davidson, 2009). Unlike the other three distributions we consider, the skewness of the data under the Pareto distribution does not fall substantially for an increasing value of the ICC (see, Figure C.3 in Appendix C), and therefore, the bias of $\widehat{G}$ continues to be high. For instance, even though the bias of $\hat{G}$ falls sharply with the increasing value of ICC, it remains noticeably high (almost 8% of G) for the largest ICC value of 0.9.

Under an *iid* framework, the bias of $\hat{G}$ is typically downward in small samples (e.g., Davidson, 2009; Deltas, 2003). We estimate the bias of $\hat{G}$ in *iid* samples drawn from populations generated with $Beta(2,5)$, $lognormal(1.5, 4^2)$ and $Pareto(5)$ distributions. Estimated biases of $\hat{G}$ from 1,000 MC samples of size 100 are -0.003%, -0.018%, and -0.002% drawn from DGPs under those three distributions respectively. These results are consistent with existing studies. Nevertheless, for our complex survey samples with a moderate size (90 sampled clusters – 10% of the population clusters), biases of $\hat{G}$ are upward when samples are drawn from finite populations with ICC values from 0 to around 0.5. Therefore, the estimated Gini coefficient from *iid* samples and complex survey samples are sometimes quite different, even when a complex survey sample comes from a population with no cluster effect.

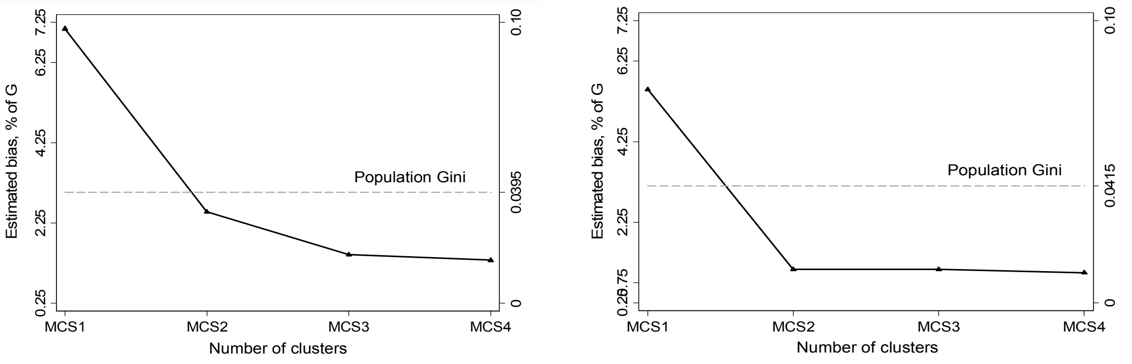
Next, we examine changes in the bias of $\hat{G}$ in samples with increasing number of sampled clusters. The bias of $\hat{G}$ in samples from four distributions and different sampling designs are shown in Figures 3- 6. We consider four sample sizes to demonstrate the changes in the bias of $\hat{G}$. For each of these distributions, we studied four cases: two from DGPs with two strata, and two from DGPs with five strata. DGPs and the number of clusters in samples are indicated in the notes for Figures 3 - 6 (for details on DGPs, see Section 3.3).

Simulation results for 1,000 MC samples on biases of $\hat{G}$, when the household effect variable has the beta distribution, are illustrated in Figure 3. Plots a-b and c-d in Figure 3 show the sampling behaviour when samples are drawn from DGPs with two strata and five strata, respectively. We see that when the sample contains a small number of clusters (approximately 3% of the population clusters, denoted by MCS1), the biases of $\hat{G}$ are 3.24% and 3.81% of G for a constant ICC of 0.1 and varying ICCs across strata, respectively, in DGPs with two strata. In contrast, the biases are 7.0% and 5.5% in samples from DGPs with five strata DGPs. The bias is relatively higher in small samples drawn from DGPs with more strata.

Figure 3 Estimated bias of $\hat{G}$, as % of G, for 1,000 MC samples, K =1,000: Beta distribution for the household effect variable.



a) DGP1B ($\rho_1 = \rho_2 = 0.1$) b) DGP1G ($\rho_1 = 0.05,\ \rho_2 = 0.1$)



c) DGP1H ($\rho_1 = \cdots = \rho_5 = 0.1$) d) DGP1I ($\rho_1 = .05, \rho_2 = .2, \rho_3 = .1, \rho_4 = .3, \rho_5 = .15$)

Notes: For DGP1B and DGP1G, MC samples are MCS1: $n_1 = 10, n_2 = 15$; MCS2: $n_1 = 25, n_2 = 25$; MCS3: $n_1 = 40, n_2 = 50$; MCS4: $n_1 = 75, n_2 = 90$. For DGP1H and DGP1I, MC samples are clusters MCS1: $n_1 = 3, n_2 = 5$, n_3=6, n_4=8, n_5=9; MCS2: $n_1 = 7, n_2 = 9$, n_3=12, n_4=15, n_5=20; MCS3: $n_1 = 12, n_2$ =20, n_3=24, n_4=32, n_5=36; MCS4: $n_1 = 25, n_2 = 35$, n_3=45, n_4=64, n_5=70.

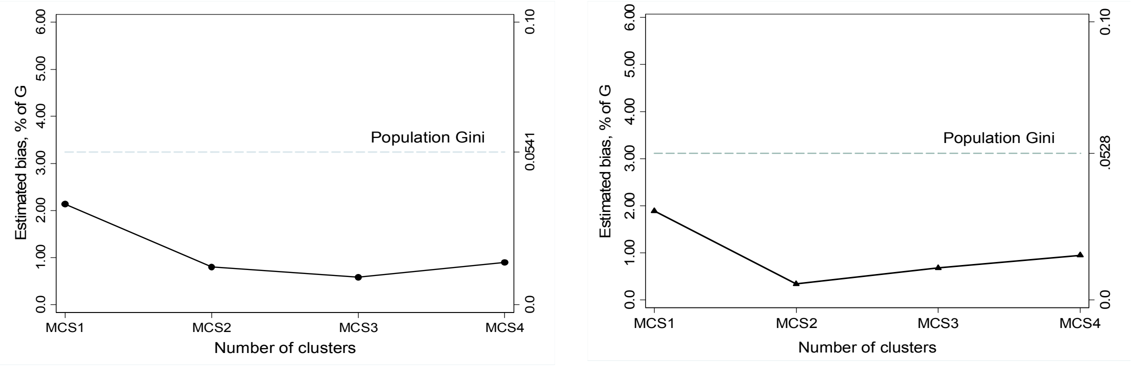
When the number of clusters in samples increases, bias of $\hat{G}$ falls rapidly. For DGPs with both constant and varying ICCs across strata, the bias falls by approximately

40% in samples with 50 clusters or less than 6% of the population clusters. The bias becomes negligible (less than 1% of G) in samples containing approximately 18% of the population clusters (165 clusters). On the other hand, for samples with about 6% of the population clusters (63 clusters) from DGPs with five strata, the bias reduces by 64% and 80%, when ICCs are constant at 0.1 and varying across strata, respectively. For the largest sample, MCS4, containing 24% of the population clusters (239 clusters), the bias is approximately 1% of G.

We also notice that the difference in bias for samples with a fixed 0.1 ICC and varying ICCs (0.05 to 0.35) across strata is trivial, especially in samples with two strata. Moreover, the bias of $\hat{G}$ falls at a faster rate with an increasing number of clusters in samples with five strata than in samples with two strata. This may occur because estimators in samples with more strata and larger number of clusters tend to behave more stably (see, e.g., Hong, 2015, p. 89).

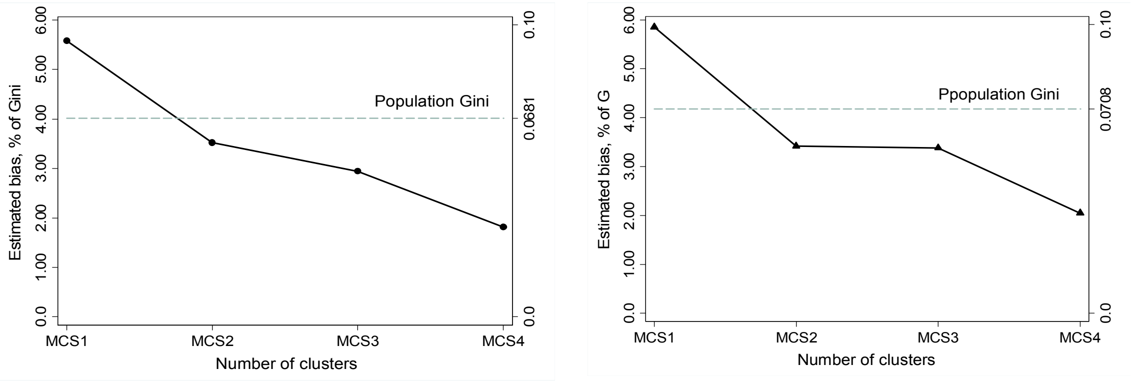
Biases of $\hat{G}$ for sampling from DGPs generated under the lognormal distributional assumption are shown in Figure 4. We see that for DGPs with two strata, the small cluster sample (MCS1) bias is relatively lower, 2-3% of G, for this distribution than for that under the beta distribution. This is because the empirical distribution of $\hat{G}$ with fewer clusters under the lognormal distribution is more balanced than the beta distribution (see Figure C.4 in Appendix C). In samples with about 6% of the population clusters (50 clusters), $\hat{G}$ is approximately unbiased (the bias is less than 1%). On the other hand, for small samples, MCS1, selected from DGPs with five strata, the bias is approximately 6% of G. The difference between the biases for data obtained from DGPs with a constant ICC (0.1) and varying ICCs across strata is trivial (less than 0.25%). Although the bias reduces with an increasing number of clusters in the sample, with the bigger sample MCS4, $\hat{G}$ remains biased, which is approximately 2% of G. Observing the trend (see plots c and d in Figure 4), a bigger sample with more clusters appears to need to be sampled to achieve an approximately unbiased estimator of G, when DGPs contain five strata and nonzero ICCs across strata.

Figure 4 Estimated bias of $\boldsymbol{\hat{G}}$, as % of $\boldsymbol{G}$, for 1,000 MC samples, $\boldsymbol{K}$ =1,000: Lognormal distribution for the household effect variable.



a) DGP2B ($\rho_1 = \rho_2 = 0.1$)

b) DGP2G ($\rho_1 = 0.05,\ \rho_2 = 0.1$)



c) DGP2H ($\rho_1 = \cdots = \rho_5 = 0.1$)

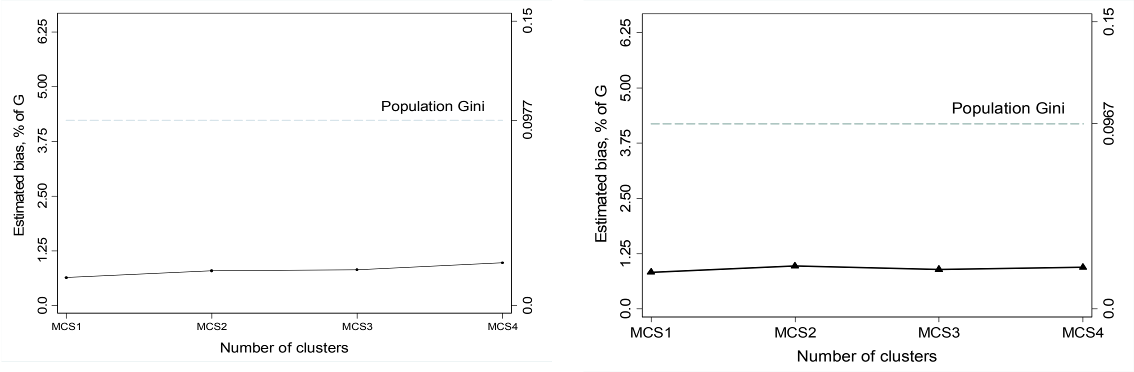
d)DGP2H ($\rho_1 = .05, \rho_2 = .2, \rho_3 = .1, \rho_4 = .3, \rho_5 = .15$)

Notes: For DGP2B and DGP2G, MC samples are MCS1: $n_1 = 10, n_2 = 15$; MCS2: $n_1 = 25, n_2 = 25$; MCS3: $n_1 = 40, n_2 = 50$; MCS4: $n_1 = 75, n_2 = 90$. For DGP2H and DGP2I, MC samples are clusters MCS1: $n_1 = 3, n_2 = 5$, n_3=6, n_4=8, n_5=9; MCS2: $n_1 = 7, n_2 = 9$, n_3=12, n_4=15, n_5=20; MCS3: $n_1 = 12, n_2$ =20, n_3=24, n_4=32, n_5=36; MCS4: $n_1 = 25, n_2 = 35$, n_3=45, n_4=64, n_5=70.

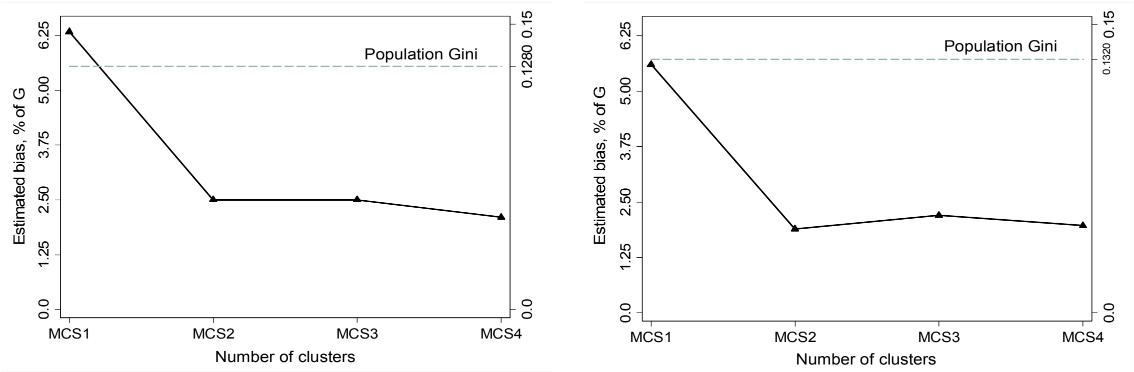
Unlike the beta and lognormal distribution, when data is generated from a chi-square distribution, a negligible level of bias (less than 1% of G) seems to be non-diminishing in the number of clusters in samples drawn from DGPs with two strata (see

Figure 5, a-b). This behaviour of bias may be because the means of empirical distributions of $\hat{G}$ for 1,000 MC samples of all four sizes from DGP3B and DGP3G are almost identical, and the distributions are highly symmetrical (see Figures C.4 - C.5 in Appendix C).

Figure 5 Estimated bias of $\hat{G}$, as % of G, for 1,000 MC samples, K =1,000: Chi-square distribution for the household effect variable.



a) DGP3B ($\rho_1 = \rho_2 = 0.1$)   b) DGP3G ($\rho_1 = 0.05, \rho_2 = 0.1$)



c) DGP3H ($\rho_1 = \cdots = \rho_5 = 0.1$)   d) DGP3I ($\rho_1 = .05, \rho_2 = .2, \rho_3 = .1, \rho_4 = .3, \rho_5 = .15$)

Notes: For DGP3B and DGP3G, MC samples are MCS1: $n_1 = 10, n_2 = 15$; MCS2: $n_1 = 25, n_2 = 25$; MCS3: $n_1 = 40, n_2 = 50$; MCS4: $n_1 = 75, n_2 = 90$. For DGP3I and DGP3J, MC samples are clusters MCS1: $n_1 = 3, n_2 = 5$, n_3=6, n_4=8, n_5=9; MCS2: $n_1 = 7, n_2 = 9$, n_3=12, n_4=15, n_5=20; MCS3: $n_1 = 12, n_2$ =20, n_3=24, n_4=32, n_5=36; MCS4: $n_1 = 25, n_2 = 35$, n_3=45, n_4=64, n_5=70.

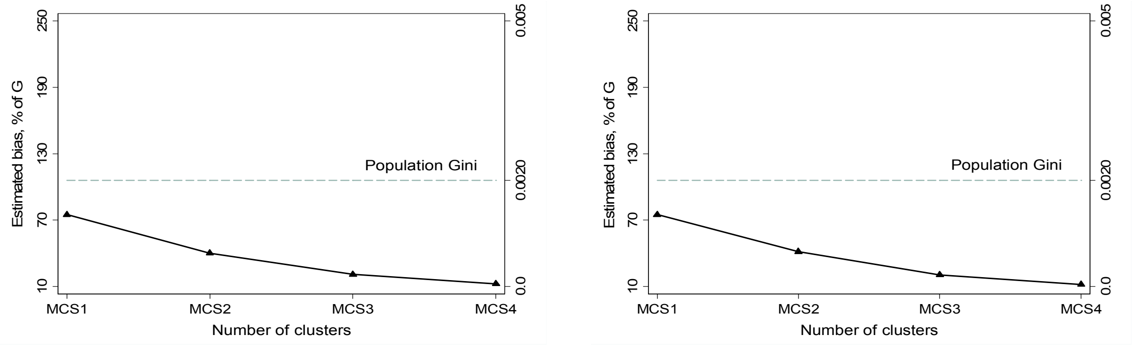
When a DGP contains more strata, however, the estimators are biased in small cluster samples under the chi-square distributions. For instance, for the sample MCS1 (containing 3% of the population clusters per stratum), the biases of $\hat{G}$ are 6.3% and 5.5% of G, when population data are generated with five strata and with a fixed 0.1 ICC and varying ICCs across strata under this distribution. However, this bias falls quickly by over

60% when we double the number of clusters in the sample. In samples with just over 6% of the population clusters (MCS2), the bias reduces to 2.5% and 1.9%. Nevertheless, the bias in samples with more clusters do not change much thereafter. This finding, along with results from the beta and lognormal distributions, suggests that, to achieve $\widehat{G}$ with tolerable bias; e.g., 2-3% of G, a survey sample should be comprised of strata having at least over 5% of the population clusters from each stratum, when the data exhibits skewness in the distribution.

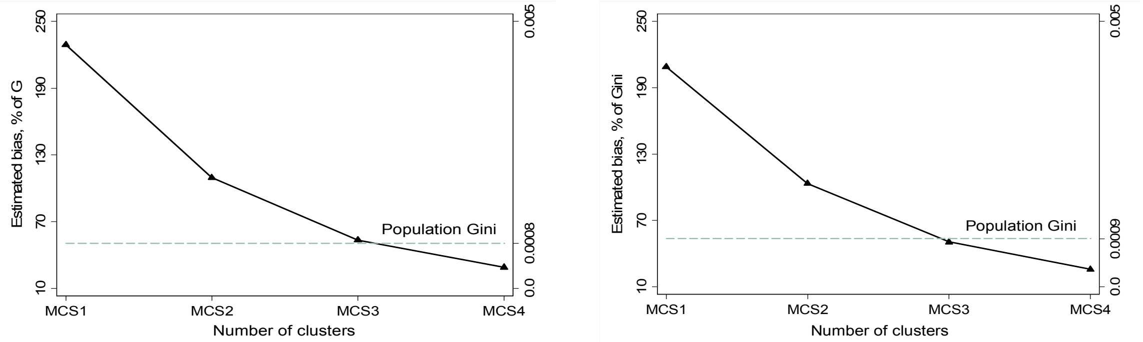
Finally, as shown in Figure 6, although the bias of $\widehat{G}$ rapidly falls when the number of clusters in the sample increases, our biggest sample is not adequate to achieve a $\widehat{G}$ with low bias, when the household effect variable has the heavy-tailed Pareto distribution.

We also see that sampling from data generated by slightly changing the ICC values across strata; e.g., 0.0 to 0.1, under the Pareto distribution does not seem to play an important role in bias reduction. Large bias issues with small sample size is well acknowledged in the literature when data are generated under this heavy-tailed Pareto distribution, and much larger samples are recommended to achieve more accurate sampling properties for an estimator (see, e.g., Sudheed and Dewan, 2013; Davidson, 2009; Gastwirth et al., 2005). A sharp decline in the bias of $\widehat{G}$ (as seen from 210-245% to 15-20% of G) when the number of clusters increased (from 3% to 24% of the population clusters per stratum) in samples from DGPs with five strata, which suggests the scope of bias reduction available with more clusters in the sample.

Figure 6 Estimated bias of $\widehat{\boldsymbol{G}}$, as % of $\boldsymbol{G}$, for 1,000 MC samples, $\boldsymbol{K}$=1,000: the Pareto distribution for the household effect variable.



a) DGP4B ($\rho_1 = \rho_2 = 0.1$) b) DGP4B ($\rho_1 = 0.05,\ \rho_2 = 0.1$).



c) DGP4H ($\rho_1 = \cdots = \rho_5 = 0.1$) d)DGP4I ($\rho_1 = .05, \rho_2 = .2, \rho_3 = .1, \rho_4 = .3, \rho_5 = .15$)

Notes: For DGP4B and DGP4G, MC samples are MCS1: $n_1 = 10, n_2 = 15$; MCS2: $n_1 = 25, n_2 = 25$; MCS3: $n_1 = 40, n_2 = 50$; MCS4: $n_1 = 75, n_2 = 90$. For DGP4H and DGP4I, MC samples are clusters MCS1: $n_1 = 3, n_2 = 5$, n_3=6, n_4=8, n_5=9; MCS2: $n_1 = 7, n_2 = 9$, n_3=12, n_4=15, n_5=20; MCS3: $n_1 = 12, n_2$ =20, n_3=24, n_4=32, n_5=36; MCS4: $n_1 = 25, n_2 = 35$, n_3=45, n_4=64, n_5=70.

Overall, the estimator $\widehat{G}$ is biased in samples with a small number of clusters, and the bias varies substantially across distributions with a much larger estimated bias under the Pareto distribution relative to the beta, lognormal, and chi-square distributions. However, the estimator is consistent, so asymptotically it is unbiased, irrespective of the underlying population DGP. We observe this feature when the number of clusters in the sample (when examining the not highly skewed distribution for the population data) increases. In some cases, particularly for the Pareto distribution, it would seem that the number of clusters in our samples was not large enough to illustrate the elimination of the bias. In addition, a slight variation in the ICC values across strata within a DGP did not appear to significantly affect the observed bias of $\widehat{G}$, compared to a fixed ICC value across strata.

3.4.2 Empirical Coverage Probabilities of 95% Confidence Intervals for G

In this subsection, we report simulation results for ECPs of interval estimators for G using data obtained from our four probability distributions and various sampling designs. Two types of experiments are conducted to evaluate the performance of an interval estimator:

1. To examine the relationship between the performance of an interval estimator for G and the ICC value in the data, we calculate ECPs of 95% CIs using samples with a specific sampling design from DGPs that are created by changing only the ICC value for each of the distributions. These experiments are referred to as the *base case.*
2. The coverage of CIs for G are estimated with an increasing number of clusters in the sample under different sampling designs and DGPs. We examine how quickly an ECP of a 95% CI estimator approaches the nominal level of confidence as the number of clusters in a sample increases.

3.4.2.1 Base Case Analysis: Sensitivity to the Intracluster Correlation Coefficient

The ICC is a key element of complex survey data. When a finite population with survey features is generated using the random effect model in expression (3.13), for a given distribution of household effect variable, the ICC determines the spread of the normally distributed cluster effect random variable and the well-being variable. Therefore, features and performance of an interval estimator for G are likely to be affected when samples are drawn from DGPs with different ICCs.

Before discussing simulation results on ECPs, we present sample variances of $\hat{G}$ using the formulae of Bhattacharya (2007), Binder and Kovačević (1995) and the standard bootstrap method in expressions (3.2), (3.3), and (3.7), respectively, along with the 95% CIs for G from a single MC sample drawn from the base case DGPs in Table 6. In addition, the 95% CI for G using the bootstrap MC percentile method is also reported.

Table 6 Estimated variances of $\widehat{\boldsymbol{G}}$, 95% CIs and widths for the Gini coefficient using a single MC sample and 99 bootstrap resamples.

DGP	G	$\hat{G}$	BTAM			BKAM			SBSM			SBPM	
			$\widetilde{Var_{BT}}(\hat{G})$	95% CI	Width	$\widehat{Var_{BK}}(\hat{G})$	95% CI	Width	$\widehat{Var_{BOOT}}(\hat{G}_b^*)$	95% CI	Width	95% CI	Width
						Beta distribution for z_{hci}							
DGP1A	0.0381	0.0382	0.167	0.0374,0.0390	0.0016	0.171	0.0374,0.0390	0.0016	0.128	0.0375,0.0389	0.0014	0.0372,0.038	0.0014
DGP1B	0.0404	0.0413	0.501	0.0399,0.0426	0.0027	0.513	0.0398,0.0427	0.0029	0.513	0.0398,0.0426	0.0028	0.0394,0.042	0.0030
DGP1C	0.0461	0.0469	1.740	0.0443,0.0494	0.0051	1.780	0.0442,0.0495	0.0053	1.590	0.0444,0.0493	0.0049	0.0440,0.048	0.0049
DGP1D	0.0548	0.0555	4.951	0.0512,0.0599	0.0087	5.066	0.0511,0.0600	0.0089	4.610	0.0513,0.0597	0.0084	0.0516,0.059	0.0078
DGP1E	0.0708	0.0720	14.500	0.0646,0.0795	0.0149	14.840	0.0645,0.0796	0.0151	14.070	0.0647,0.0794	0.0147	0.0654,0.079	0.0137
DGP1F	0.1226	0.1266	72.720	0.1099,0.1433	0.0334	74.430	0.1097,0.1435	0.0338	73.730	0.1098,0.1435	0.0337	0.1129,0.144	0.0313
						Lognormal distribution for z_{hci}							
DGP2A	0.0498	0.0524	0.936	0.0505,0.0543	0.0038	0.959	0.0505,0.0543	0.0038	0.814	0.0506,0.0541	0.0035	0.0504,0.053	0.0033
DGP2B	0.0542	0.0531	1.459	0.0508,0.0555	0.0047	1.494	0.0507,0.0555	0.0048	1.424	0.0508,0.0555	0.0047	0.0508,0.055	0.0044
DGP2C	0.0636	0.0626	4.329	0.0585,0.0667	0.0082	4.434	0.0585,0.0667	0.0082	4.261	0.0586,0.0666	0.0080	0.0585,0.066	0.0084
DGP2D	0.0872	0.0862	17.520	0.0780,0.0944	0.0164	17.920	0.0780,0.0945	0.0165	18.480	0.0778,0.0947	0.0169	0.0783,0.095	0.0170
DGP2E	0.1013	0.1023	33.850	0.0909,0.1137	0.0228	34.670	0.0907,0.1138	0.0231	33.680	0.0909,0.1136	0.0227	0.0909,0.114	0.0234
DGP2F	0.1756	0.1851	182.910	0.1586,0.2116	0.0530	187.360	0.1583,0.2120	0.0537	182.620	0.1587,0.2116	0.0529	0.1613,0.211	0.0504
						Chi-square distribution for z_{hci}							
DGP3A	0.0922	0.0926	1.263	0.0904,0.0948	0.0044	1.290	0.0904,0.0948	0.0044	1.074	0.0906,0.0946	0.0040	0.0902,0.094	0.0040
DGP3B	0.0976	0.0983	4.766	0.0940,0.1026	0.0086	4.873	0.0940,0.1026	0.0086	4.872	0.0940,0.1026	0.0086	0.0938,0.103	0.0096
DGP3C	0.1091	0.1090	13.750	0.1017,0.1162	0.0145	14.060	0.1016,0.1163	0.0147	13.660	0.1017,0.1162	0.0145	0.1012,0.117	0.0161
DGP3D	0.1262	0.1260	30.950	0.1151,0.1369	0.0218	31.660	0.1149,0.1368	0.0219	30.710	0.1151,0.1368	0.0217	0.1161,0.137	0.0216
DGP3E	0.1578	0.1592	77.790	0.1419,0.1764	0.0345	79.560	0.1418,0.1766	0.0348	78.950	0.1418,0.1766	0.0348	0.1449,0.176	0.0318
						Pareto distribution for z_{hci}							
DGP4A	0.0020	0.0026	0.062	0.0021,0.003	0.0010	0.063	0.0021,0.0031	0.0010	0.065	0.0021,0.003	0.0010	0.0018,0.00	0.0010
DGP4B	0.0021	0.0027	0.081	0.0021,0.003	0.0012	0.083	0.0021,0.0033	0.0012	0.087	0.0021,0.003	0.0012	0.0017,0.00	0.0011
DGP4C	0.0022	0.0028	0.086	0.0023,0.003	0.0011	0.088	0.0023,0.0034	0.0011	0.092	0.0022,0.003	0.0011	0.0019,0.00	0.0011
DGP4D	0.0024	0.0030	0.093	0.0024,0.003	0.0012	0.095	0.0024,0.0036	0.0012	0.098	0.0024,0.003	0.0012	0.0021,0.00	0.0011
DGP4E	0.0027	0.0034	0.106	0.0027,0.004	0.0013	0.108	0.0027,0.0040	0.0013	0.110	0.0027,0.004	0.0013	0.0023,0.00	0.0012
DGP4F	0.0038	0.0044	0.154	0.0037,0.005	0.0015	0.158	0.0037,0.0052	0.0015	0.157	0.0037,0.005	0.0015	0.0033,0.00	0.0015

Notes: Variance estimates have been scaled by 10^6. $\widetilde{Var_{BT}}(\hat{G})$ and $\widehat{Var_{BK}}(\hat{G})$ are Bhattacharya's (2007) and Binder and Kovačević 's (1995) estimators, respectively. $\widehat{Var_{BOOT}}(\hat{G}_b^*)$ is the bootstrap variances estimator using expression (3.7) for 99 replications.

As discussed earlier, the base case consists of samples from six DGPs with two strata under each of the probability distributions for the household effect variable. The ICC values assumed to determine variances in the normally distributed cluster effect variable in those DGPs are 0.0, 0.1, 0.3, 0.5, 0.7 and 0.9. We exclude the DGP under the chi-square distribution with ICC equal to 0.9 (DGP3F), as some household values are negative, resulting in a total of 23 DGPs in the base case. A single MC sample with 10% of the population clusters from each stratum (n_1=40 and n_2=50) is drawn from each DGP with a common sampling design.

The larger the value of ICC within strata, the higher $\hat{G}$ and its variance estimators. In the DGP, when the ICC is increased, keeping other survey features the same (e.g., the number of strata, clusters and households, the distribution of z_{hci}), both the variances of the normally distributed cluster effect variable and the well-being variable ($\psi_h^2 = \gamma_h^2 + \sigma_h^2$) rise. This may result in larger sampling variances of $\hat{G}$.

The variance estimators for $\hat{G}$ using the formulae provided by Bhattacharya (2007) and Binder and Kovačević (1995) are very similar across DGPs, and the small difference between them is due to the correction factor $n_h/(n_h - 1)$ associated with the latter formula. For more sampled clusters across strata in a sample the difference turns out to be negligible. The conventional bootstrap variance estimator of $\hat{G}$ based on 99 replications leads to a slightly smaller estimate than the two asymptotic variance estimates. As such, the 95% SN approximation CIs formed using Bhattacharya's (2007) and Binder and Kovačević's (1995) variance estimators, BTAM and BKAM methods, respectively, are somewhat wider than those for the bootstrap variance estimator (SBSM) (see Table 6). On the other hand, the width of the 95% CI formed using the bootstrap MC percentile method (SBPM) is similar to that of the SBSM method.

The extent of effects of the complex sampling design on estimation compared to the simple random sampling (SRS) is typically measured by the design effect. A sampling variance of $\hat{G}$ obtained under the complex survey design differs from that obtained from a SRS framework with the magnitude of the difference depending on the sampling features – stratification, clustering, weighting, etc. In addition, the underlying probability distribution of the data is a key factor, along with the ICC values across strata within the finite population. Skinner et al. (1989, p. 28-35), for example, discuss the implications of the design effect on the performance of CIs.

Following Kish (1965), the measure of efficiency of the complex survey design in estimating $\hat{G}$, the design effect, is defined as $DEFF(\hat{G}) = \frac{Var_{CS}(\hat{G})}{Var_{SRS}(\hat{G})}$, where $Var_{CS}(\hat{G})$ and $Var_{SRS}(\hat{G})$ are variances of $\hat{G}$ under a complex sampling (CS) design and a simple random sampling design with weights (SRS), respectively. For $DEFF(\hat{G}) > 1$, the CS design is less efficient than the SRS, whereas when $DEFF(\hat{G}) < 1$, the CS design is more efficient than an SRS design; when $DEFF(\hat{G}) = 1$, the CS and SRS designs are equally efficient.

For a statistic of interest obtained from a complex survey sample, Park et al. (2004, 2003) derive a formula for $DEFF$ in terms of different survey components – stratification, clustering, intracluster correlation coefficient, and weights. Further elaboration on the design effect is beyond the scope of our study. We estimate $DEFF(\hat{G})$ from a single MC sample, using Bhattacharya's (2007) variance estimator of $\hat{G}$ formula in expression (3.2), to examine the effects of the design effect on the performance of interval estimators for G. This design effect is:

$$\text{deff}\left(\hat{G}\right) = \frac{\widetilde{Var_{BT}}\left(\hat{G}\right)}{\widetilde{Var_{SRS}}\left(\hat{G}\right)}, \tag{3.15}$$

where $\widetilde{Var_{BT}}\left(\hat{G}\right)$ is Bhattacharya’s (2007) variance estimator for $\hat{G}$ under the complex survey design, $\widetilde{Var_{SRS}}\left(\hat{G}\right)$ is the variance estimator of $\hat{G}$ under SRS with weights (first component on the left side of expression (3.2)). We report estimated design effects for the base case in Table 7.

When the ICC is zero across strata in the population, households within clusters are no different from households across clusters; e.g., DGP1A. Then, the design effect is less than one for each of the four considered DGPs. A design effect less than one implies that the complex survey sampling is more efficient than SRS, which makes sense here because the stratification effect can be larger than the cluster effect according to expression (3.2). Recall, the stratification reduces the sampling variance of $\hat{G}$ as households within a stratum are homogeneous. For ICC=0, we also observe some negative cluster effects in Table 7. As the cluster effect is the contribution of clustering to the complex sample variance of $\hat{G}$, the second component in expression (3.2), which is essentially a covariance expression, can be negative. When the covariances between values obtained from the same cluster are negative, that is greater values of some households correspond to the lesser values of other households within a cluster, the term can be negative.

Table 7 Estimated cluster effects, stratum effects and design effects using the variance formula of Bhattacharya (2007).

DGP	ICC $\rho_1 = \rho_2$	$\widetilde{Var_{BT}}(\hat{G})$	$\widetilde{Var_{SRS}}(\hat{G})$	Cluster effect	Stratum effect	Design effect
			Beta distribution			
DGP1A	0.0	0.167	0.254	-0.062	0.025	0.66
DGP1B	0.1	0.501	0.294	0.226	0.019	1.70
DGP1C	0.3	1.740	0.410	1.370	0.040	4.24
DGP1D	0.5	4.951	0.580	4.471	0.100	8.53
DGP1E	0.7	14.500	1.000	13.800	0.300	14.50
DGP1F	0.9	72.720	3.101	71.110	1.491	23.45
			Lognormal distribution			
DGP2A	0.0	0.936	0.984	0.169	0.217	0.95
DGP2B	0.1	1.459	0.880	0.800	0.221	1.66
DGP2C	0.3	4.329	1.030	3.910	0.611	4.20
DGP2D	0.5	17.520	1.501	16.110	0.091	11.67
DGP2E	0.7	33.850	2.202	35.403	3.755	15.37
DGP2F	0.9	182.910	7.994	191.917	17.001	22.88
			Chi-square distribution			
DGP3A	0.0	1.263	1.684	-0.230	0.191	0.75
DGP3B	0.1	4.766	2.081	3.110	0.425	2.29
DGP3C	0.3	13.750	2.491	11.760	0.501	5.52
DGP3D	0.5	30.950	3.194	28.401	0.645	9.69
DGP3E	0.7	77.790	4.999	73.500	0.709	15.56
			Pareto distribution			
DGP4A	0.0	0.062	0.069	0.023	0.030	0.90
DGP4B	0.1	0.081	0.086	0.023	0.028	0.94
DGP4C	0.3	0.086	0.086	0.027	0.027	1.00
DGP4D	0.5	0.093	0.086	0.033	0.026	1.08
DGP4E	0.7	0.106	0.086	0.046	0.026	1.23
DGP4F	0.9	0.154	0.087	0.090	0.023	1.77

Notes: Variance estimates have been scaled by 10^6. $\widetilde{Var_{SRS}}(\hat{G})$ is the first term of expression (3.2), the variance estimator under an assumption of SRS with weights. The "design effect" provides the ratio of $\widetilde{Var_{BT}}(\hat{G})$ to $\widetilde{Var_{SRS}}(\hat{G})$. The cluster and stratum effects are estimated using the second and third components, respectively, on the right side of expression (3.2).

Estimated design effects are greater than one in samples from DGPs with nonzero ICCs, and they increase with larger ICC values under the beta, lognormal, and chi-square distributions. As such, $\hat{G}$ is less efficient under the survey design for larger ICC values compared to that under the SRS design. For the heavy-tailed Pareto distribution, the estimated design effects are somewhat different than what we observe with the other distributions. As reported in Table 7, for a small ICC value (less than 0.3), the stratum effect is higher in magnitude than the cluster effect, so that the complex sampling design appears to be more efficient than the SRS method, which is rather unusual. For a large value of ICC; e.g., 0.50, when the design effect is around 10 under other distributions, it is barely greater than one in the case of the Pareto distribution. It seems that, unlike other distributions, a larger ICC value does not reduce the variability among households within clusters in such a skewed distribution.

The simulation results on ECPs of 95% CIs for G for the base cases are reported in Table 8. As expected, the performance of the two SN approximation intervals using the analytical asymptotic variances, BTAM and BKAM, are highly comparable. For samples of 90 clusters (10% of the population clusters) drawn from DGPs with zero ICC value, ECPs of the two 95% SN approximation CIs are on average 81.5%, 88.5%, 93.5%, and 51% under the beta, lognormal, chi-square, and Pareto distributions, respectively. The SN approximation interval using the bootstrap variance (SBSM) is not always more efficient than the two other SN approximation methods, except for the Pareto distribution. Under the highly skewed Pareto distribution, the ECP of this bootstrap interval estimator is almost 18% higher than that of the BTAM and BKAM. On the other hand, the bootstrap MC percentile CIs (SBPM) performs better than the SN approximation interval estimators for samples with few clusters under the beta and lognormal distributions and the heavy-tailed Pareto distribution. Our findings accord with Langel and Tillé (2013) and Davidson (2009) who report that the variance approximation based inference is very good if the tail of the underlying distribution is not heavily skewed.

Table 8 Empirical coverage probabilities (CP%), lower (L%), and upper (U%) tail coverage rates of the five 95% CIs for the Gini coefficient.

DGP	BTAM			BKAM			SBSM			SBPM			WSPM		
	CP	L	U	CP	L	U	CP	L	U	CP	L	U	CP	L	U
							Beta distribution for z_{hci}								
DGP1A	81.70	18.00	0.30	83.00	16.80	0.20	83.10	16.70	0.20	92.90	4.90	2.20	93.70	4.10	2.20
DGP1B	88.60	11.10	0.30	89.10	10.60	0.30	88.20	11.50	0.30	94.00	3.90	2.10	94.70	4.50	0.80
DGP1C	93.30	4.20	2.50	93.70	3.90	2.40	93.60	3.90	2.50	92.50	2.20	5.30	94.20	3.50	2.30
DGP1D	94.40	2.10	3.50	94.80	1.90	3.30	93.70	2.40	3.90	91.10	2.30	6.60	94.90	3.00	2.10
DGP1E	93.70	1.70	4.60	94.00	1.70	4.30	93.20	2.10	4.70	92.00	1.70	6.30	95.40	2.20	2.40
DGP1F	92.50	1.20	6.30	92.10	1.10	5.80	92.00	1.30	6.70	91.10	1.60	7.30	95.40	1.70	2.90
							Lognormal distribution for z_{hci}								
DGP2A	88.30	11.20	0.50	88.90	10.70	0.40	87.90	11.30	0.80	92.90	5.70	1.40	95.10	3.70	1.20
DGP2B	90.90	6.30	2.80	91.40	5.90	2.70	91.60	5.70	2.70	92.00	3.40	4.60	93.00	3.70	3.30
DGP2C	92.70	2.60	4.70	92.80	2.60	4.60	93.60	2.20	4.20	92.20	1.90	5.90	94.50	2.60	2.90
DGP2D	93.10	1.60	5.30	93.20	1.60	5.20	92.60	1.90	5.50	92.10	1.80	6.10	95.00	2.10	2.90
DGP2E	93.10	1.60	5.30	93.60	1.20	5.20	92.90	1.40	5.70	91.50	2.10	6.40	95.10	2.10	2.80
DGP2F	92.10	0.70	7.20	92.20	0.70	7.10	91.70	0.90	7.40	90.80	1.60	7.60	95.10	1.80	3.10
							Chi-square distribution for z_{hci}								
DGP3A	93.30	5.10	1.60	93.90	4.80	1.30	93.50	4.80	1.70	93.40	2.90	3.70	95.70	1.30	3.00
DGP3B	93.20	5.00	1.80	93.70	4.60	1.70	93.00	4.80	2.20	93.30	3.70	3.00	93.90	4.00	2.10
DGP3C	93.00	3.50	3.50	93.10	3.50	3.40	92.70	3.70	3.60	91.40	3.80	4.80	94.40	3.40	2.20
DGP3D	92.60	2.70	4.70	93.10	2.50	4.40	92.70	2.30	5.00	91.30	2.50	6.20	93.90	3.00	3.10
DGP3E	92.70	1.70	5.60	93.20	1.40	5.40	92.40	1.70	5.90	91.40	2.10	6.50	94.60	2.10	3.30
							Pareto distribution for z_{hci}								
DGP4A	50.60	49.40	0.00	51.40	48.60	0.00	60.10	39.90	0.00	84.20	0.10	15.70	96.30	1.80	1.90
DGP4B	46.50	53.50	0.00	47.70	52.30	0.00	54.40	45.60	0.00	85.90	0.10	14.00	96.70	0.80	2.50
DGP4C	48.50	51.50	0.00	50.30	49.70	0.00	57.60	42.10	0.00	85.40	0.10	14.50	96.30	0.70	3.00
DGP4D	53.00	47.00	0.00	54.70	45.30	0.00	60.90	39.10	0.00	84.80	0.10	15.10	96.00	0.60	3.40
DGP4E	61.10	38.90	0.00	62.90	37.10	0.00	67.50	32.50	0.00	85.80	0.10	14.10	95.90	0.80	3.30
DGP4F	78.30	21.50	0.20	79.50	20.30	0.20	80.00	19.80	0.20	88.80	0.70	10.50	95.80	0.90	3.30

As seen in Table 6, the sample variance of $\hat{G}$ increases for larger values of the ICC parameter, and the corresponding 95% CIs for G using any of the SN approximation methods become wider. A wider interval usually leads to a higher coverage probability. Nevertheless, we do not see a linear positive relationship between the ICC value and the performance of an interval estimator for G.

For samples from the beta and lognormal distributions, ECPs of the SN approximation intervals improve with increasing ICC values up to approximately one half, after which they start to fall. The coverages of these interval estimators seems to be non-responsive to the ICC parameter under the chi-square distribution. The ECPs, however, increases for samples from the DGP with a larger ICC value created from the Pareto distribution. These results seem to be related to the skewness of the underlying distribution of the population data.

As noted earlier, when the ICC value in the DGP is increased, the cluster effect distribution dominates the household effect distribution, and a positively skewed distribution of the well-being variable y_{hci} becomes more symmetrical under the beta, lognormal, and chi-square distributional assumptions. Nevertheless, for a large value of ICC; e.g., 0.7 or 0.9, the distribution of y_{hci} converts into a negatively skewed distribution (see the illustration with the lognormal distribution in Figure C.2, in Appendix C). Such a phenomenon may worsen the performance of the SN interval estimator for larger ICCs. A much lower ECPs of the SN approximation CIs the Pareto distribution is considered to be due to the *heavy-tailed* nature of the distribution; see, e.g., Langel and Tillé (2011), and Gastwarth et al. (2005).

In addition, we recall that the estimator $\hat{G}$ is biased, often severely, when the data are obtained from the Pareto distribution, although the bias falls with the ICC value (see Table 5). When an estimator is highly biased, not surprisingly, the performance of a SN approximation interval estimator can be low. These findings with the Pareto distribution are not unusual, and are often reported in the literature even under an *iid* sampling framework (see, e.g., Peng, 2011; Davidson, 2009; Gastwarth et al., 2005). Nevertheless, unlike the SN approximation for CIs, the performance of the SBPM method does not exhibit any resilient dependency on the ICC value. This is because the SBPM CIs are formed using the empirical distribution of $\hat{G}$ only, rather than using a standard error.

Our MC bootstrap percentile method (SBPM) is akin to the double-bootstrap method. One of the purposes of a double-bootstrap method is to improve the coverage accuracy of CIs by reducing the coverage error; see, e.g., Chang and Hall (2015). Recall that exploring the sampling properties using such a double bootstrap method is computationally intensive, which led Giacomini et al. (2013) to propose a time saving, so called warp-speed MC method (WSPM) to estimate ECPs in a MC simulation study.

For our base cases, on average, the WSPM method produces ECPs around the 95% nominal level regardless of the distribution and ICC value of the data. On the other hand, ECPs of the WSPM method are higher than those of the SBPM method, and the difference between the ECPs are noticeable, especially under the Pareto distribution. This result may be because the SBPM interval is based on the empirical distribution of bootstrap $\widehat{G}$, and in the WSPM method each bootstrap resample is drawn from a different empirical distribution (see, e.g., Giacomini et al., 2013). As such, we do not think that the warp-speed MC method can always assess the properties of a double bootstrap method to make inferences about G, at least when data are obtained using complex survey designs.

We also report the simulated values of the lower tail and upper tail coverage probabilities (in %) in Table 8. When the tail coverage probabilities are not similar, the sampling distribution of $\hat{G}$ is skewed. For a positively (negatively) skewed distribution the lower (upper) tail coverage rate is higher. The estimated tail probabilities are said to be symmetric or balanced, when the larger tail probability is less than 1.5 times the smaller one (see, e.g., Arasan and Adam 2014). In our base case simulations, tail CPs are not balanced for the interval estimators across our four studied distributions, which likely follow from the skewness in the distribution of data; see, e.g., Qin et al. (2010).

The three SN approximation estimators (BTAM, BKAM, and SBSM) are comparable when evaluated using the tail ECPs, except for that under the Pareto distribution. The lower tail coverage rate is higher than that of the upper tail for a zero ICC value for these three methods, implying that the empirical distribution of $\hat{G}$ is positively skewed. Nevertheless, when the ICC value increases, the lower tail coverage rate decreases accompanied by an increasing upper tail coverage rate, indicating that the empirical distribution of $\hat{G}$ becomes more symmetrical with an increasing ICC value. For large ICCs; however, for example, 0.7 and 0.9, the distribution becomes negatively skewed, which we also observed earlier with the underlying distributions of data. In the case of the Pareto distribution, although the lower tail error rate falls with the ICC parameter, no upper tail error coverage occurs. This is due to the heavy-tailed nature of the distribution. We also observe that for such a distribution, the SBSM method performs relatively better with a smaller total tail coverage rate.

In terms of the tail ECP, the SBPM performs better than the three SN approximation interval estimators with more balanced tail error rates under the beta, lognormal, and chi-square distributions. This may be because the CI using the SBPM method uses the resampling distribution of $\hat{G}$ rather than the standard errors, producing more balanced tail ECPs; see, e.g., Wu and Rao (2006). The tail ECPs of the WSPM method are not balanced, and we do not find them similar to the SBPM interval estimators either, again suggesting that the warp-speed method is not adequate at representing the MC performance of the double-bootstrap for our cases.

3.4.2.2 Sensitivity of the Performance of Interval Estimators to the Sample Size

We now examine the performance of 95% CI estimators for G with increasing sample sizes, measured by the number of clusters in the sample. Samples are drawn from four DGPs, two DGPs with two strata and two DGPs with five strata, under each of the four probability distributions (beta, lognormal, chi-square, and the Pareto). In addition, given the difficulties of the estimators to perform well when data are from the heavy-tailed Pareto distribution, we consider an additional DGP containing three strata, resulting in a total of 17 DGPs for the simulation experiments. The examined DGPs are DGP1B-4B, DGP1G-4G, DGP1H-4H, DGP1I-4I, and DGP4J.

Four samples, MCS1-4, of different combinations of clusters across strata are assumed from each of the aforementioned DGPs (see the descriptions in Section 3.3). For the additional experiment with DGP4J from the Pareto distribution, the results of the 95% CI estimators are examined with five samples MCS1-5. The first sample, MCS1, is selected with 5% of the population clusters from each stratum, resulting in a total of 38 clusters in the sample ($n_1 = 10$, $n_2 = 13$, $n_3 = 15$), with a fixed 30 households per cluster. The successive four samples contain 10% ($n_1 = 20$, $n_2 = 25$, $n_3 = 30$), 20% ($n_1 = 40$, $n_2 = 50$, $n_3 = 60$), 30% ($n_1 = 60$, $n_2 = 75$, $n_3 = 90$), and 40% ($n_1 = 80$, $n_2 = 100$, $n_3 = 120$) of the population clusters from each stratum.

The performance of 95% CIs for G are evaluated under two cases: a) for samples drawn from DGPs with a fixed 0.1 ICC across strata; and b) for samples drawn from DGPs with varying ICCs across strata. For each case, four interval estimators, BTAM, BKAM, SBSM, and SBPM, are examined; ECPs are shown in Figures 7 and 8. In addition, to assess the outcome of the SBPM estimators, we also report ECPs of interval estimators using the warp-speed method for each case.

a) Empirical Coverage probabilities of 95% CIs for G with the increasing number of clusters drawn from DGPs with ICC=0.1.

The coverage probabilities of 95% CIs for G, when samples are drawn by increasing the number of clusters from DGPs with a constant ICC equal to 0.1 across strata generated under the beta, lognormal, chi-square, and Pareto distributions, are illustrated in Figure 7. The results suggest that the level of performance of an interval estimator is highly distribution specific, especially for samples containing fewer clusters. For instance, in a small[23] sample containing approximately 3% of the population clusters (MCS1), ECPs of the three SN approximation and the bootstrap MC percentile interval estimators are 75.3-78.1% and 92.7%, respectively, when the sample is drawn from DGP1B created under the beta distribution. Whereas, these probabilities are 88.7-89.7% and 89%, when the sample MCS1 is obtained from the DGP2B with the lognormal distribution, and they are less than 5% for the sample MCS1 obtained from the Pareto distribution (DGP4B). The coverage of a SN approximation estimator deteriorates with more skewed distributions of the well-being variable.

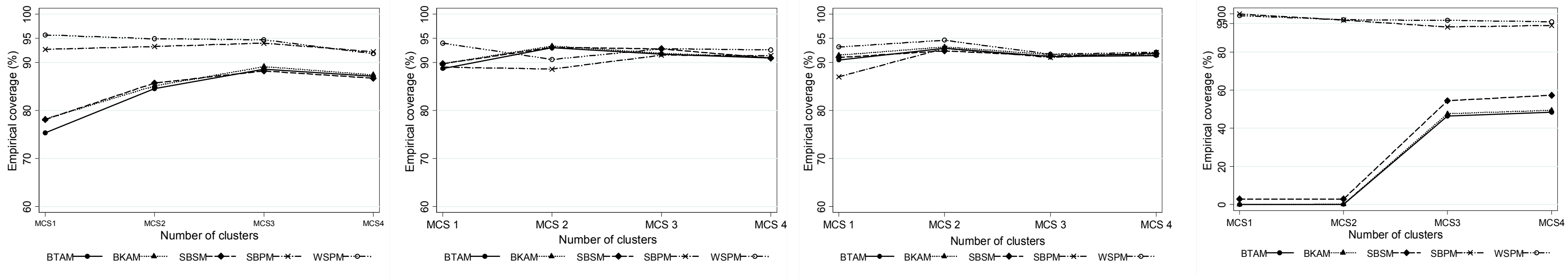
On the other hand, we see that an interval estimator does not always perform better when more clusters are included in the sample. For example, the ECP of BKAM interval estimator increased from 83.2% to 85.90% when the number of clusters in the sample increased from 63 (MCS2) to 241 (MCS4) drawn from the DGP with five strata and lognormal distribution (DGP2H). When the samples were obtained from data generated with the chi-square distributional assumption (DGP3H), the coverage decreased from 81.2% to 70.6%. In the case of the Pareto distribution (DGP4H), none of the interval estimators responded to the number of clusters in the sample.

---

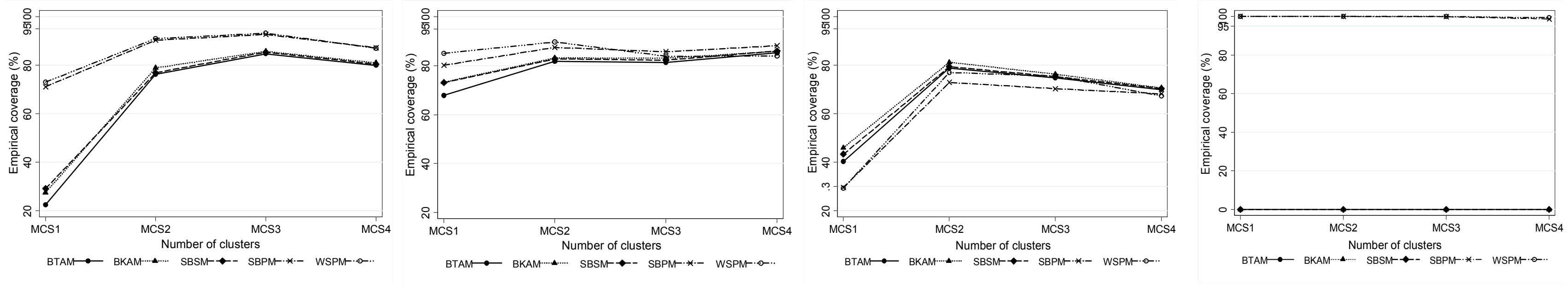
[23] In our simulations, a small sample refers to the sample containing 3-5% of the population clusters.

Figure 7 Empirical coverage probabilities of the 95% CI estimators for G: samples with disproportional increase in the number of clusters from DPGs with ICC equals 0.1 across strata.

Beta distribution	Lognormal distribution	Chi-square distribution	The Pareto distribution



a) DGP1B b) DGP2B c) DGP3B d) DGP4B



e) DGP1H f) DGP2H g) DGP3H h) DGP4H

Notes: First row shows results for DGPs with two strata. Samples drawn from these DGPs are MCS1: MC sample 1 with n_1=10, n_2=15; MCS2: MC sample 2 with n_1=25, n_2=25; MCS3: MC sample 3 with n_1=40, n_2=50; and MCS4: MC sample 4 with n_1=75, n_2=90. Second row shows results for DGPs with five strata. Samples drawn from these DGPs are MCS1: MC sample 1 with n_1=3, n_2=5, of n_3=6, n_4=8, n_5=9; MCS2: MC sample 2 with n_1=6, n_2=10, n_3=12, n_4=15, n_5=20; MCS3: MC sample 3 with n_1=12, n_2=20, n_3=24, n_4=32, n_5=36; and MCS4: MC sample 4 with n_1=25, n_2=35, n_3=45, n_4=64, n_5=70. BTAM: SN approximation CI using Bhattacharya's (2007) standard error, BKAM: SN approximation CI using Binder and Kovačević's (1995) standard error, SBSM: SN approximation CI using bootstrap standard error, SBPM: standard MC bootstrap percentile CI and WSPM: warp-speed MC percentile CI.

The ECPs of BTAM and BKAM estimators using asymptotic variances, as before, are highly comparable to each other. For example, the ECPs of the BTAM and BKAM interval estimators are 90.1% and 90.8%, respectively, for the sample MCS3 from DGP1I, and 88.7% and 89.7% for the sample MCS1 drawn from DGP2B. The other SN approximation interval estimator using the bootstrap variance, SBSM, is not always more efficient than the BKAM and BTAM estimators, except for some small samples, as expected. For instance, the ECP of the SBSM is 3-7% higher than that of BTAM and BKAM estimators for the sample MCS1 (with less than 3% of the population clusters), whereas it is slightly lower (about 1%) in the larger sample MCS4 drawn from DGP1G.

With some exceptions, the MC bootstrap percentile interval estimators (SBPM) performed quite well, being close to the nominal level, and in some cases with ECPs higher than the assigned nominal level. The latter result typically arises when the samples are obtained from the highly skewed Pareto distribution. More often, the coverage of the interval estimator is similar across changes in the number of strata in the DGP and the number of sampled clusters under the distribution.

The SBPM estimators also behave differently than the three SN approximation interval estimators. For samples from finite populations under the beta and Pareto distribution, the ECP of a SBPM interval estimator is substantially higher than the ECPs from the SN approximation interval estimators across samples MCS1 to MCS4. For DGPs under the lognormal and chi-square distributions, ECPs of SBPM interval estimators in the small sample MCS1 are slightly different than from the SN approximation interval estimators, but this difference becomes negligible in MCS4, samples with 18-25% of the population clusters. In any case, the performance of the SBPM interval estimator does not improve noticeably by increasing the number of sampled clusters in samples containing more than 5% of the population clusters. We notice that when the cluster sample size increases after the sample MCS2, the lower tail coverage rate of SBPM increases, implying a shift in the sampling distribution of $\hat{G}$ to the left when the sample contains more clusters. Overall, a double bootstrap method (SBPM) seems to be a good choice for small cluster samples, and for data obtained from DGPs with more strata or when the underlying data distribution is highly skewed.

Empirical coverage probabilities of the WSPM interval estimators are around, and often above, the 95% nominal level, irrespective of the number of clusters in the sample across DGPs. These ECPs are 1-7% higher than those of the SBPM interval estimators for samples containing 3-12% of the population clusters (MCS1-MCS3), except for when data are from the Pareto distribution, where both methods perform similarly with the coverage rates being above the nominal level. Nevertheless, the difference between the ECPs of the two percentile interval estimators becomes negligible (less than 1%) when samples contain more clusters (20-24% of the population clusters; MCS4). In such cases, the WSPM method seems to better reflect the performance of the SBPM interval estimators than in previously discussed cases.

The outcome of a 95% CI for G in samples drawn from DGPs with five strata differs largely from the samples from DGPs with two strata (see Figure 7). We see that ECPs of the three SN approximation interval estimators are substantially lower than those for samples from DGPs with two strata across distributions, and the difference is much higher for samples with fewer clusters. For example, the coverages of the BTAM interval estimators are 84.5% and 75.8% for the sample MCS2 obtained from DGP1B and DGP1H, respectively. This may be because more strata in small samples reduces the sampling variance of $\hat{G}$ and the width of the CI, resulting in a lower ECP. In the case of the Pareto distribution, these interval estimators fail to cover G with any of the samples we considered.

For DGPs with five strata, when the number of clusters in a sample is increased from 31 to 63 (MCS1 to MCS2), the coverages of all five interval estimators rise markedly. After that, however, the rate of increase in the coverage for including more clusters in the sample is not noteworthy under the beta, lognormal, and chi-square distributions. In fact, for the beta and chi-square distributions, the ECP for the sample MCS4 is 5-7% lower than that for sample MCS3. In these cases, the lower tail ECPs of the interval estimators increase, although both the bias and estimated variance of $\hat{G}$ reduce with the rising number of clusters in the samples. This may be because of a more spiked and slightly left-shifted sampling distribution of $\hat{G}$ for MCS4 (see, e.g., Figure C.4-5 in Appendix C).

The SN approximation interval estimators do not perform well under the heavy-tailed Pareto distribution. For example, for DGPs with two strata, the ECP in samples with less than 6% of the population clusters is zero, and it fails to exceed 50% in our biggest sample MCS4. The high bias of $\hat{G}$, and the skewed shape of the data distribution are probable causes of such a low performance. The situation deteriorates further for the DGP with five strata, DGP4H, when these interval estimators fail to provide any coverage for any sample (see Figure 7h). This is likely arising for data from a highly skewed distribution that can cause difficulties in forming the SN approximation CIs for G. To provide interval estimators for G with reasonable coverage, a larger sample with more clusters may be needed, as is often suggested in the literature. For instance, under an *iid* framework, Gastwirth et al. (2005) show that to obtain a similar coverage probability of the 95% CI for G, the sample obtained from a Pareto distribution should be as large as 25 times than that from an exponential distribution.

On the other hand, ECPs of the SBPM interval estimators are around the 95% nominal level and approximately 100% level across samples from DGP4B and DGP4H, respectively, created under the Pareto distributions. For these DGPs, performance of the warp-speed MC percentile interval estimators are similar to the SBPM method. An exact explanation for such high ECPs of the bootstrap percentile estimators is subject to future research. Recall, $\hat{G}$ in samples from DGPs under the Pareto distribution are highly biased, even in our biggest sample, MCS4. However, as these percentile methods are formed based on the resampling distribution of $\hat{\hat{G}}$, which is highly symmetrical, may result in such high ECPs.

b) Empirical Coverage probabilities with increasing number of clusters in the samples drawn from DGPs with varying values of ICC

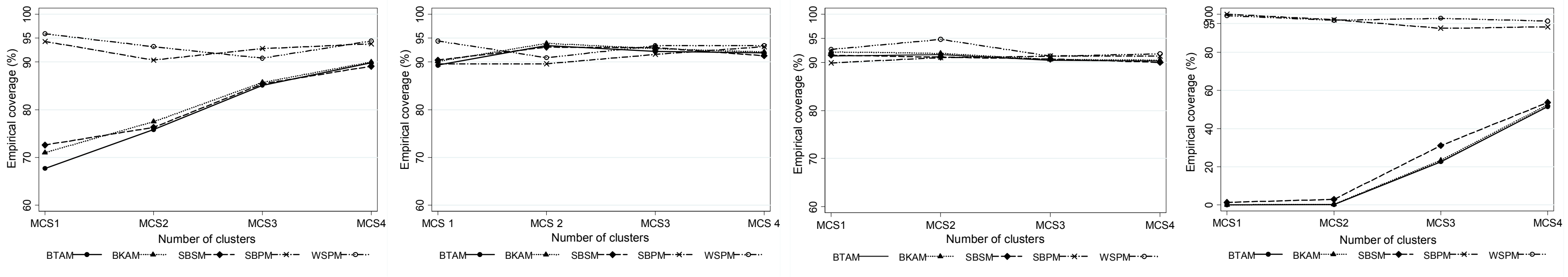
Intracluster correlation coefficients across different strata are not the same in real-world survey data. To see the effects of varying ICC values across strata in data on the ECP of a CI for G, we examine eight DGPs under four distributional assumptions (beta, lognormal, chi-square, and Pareto distributions). Empirical CPs of the 95% interval estimators are reported in Figure 8 for samples with an increasing number of clusters from those DGPs. Recall, the varying ICCs in DGPs with two strata are 0.05 and 0.1, in strata 1 and 2, respectively, and they range from 0.05 to 0.3 in strata when DGPs contain five strata.

This case differs from the previous case by the values of ICC parameter across strata in a DGP, which we previously assumed to be fixed at 0.1 in all strata. However, the way in which an interval estimator responds to the increase in the number of clusters in a sample obtained from varying ICCs across strata is fairly similar to the formerly discussed case. For example, ECPs of the three SN approximation interval estimators are, on average, higher across samples from the lognormal distribution than from the other distributions. In the case of the Pareto distribution, for samples MCS1 and MCS2 from DGP4B, the coverage is zero and after that it rises with more clusters in the sample, although coverage fails to exceed 60% for MCS4. In contrast, the ECPs of these interval estimators are zero across samples drawn from a DGP with five strata. The performance of the SBPM interval estimator is parallel to the previous case outcomes across samples and DGPs, where we saw that the ECP of the interval estimator was around the 95% nominal level.

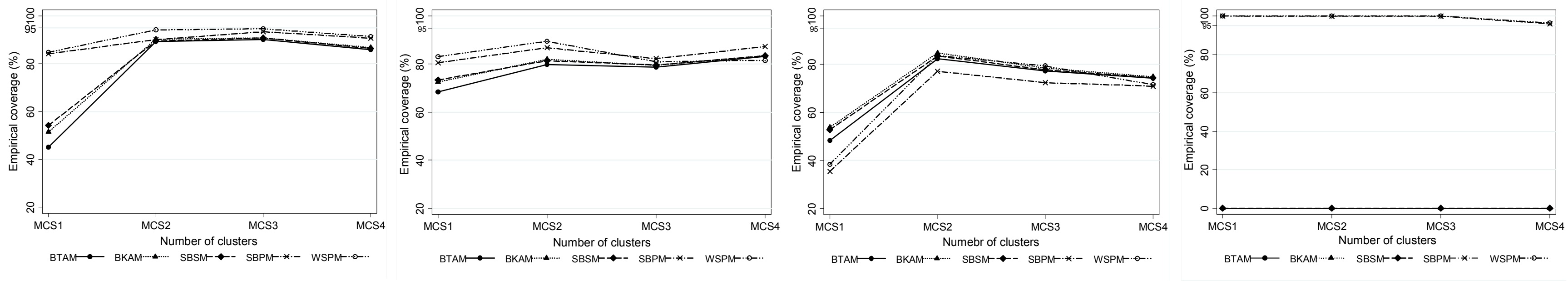
We see some differences in the level of coverage of an interval estimator for data containing different ICCs across strata, compared to the base case, except for the samples from DGP4I. When contrasted with the base case results, we see that the ECPs of the SN approximation interval estimators are slightly lower for samples from DGPs with two strata created under the chi-square distribution, whereas they are somewhat higher for larger samples (MCS4) from beta and lognormal distributions. On the other hand, when the outcomes of these interval estimators for a larger sample (e.g., MCS4) from DGPs with five strata and varying ICCs across strata are compared with the DGPs with five strata and a constant ICC, we see, the coverage of the interval estimators improves under the beta and chi-square distributional assumptions, but marginally deteriorates for data from the lognormal distribution. This may be due to an increase in the bias of the estimator in this sample (1.81% to 2.05% of G) when drawn from the DGP with varying ICCs than the fixed ICC under the lognormal distribution, whereas for the other distributions the empirical bias decreased.

Figure 8 Empirical coverage probabilities of the 95% CI estimators for *G*: samples with disproportional increase in the number of clusters from DPGs with varying ICCs across strata.

Beta distribution | Lognormal distribution | Chi-square distribution | The Pareto distribution



a) DGP1G b) DGP2G c) DGP3G d) DGP4G



e) DGP1I f) DGP2I g) DGP3I h) DGP4I

Notes: First row shows results for DGPs with two strata. Samples drawn from these DGPs are MCS1: MC sample 1 with n_1=10, n_2=15; MCS2: MC sample 2 with n_1=25, n_2=25; MCS3: MC sample 3 with n_1=40, n_2=50; and MCS4: MC sample 4 with n_1=75, n_2=90. Second row shows results for DGPs with five strata. Samples drawn from these DGPs are MCS1: MC sample 1 with n_1=3, n_2=5, of n_3=6, n_4=8, n_5=9; MCS2: MC sample 2 with n_1=6, n_2=10, n_3=12, n_4=15, n_5=20; MCS3: MC sample 3 with n_1=12, n_2=20, n_3=24, n_4=32, n_5=36; and MCS4: MC sample 4 with n_1=25, n_2=35, n_3=45, n_4=64, n_5=70. BTAM: SN approximation CI using Bhattacharya's (2007) standard error, BKAM: SN approximation CI using Binder and Kovačević's (1995) standard error, SBSM: SN approximation CI using bootstrap standard error, SBPM: standard MC bootstrap percentile CI and WSPM: warp-speed MC percentile CI.

Overall, for small samples containing 10-15 clusters per stratum and 3-9 clusters per stratum obtained from DGPs with two strata and five strata, respectively, substantial under-coverage problems occur with the three SN approximation interval estimators, especially with samples from five strata DGPs. Although the performance of an interval estimator does not always improve when more clusters are included in the sample, our results suggest that to achieve at least 80% CP with a SN approximation interval estimator, the sample should contain around 10% and 12% of the population clusters from every stratum when samples are drawn from DGPs with two strata and five strata, respectively, created under the typical skewed distributions (beta, lognormal, and chi-square).

In terms of ECPs, of the three SN approximation interval estimators, those using the asymptotic variances, BTAM and BKAM, behave in a fairly similar manner across samples and DGPs. The difference in ECPs of these interval estimators, which arises in smaller cluster samples, is due to the correction factor associated with the variance estimator used in the BKAM interval construction, as it disappears for samples with more clusters. On the other hand, in some small cluster samples, the SN approximation interval estimators using the bootstrap variance, SBSM, provide higher ECPs than when the asymptotic variances are adopted. Nevertheless, the dominance of the SN bootstrap interval estimators fades with the increasing number of sampled clusters in the sample.

These findings suggest that practitioners may consider using the SBSM interval estimator over the analytical variance based interval estimators, when the sample contains a small number of clusters. However, for complex survey samples containing around 6% of the population clusters, the SBSM variance does not seem to be more efficient than the analytical variances. In addition, when accounting for the computational expense, BTAM or BKAM interval estimators are recommended over the SBSM interval estimators. As such, the SBSM interval estimator offers an alternative estimator, rather than a generally more efficient CI estimator, compared to that based on asymptotic variance formulae for $\hat{G}$.

The bootstrap MC percentile interval estimator often performs reasonably well, with ECPs close to, and in some cases exceeding, the assigned nominal level. In particular, for small cluster samples or samples from the highly skewed Pareto distribution, this percentile method might be a good choice for applied researchers. Nevertheless, for samples with large numbers of clusters obtained from typical skewed distribution, the BTAM or BKAM interval estimators can be preferred, especially as these methods are computationally time saving.

Coverage probabilities of the WSPM estimators in samples with fewer clusters are generally higher than those of the SBPM estimators across DGPs, except for the Pareto distribution. Giacomini et al. (2013) report that the ECPs of the WSPM estimators are highly comparable to that of the SBPM interval estimators, and the differences in these coverages occur only in the third decimal place (less than 1%) for an *iid* sample of moderate size. We also find small differences between these two estimators in some cases for our simulations with complex survey samples, when samples contain a large number of clusters, such as for the sample MCS4. In these cases, the WSPM method reflects the performance of the SBPM interval estimators closely; however, when considering cases with samples containing 10-15% of the population clusters (MCS2 and MCS3), and are obtained from usual skewed distributions, we do not think that the method well compares with the coverage probability of the double bootstrap interval estimator.

Both the SN approximation and the SBPM interval estimators under the Pareto distribution, especially in samples from DGPs with five strata, produce ECPs that are far from desirable. We assume that the size of the sample and the high bias of $\hat{G}$ may be the causes for the zero ECPs of these interval estimators. Although, the coverage of the SBPM interval estimator, which is formed using the resampling distributions of $\hat{G}$ for 1,000 MC samples (see Figure C.7d, C.8d, and C.9d in Appendix C) can be 100%, the lack of sensitivity of the outcome to increasing the number of clusters in the sample is unexpected. To further investigate this issue, we undertook simulations from an additional DGP, DGP4J, with three strata and a constant ICC value of 0.1 across strata under the Pareto distribution. Estimated Gini coefficients, variance estimates resulting from the three SN approximation methods, and their corresponding 95% CIs for G and the 95% CI for the SBPM method are reported in Table 9. The ECPs of the interval estimators along with upper and lower tail CPs are shown in Table 10.

The Gini coefficient of the finite population DGP4J with 750 clusters, containing 105,506 households, is 0.002. As reported in Table 9, the difference between $\hat{G}$ and G in the sample MCS1 with 5% of the population clusters is 30% of G, although it declines when the number of sampled clusters is increased. We also see that the 95% SN interval estimators using the analytical variance estimators are similar in each sample. The SBSM estimator is slightly wider, but becomes similar to these two SN approximation intervals as the number of sampled clusters increases. On the other hand, the width of the SBPM interval estimate is comparable to the SBSM interval estimator, although the SBPM interval estimate is slightly shifted to the right.

Table 9 Estimated G and variance using SN approximation methods and 95% CIs for G using a single MC sample and 99 bootstrap replications for DGP4J.

DGP	G	MC Sample	$\hat{G}$	BTAM			BKAM			SBSM			SBPM	
				$\widetilde{Var_{BT}}(\hat{G})$	95% CI	Width	$\widehat{Var_{BK}}(\hat{G})$	95% CI	Width	$Var_{BOOT}(\hat{G}_b^*$	95% CI	Width	95% CI	Width
DGP4K	0.002	1	0.0026	0.0765	0.0022,0.0030	0.0008	0.0778	0.0022,0.0030	0.0008	0.0806	0.0021,0.0030	0.0009	0.0019,0.0028	0.0009
		2	0.0024	0.0331	0.0020,0.0027	0.0007	0.0348	0.0020,0.0027	0.0007	0.0513	0.0019,0.0028	0.0009	0.0014,0.0023	0.0009
		3	0.0021	0.0122	0.0019,0.0023	0.0004	0.0125	0.0019,0.0023	0.0004	0.0120	0.0019,0.0023	0.0004	0.0017,0.0021	0.0004
		4	0.0020	0.0062	0.0018,0.0022	0.0004	0.0063	0.0018,0.0022	0.0004	0.0056	0.0018,0.0021	0.0003	0.0017,0.0020	0.0003
		5	0.0019	0.0042	0.0018,0.0021	0.0003	0.0042	0.0018,0.0021	0.0003	0.0042	0.0018,0.0021	0.0003	0.0017,0.0020	0.0003

Notes: Notes: Variance estimates have been scaled by 10^6. $\widetilde{Var_{BT}}(\hat{G})$ and $\widehat{Var_{BK}}(\hat{G})$ are Bhattacharya's (2007) and Binder and Kovačević 's (1995) estimators, respectively. $\widehat{Var_{BOOT}}(\hat{G}_b^*)$ is the bootstrap variances estimator using expression (3.7) for 99 replications. MC Samples 1-5 are samples with 5%, 10%, 20%, 30% and 40% of the population clusters respectively. BTAM: SN approximation CI using Bhattacharya's (2007) standard error, BKAM: SN approximation CI using Binder and Kovačević's (1995) standard error, SBSM: SN approximation CI using bootstrap standard error, SBPM: standard MC bootstrap percentile CI and WSPM: warp-speed MC percentile CI.

Table 10 Empirical coverage probabilities (CP%), lower (L%) and upper (U%) tail error rates of 95% confidence intervals for the Gini coefficient for DGP4J.

MC Sample	BTAM			BKAM			SBSM			SBPM			WSBM		
	CP	L	U	CP	L	U	CP	L	U	CP	L	U	CP	L	U
1	0.30	99.70	0.00	0.30	99.70	0.00	2.20	97.80	0.00	96.00	0.00	4.00	96.10	0.50	3.40
2	10.90	89.10	0.00	11.90	88.10	0.00	23.60	76.40	0.00	89.90	0.20	9.90	98.00	1.20	0.80
3	82.40	16.40	1.20	83.60	15.20	1.20	85.90	13.50	0.60	72.50	0.10	27.40	85.80	0.50	13.70
4	89.70	8.30	2.00	89.90	8.10	2.00	90.30	8.10	1.60	75.90	0.00	24.10	87.80	0.40	11.80
5	89.90	9.30	0.80	91.40	7.80	0.80	91.50	7.80	0.70	87.10	0.60	12.30	95.30	0.90	3.80

Notes: MC Samples 1-5 are samples with 5%, 10%, 20%, 30% and 40% of the population clusters respectively. BTAM: SN approximation CI using Bhattacharya's (2007) standard error, BKAM: SN approximation CI using Binder and Kovačević's (1995) standard error, SBSM: SN approximation CI using bootstrap standard error, SBPM: standard MC bootstrap percentile CI and WSPM: warp-speed MC percentile CI.

ECPs of the three 95% SN approximation interval estimators for G are highly similar and increase for samples with more clusters, although they are below the assigned nominal level. For the sample MCS4 that contains 40% of the population clusters, the ECP is 90-92%. Coverage for the SBPM interval estimator is better than for the SN approximation interval estimator, but only when samples contain 5-10% of the population clusters. For larger proportion of clusters in the samples, we see that the SN approximation interval estimators achieve higher ECPs than the SBPM interval estimator, although the ECPs remain below the nominal level. As such, we conclude that the SN approximation methods using analytical variances are more convenient ways to assess inequality in data, even with heavy-tailed distributions, compared to computationally cumbersome bootstrap techniques, when samples contain a large proportion of clusters from few strata.

The coverage of the warp-speed CI estimator is lower than the nominal level for samples with 20-30% of the population clusters, but for other small (5-10% clusters) or large (40% clusters) samples, it is higher than the 95% nominal coverage rate. In the previous cases, especially for DGPs with five strata, we found that the ECPs of the warp-speed interval estimator were similar to those of the SBPM interval estimator, regardless of the number of clusters in the sample. Since results from using the warp-speed approach performs substantially differently from the SBPM interval estimator, we again argue that the warp-speed method is unable to well replicate the CPs of our double bootstrap interval estimator.

In terms of the tail ECPs, the three SN interval estimators lead to similar outcomes. For example, for small cluster samples, the interval estimators have larger lower tail ECPs, but these fall when the cluster sample size increases, indicating a more symmetrical empirical distribution of $\widehat{G}$ in samples with a larger proportion of the clusters. Overall, when data are generated from the Pareto distribution, the proportion of clusters selected in the sample from the population matters. It should be relatively large, containing at least 20% of the clusters in the population, to achieve a moderate ECP for the 95% SN approximation CI estimators. A sample as large as 40% of the population clusters provides ECPs of around 90%. For samples containing fewer clusters, e.g., under 20% of the population clusters, it is likely preferable to use the SBPM interval estimator to generate estimates rather than adopting any of the SN approximation interval estimators.

3.5 Concluding Remarks

We have performed MC simulation experiments using 36 DGPs created under four distributional assumptions (beta, lognormal, chi-square, and the Pareto distributions) to examine the finite sample properties of $\widehat{G}$, and the ECP of the 95% CIs for G, when samples are obtained under the complex survey design. We find that the population features, such as distribution, number of strata, ICC values across strata, and the proportion of sampled clusters, affect such features.

Our simulation results suggest that, in general, the bias of $\widehat{G}$ decreases with an increasing ICC value in the data and with the proportion of sampled clusters. The bias of $\widehat{G}$ for data under the Pareto distribution is high; however, for a fixed number of clusters in the sample (10% of the population clusters from two strata DGPs), the bias decreases with increasing ICC values across strata in a DGP. The estimator is approximately unbiased for an ICC value of around 0.5 in DGPs created from two strata, and the beta, lognormal, and chi-square distributional assumption. On the other hand, the bias of $\hat{G}$ becomes negligible when samples are selected with more than 18% and 24% of population clusters from DGPs with two strata and five strata, respectively, except for the Pareto distribution. Perhaps, due to the high skewness of data under the Pareto distribution, for these cluster sample sizes, the bias remains rather large, above 10% of G.

Outcomes of the two 95% SN approximation interval estimators for G using analytical variances are highly similar across samples and DGPs. In some cases, particularly for samples with few clusters (e.g., the sample MCS1), the SN approximation interval estimators using the bootstrap variance lead to slightly higher ECPs than the other two SN approximation interval estimators. Nevertheless, results from this bootstrap interval estimator is similar, sometimes lower, to that from the two SN approximation interval estimators, for samples with more clusters.

The SBPM interval estimator generally produces higher ECPs for small samples and the heavy-tailed Pareto distribution, compared to those generated using the three SN interval estimators. For samples with larger number of clusters, this approach to forming an interval estimator results in similar ECPs to the SN approximation interval estimator, often lower for heavily skewed distributions; e.g., DGP with three strata under the Pareto distribution.

We have considered Giacomini et al.'s (2013) recommendation of the approximate equivalence of coverages of the warp-speed method to the double-bootstrap interval estimator for our complex survey framework. With few exceptions, we found that the WSPM interval estimator leads to different outcomes (higher ECPs, as high as

14% under the Pareto distribution) from the double-bootstrap technique. As such, we are not convinced that the warp-speed approximation works well at reflecting outcomes from the SBPM interval estimators under a complex survey design.

Finally, samples with less than 5% of the population clusters per stratum a DGP are too small to lead to satisfactory results using SN approximation interval estimators. Nevertheless, our simulations suggest that the performance of an interval estimator does not always improve by just including more clusters in the sample. Outcomes also depend on two crucial population features: the number of strata and the underlying distribution of the well-being data. For reasonably skewed distributions of data (beta, lognormal and chi-square), achieving a reasonable coverage probability (say an ECP of around 90%) using SN approximation interval estimators is possible with samples containing around 10% of the population clusters drawn from DGPs with two strata and 0.05-0.1 ICC across strata. When the number of strata in the population increases, the overall coverage probability may decrease. To obtain moderate ECPs of around 80%, samples from the five strata DGPs should have approximately 12% of the population clusters.

Overall, when the distribution of a well-being variable is not too heavy-tailed, for samples with over 10% of the population clusters, the SN approximation interval estimators using Bhattacharya's (2007) and Binder and Kovačević's (1995) variance estimators perform well. When computation time involved in estimation of CIs for G is accounted for, even for samples with fewer clusters, these asymptotic variance estimators can be adopted in inequality analysis, as their outcomes are often comparable to resampling alternatives.

CHAPTER FOUR: HEALTH WELL-BEING IN BANGLADESH

4.1 Introduction

Inequality in income (or consumption expenditure) as a measure of well-being across individuals and households is well documented. In contrast, inequality studies on other well-being attributes such as health and healthcare service utilization, have been less frequent but are increasing recently (see, e.g., Clarke and Roy, 2012; Araar et al., 2009, Slater et al., 2009). Our work continues this research on inequality in health well-being.

Health is a fundamental component of well-being and many countries use different indicators in regular reporting of the health status of their populations. These include birth- and death-related indicators (e.g., life expectancy at birth, infant mortality rate), morbidity indicators (e.g., prevalence of diabetics, obesity) and health service indicators (e.g., antenatal care utilization, immunization rates). They are also used by major organizations such as the World Health Organization (WHO) and the United Nations Children's Fund (UNICEF), for research and guidance regarding global health development. Nevertheless, these health indicators are focused on national averages, and individually, none can provide a complete picture of a country's health status. For instance, birth- and death-related indicators do not account for the prevalence of preventable deaths; nor do health service indicators account for the quality of the services.[24] Moreover, when setting policy agendas, summary statistics such as the average level of health would not likely be considered as a sufficient indicator of the

---

[24] See, for example, Eurostat: Statistics Explained; available from http://ec.europa.eu/eurostat/statistics-explained/, last accessed November 7, 2015.

health status of a population or the performance of the health system (see, e.g., Murrary and Evans, 2003, p. 471-481).

We can better understand the distribution of health in a country by examining the inequality in health well-being. Inequality in health may be defined as variations in health status across individuals or groups in the population, allowing for the natural variations that arise across a healthy population, that may be arising from differences in, for instance, income, education, living conditions, ethnicity, gender, and lack of access to healthcare services. Health inequality, above a naturally arising level of inequality in a healthy population, is a serious socioeconomic concern that may undermine developmental efforts. Therefore, the monitoring and measuring of inequalities in health is important. Reliable health inequality measures are crucial for policy creation, and they are important indicators of the health of a population.

To contribute to the broadening literature on empirical explorations of inequalities in health status, we use the Gini coefficient (G). We consider two health indicators: Body Mass Index (BMI) of ever-married women aged 15-49 years and Hemoglobin (Hb) level among children aged 6-59 months to assess the health of women and children, respectively, in Bangladesh. Our choices for the two health indicators were made because they may indicate the well-being of all Bangladeshis. For instance, the BMI of a married woman, which indicates her physical health status, may also influence her reproductive capability and ability to give birth to healthy babies (see, e.g., Hossain et al., 2012; Lake et al., 1997), raising the health quotient of the next generation. Similarly, the Hb level of a child may also indicate potential participation in the future labour market and contribution to economic and social outcomes.

Over the last four decades, Bangladesh has achieved significant progress in health, as demonstrated by some average health indicators. For example, life expectancy at birth increased from 60 to 70 years of age between 1990 and 2012. The maternal mortality ratio[25] reached 194 in 2010, a 40% drop from 2001, which is an average decline of 3.3% per year (BPC 2014, p. 62-63). The country has also achieved targets of the Millennium Development Goal (MDG) 4, particularly Target 4A.[26]

However, inequalities in the health indicators between or across categories remain troubling. For example, the under-five mortality rate for children belonging to the lowest wealth quintile is double that of children in the highest wealth quintile (WHO 2014, p. 155). In addition, while some studies on health inequality in Bangladesh use descriptive statistics or compare the trends in health and health care provision of rich and poor; e.g., Uddin et al. (2012), Khan et al. (2011), Mahabub-ul-Anwar et al. (2008), Hong et al. (2006), measuring inequality in key indicators of health using formal inequality measures is still unexplored.

Our examination of inequality in two health indicators using G can improve our understanding of the underlying status of health and may assist in leading to policy interventions for improved health care in Bangladesh. Besides estimating G and associated sampling variances for ever-married women, we also report Gini coefficient estimates $(\hat{G})$ when the data are divided by place of residence (urban; rural), wealth category (poorest; poorer; middle; richer; richest), and educational attainment (no education; incomplete primary; primary complete; incomplete secondary; secondary

---

[25] The maternal death ratio is the ratio of the number of maternal deaths in a period to live births during the same period, per 100,000 live births (e.g., BPC, 2014, p. 64).

[26] MDG 4; Target 4.A: Reduce the under-five mortality rate by two-thirds, between 1990 and 2015. In 2012, the under-five mortality rate (per 1,000 live births) and the infant mortality rate (per 1,000 live births) were 41 and 33 respectively, which are equal to or less than one-third of these rates in 1990 (WHO, 2014, p. 61)

complete or higher). For children, $\hat{G}$, and associated sampling variances are reported when they are classified by gender, place of residence, age group (6-23 and 24-59 months), wealth category, and division/region (Barisal; Chittagong; Dhaka; Khulna; Rajshahi; Rangpur; Sylhet). These breakdowns into subgrouops enable us to explore whether or not such features contribute to differences in health inequality. We also report 95% confidence interval (CI) for G for each case.

The remainder of this chapter is organized as follows. In section 4.2, we provide details of the data and survey design, including issues related to sampling weights and estimation methodology. In section 4.3, we document our first application of examining inequality among Bangladeshi women 15-49 years of age, using BMI data. Section 4.4 presents our second application regarding inequality among Bangladeshi children, 6-59 months of age, using Hb data. Section 4.5 concludes the chapter with remarks on the research.

4.2 Data, Sampling Design and Methodology

4.2.1 Data and Sampling Design

The data are taken from three of the Bangladesh Demographic and Health Surveys (BDHS 2011, 2007 and 2004), which are nationally representative sample surveys, useful for ascertaining the changes in key areas of development, including maternal and child health, domestic violence, education and poverty reduction (see NIPORT et al., 2013, 2009, 2005). These are the fourth, fifth, and sixth surveys of their kind for this developing country, and they were conducted in 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.

For the first application, examining health inequality among Bangladeshi ever-married women, 15-49 years of age, the raw BMI data is taken from each BDHS. We limit our attention to this age group for the sake of consistency across surveys. The 2007 survey originally included ever-married women age 10-49 years, but the 10-14 years age group was dropped because of the small number in this group. For the second application, health inequality among Bangladeshi children, 6-59 months old, the Hb data is only taken from the BDHS 2011 as earlier surveys did not include this health measure. Hb testing was carried out by the HemoCue system, a WHO recommended rapid testing technology for determining Hb levels in field surveys, for eligible children, 6-59 months old, from every third household selected for interviews with ever-married men, 15-54 years of age.

The Demographic and Health Survey (DHS) program collects data using a stratified two-stage cluster sampling method with somewhat similar designs across surveys. We provide a brief outline of the BDHS here, referring readers to NIPORT et al. (2013, 2009, 2005) for more details.

Bangladesh is comprised of administrative divisions[27] that are split into zilas, which are further subdivided into upazilas. In rural regions, upazilas are split into union parishads, which are further divided into mouzas. Urban region upazilas contain wards, which are then 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 to mouzas in rural regions, with 100-120 households, on average, in each EA. Large EAs

---

[27] The 2004 and 2007 surveys involved six administrative divisions (Barisal, Chittagong, Dhaka, Khulna, Rajshahi and Sylhet), with an additional division included in the 2011 survey (Rangpur).

are further divided with the segmented areas forming the PSUs. With the country being stratified into rural and urban areas for each division, the urban areas are further stratified into city corporations or non city corporations for the 2011 survey, and into statistical metropolitan areas, municipality areas and other urban areas for the 2004 and 2007 surveys. This results in 20 strata for the 2011 survey and 22 strata for the 2004 and 2007 surveys.[28]

For sampling, in the first stage, the desired number of EAs (clusters) was selected, independently across the strata, with probability proportional to the number of households in the EA. Upon selection, for each of the chosen clusters, a household listing and mapping operation was conducted to list households residing in each EA with a location map that also updated the population information for the sampling frame. In the second stage, from each of the sampled clusters, 30 households were selected using an equal probability systematic selection technique, and all household members of a certain age group (e.g., all ever-married women,15-49 years of age) in the sampled households were selected for the survey. Such a stratified two-stage cluster sampling has some advantages; for instance, by providing good coverage and being easier to implement than a simple random sample (e.g., see Aliaga and Ren, 2006, for discussion). As most of the population in Bangladesh resides in rural areas, urban households were over-sampled to obtain a level of statistical precision that was similar to the rural regions.[29]

For the BMI data, in the BDHS 2011 (2007; 2004) survey, 17,309 (10,836; 11,166) ever-married women, 15-49 years of age, were selected from 600 (364; 361) clusters across 20 (22; 22) strata. The number of clusters were allocated proportionally

---

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

[29] In addition, as one of the BDHS objectives is to obtain estimates for each division, and for the country as a whole, some divisions were also oversampled. Thus, the BDHS samples were not self-weighting.

across strata; in 2011 (2007; 2004), 393 (228; 239) clusters were selected from rural areas, and 207 (136; 122) were selected from urban areas. The range in number of clusters within a stratum varied from 5 (3; 2) to 61 (46; 57).

As previously mentioned, the Hb data was from a portion of the selected households in the BDHS 2011 survey. The sample was comprised of 2,432 children, which was 92% of the eligible children selected for measuring Hb levels. These children were selected from 583 clusters across the 20 strata, with a wide range in number of clusters across strata – from 5 to 60. Given that all households do not have eligible children under five years of age, and only one-third of the households were selected for Hb collection, the number of children in a cluster was small, ranging from only 1 to 16 children. This issue and its effect on our analysis is discussed in subsection 4.4.3.

Potentially, we have more than one observation (an ever-married woman or a child under five) from each household in the sample. Thus, the “i” (or “t”, as appropriate) subscript, in the expressions from Chapter 2, are for all women and children in the sampled cluster, rather than for households. The total number of sample observations is now the number of ever-married women age 15-49 years or children 6-59 months old. Such changes are irrelevant as long as the “correct” sampling weight is adopted. Here, we use the individual weight from the ever-married women and men survey files. Let W_{hci} be an individual’s sampling weight, normalized so that $\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m}W_{hci}$ is equal to the number of sampled individuals; we assume that the weights have been scaled to ensure that $\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m}W_{hci} = M$, the number of sampled individuals. In addition, it simplifies the algebra to adopt standardized weights $w_{hci} = \frac{W_{hci}}{\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m}W_{hci}}$ such that $\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m}w_{hci} = 1$.

4.2.2 Methodology of Estimation

Our goal is to estimate G, and sampling variances to make statistical inferences about inequalities in BMI among women, and Hb levels among children in Bangladesh. Nevertheless, we also detail features of the BMI and Hb data using qualitative, quantitative and distribution analyses before providing $\hat{G}$ and their sampling variances using the formulae and techniques described in Chapters 2 and 3.

4.2.2.1 Calculating the Gini Coefficient

To measure inequality in the BMI or Hb distribution, we calculate the plug-in estimator for G as described in Chapter 2. Let y_{hci} denote the BMI (Hb level) of the ith woman (child) in cluster c of stratum h with plug-in estimator of G:

$$\hat{G} = \frac{2}{\hat{\mu}} \sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci}\, y_{hci} \hat{F}(y_{hci}) - 1, \tag{4.1}$$

where $\hat{\mu} = \sum_{h=1}^{H} \sum_{c=1}^{n_h} \sum_{i=1}^{m} w_{hci}\, y_{hci}$ and $\hat{F}(y_{hci}) = \sum_{r=1}^{H} \sum_{s=1}^{n_r} \sum_{t=1}^{m} w_{rst} I\{y_{rst} \leq y_{hci}\}$.

To calculate $\hat{G}$ using BMI, we need to ensure comparability across different ages. We allow for natural age effects for younger women by standardizing the BMIs using growth curves provided by the WHO (e.g., see, de Onis et al., 2007). The WHO finds that the distribution of BMIs for 19 year olds matches that for adult women, thus we use WHO growth curves to convert the BMIs for women from 15 to 18 years of age 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 by 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 $a = 15, \ldots, 18$, we generate $BMI_{19,i}$ such that $F_a(BMI_i) = F_{19}(BMI_{19,i})$. Our notion of "equivalent" with these "standardized" BMIs is used in the statistics generated for the sample.[30] The number of observations is 1295 (954; 1147) for the 2011 (2007; 2004) data-sets. Hereafter, we use "BMI" to denote the series that includes the standardized BMI numbers for the women younger than 19 years of age.

4.2.2.2 Inference: Variance Estimators Accounting for Complex Surveys and Design Effects

To make inferences about G, the sampling variance or the standard error is estimated using two asymptotically equivalent variance expressions provided by Binder and Kovačević (1995) and Bhattacharya (2005, 2007). These are:

$$\widehat{Var_{BK}}(\hat{G}) = \sum_{h=1}^{H} \sum_{c=1}^{n_h} (u_{hc}^{*} - \bar{u}_{h}^{*})^2, \tag{4.2}$$

and

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

respectively (see Chapter 2 for details). We also calculate the bootstrap estimator for the variance of $\hat{G}$, following Wolter (2007, p. 215), given by:

---

[30] 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 10 June 2015.

$$\widehat{Var}_{BOOT}\left(\hat{G}^*\right) = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{G}_b^* - \bar{\hat{G}}^*)^2, \tag{4.4}$$

where B is the number of bootstrap replications, $\bar{\hat{G}}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{G}_b^*$ and $\hat{G}_b^*$ is an estimator of G calculated using the formula in expression (4.1) in bootstrap replication b. The purpose of the bootstrap variance calculation is to see whether or not the variance estimator is similar to those from the asymptotic expressions in practical cases. It would be useful if the bootstrap estimates had a magnitude that was similar to the asymptotic estimates, as the latter are far easier to generate in practice. The programs STATA 9.2 (StataCorp, 2007), STATA 11 (StataCorp, 2007) and EViews 8 (Quantitative Micro Software, 2013) were used throughout. The number of bootstrap replications is $B = 799$.

We report the breakdown of the variance estimators into three components using Bhattacharya's (2007) formula in (4.3): the variance arising from naively assuming a simple random sample (SRS) design with weights, and the effects of clustering and stratification on the variance estimate. Typically, stratification increases the precision and clustering reduces the precision of an estimator. If the cluster effect is not fully offset by the stratum effect, variances associated with a complex survey design are larger than those obtained from standard formulae based on simple random sampling.

To measure the effect of the survey design (stratification, clustering, unequal selection probabilities, and sample weighting adjustments for non-response) on a variance estimator, we estimate the design effect (*deff)*, a widely used statistical measure (discussed in Chapter 3). Design effects are calculated for $\hat{G}$ using Bhattacharya's (2007) variance formula as the ratio of $\widetilde{Var}_{BT}\left(\hat{G}\right)$ and $\widetilde{Var}_{SRS}\left(\hat{G}\right)$. The larger the design effect, the less precise is the estimate relative to that which would be obtained using a simple random sample.

As clustering typically increases the sample variance of an estimator, and hence the design effect, further discussion is warranted on the properties of the cluster. The intracluster correlation coefficient (ICC) is an important aspect of clustering, and in Chapter 3, we clearly demonstrate how the ICC affects the finite sample properties of $\hat{G}$. An often used expression to show the relationship between the design effect and ICC due to clustering within a stratum h, is:

$$\mathrm{deff}^{*}(\hat{\theta}) \approx 1 + (\bar{m} - 1)\rho_h, \tag{4.5}$$

where $\bar{m}$ is the average number of individuals in a cluster and ρ_h is the ICC within stratum h (see, e.g., Cochran, 1977, p. 242). Expression (4.5) is an approximation only because clusters are not uniform in size.

The ICC for stratum h, ρ_h, measures how strongly individuals within a cluster resemble each other and their difference from other clusters within the stratum. In other words, the ICC compares the variance between clusters (γ_h^2) with the total variance within a stratum ($\gamma_h^2 + \sigma_h^2$, where σ_h^2 is the within cluster variance). A large ICC implies larger sampling variance. For DHS surveys, necessary precautions are taken to keep the ICC as small as possible when the sampling design is implemented (see, e.g., Aliaga and Ren, 2006; UN, 2005). Although clusters are assumed to be homogeneous, in practice, this may not be a reasonable assumption for many reasons. For instance, some clusters may be wealthier than others, some may have better access to education opportunities or health facilities, better living conditions, and so on. In such cases, within cluster homogeneity can be larger, resulting in higher ICC and variance estimates for $\hat{G}$.

4.2.2.3 95% Confidence Interval (CI) Estimators for the Gini Coefficient

To assess the extent of inequality in BMI among ever-married women, 15-49 years of age, and Hb among children, 6-59 months old, in Bangladesh, we construct 95% CIs for G. These intervals vary in how they are constructed; two of the intervals are constructed using asymptotic standard normal cut-offs with standard errors from Binder and Kovačević's and Bhattacharya's variances. The other two intervals are based on bootstrapping: the bootstrap percentile and percentile-t CIs, briefly discussed below.

The $(1-\alpha)100\%$ standard normal (SN) approximation CIs for G using the standard error from Binder and Kovačević's (1995) variance, is:

$$\left[\hat{G} - z_{(1-\alpha/2)} se_{BK};\ \hat{G} + z_{(1-\alpha/2)} se_{BK}\right], \tag{4.6}$$

where $se_{BK} = \sqrt{\widetilde{Var_{BK}}(\hat{G})}$ and z is the standard normal value. Similarly, using Bhattacharya's (2007) variance, the interval estimator is:

$$\left[\hat{G} - z_{(1-\alpha/2)} se_{BT};\ \hat{G} + z_{(1-\alpha/2)} se_{BT}\right], \tag{4.7}$$

where $se_{BT} = \sqrt{\widetilde{Var_{BT}}(\hat{G})}$. Such SN approximation CIs are justified as the estimator for G is asymptotically normally distributed, as shown by Bhattacharya (2007, 2005) (and which we discussed in Chapter 2).

Alternatively, given that the bootstrap variance estimator may provide a more accurate interval estimator, especially in small samples (as illustrated in Chapter 3), we report two such bootstrap interval estimators. The percentile CI is constructed using the sampling distribution of the bootstrap estimates $\left\{\hat{G}_j^*\right\}_{j=1}^{B}$ as the empirical distribution function for G, where, B is the number of bootstrap replications. To obtain a sample value using this interval estimator, we arrange the bootstrap estimates, $\hat{G}_j^*$, in ascending order, select the $\alpha/2$ and $(1-\alpha/2)$ centiles of $\hat{G}_j^*$, denoted by $G^*(\alpha/2)$ and $G^*(1-\alpha/2)$, respectively, which then form the lower and upper limits of the interval estimator:

$$[G^*(\alpha/2), G^*(1-\alpha/2)]. \tag{4.8}$$

For the endpoints of the interval to be exact, $\alpha(B+1)/2$ must be an integer (e.g., Davidson and MacKinnon, 2000); we select $B = 799$.

Finally, following Davidson and McKinnon (2004, p. 185-196), we construct a percentile-*t* or bootstrap-*t* CI for G, based on the bootstrap *t* statistic $t_j^* = (\hat{G}_j^* - \hat{G})/se_j^*$, where $\hat{G}_j^*$ is the estimate of G from the jth bootstrap sample, $j = 1, \dots, B$. The estimator $\hat{G}$ is formed from the original sample. After sorting the t_j^* statistics from smallest to largest, we determine the $(\alpha/2)100\%$ and $(1-\alpha/2)100\%$ quintiles, denoted by $t_{\alpha/2}^{*H}$ and $t_{(1-\alpha/2)}^{*L}$ respectively. The lower and upper limits of the interval estimator for G are then: $\left[\hat{G} - se_B t_{(1-\alpha/2)}^{*L};\ \hat{G} - se_B t_{\alpha/2}^{*H}\right]$, where se_B is the bootstrap standard error of $\hat{G}_j^*$.

One of the limitations of the percentile-*t* interval estimator is that the standard error of the bootstrap estimator must be calculated for every replication, which typically requires a double bootstrap,[31] a computer intensive process as it requires $B \times B_w$ bootstrap replication, where B_w is the number of replications within each bootstrap replication used to generate a standard error estimate. We thus consider an alternative method of using one of the asymptotic standard errors rather than calculating a bootstrap standard error for each replication. Specifically, we use Binder and Kovačević's (1995) variance formula, to obtain the empirical *t* statistic, t_j^*, such that $t_j^* = \left(\hat{G}_j^* - \hat{G}\right)/se_{BK_j}^*$. The resulting percentile-*t* CI is then:

$$\left[\hat{G} - se_{BK} t_{(1-\alpha/2)}^{*L}, \hat{G} - se_{BK} t_{\alpha/2}^{*H}\right]. \tag{4.9}$$

---

[31] When the standard error of an estimator for each bootstrap replication is calculated by another full bootstrap (a bootstrap within a bootstrap replication), this is known as the double bootstrap method.

4.3 APPLICATION ONE: BMI INEQUALITY AMONG BANGLADESHI WOMEN, 15-49 YEARS OF AGE

4.3.1 Introduction

We examine health inequality among Bangladeshi women using the body mass index (BMI) as our gauge of well-being. The BMI is calculated as weight in kilograms divided by height in metres squared (kg/m^2). A woman is said to have normal weight, consistent with her height (good or normal physical health), when her BMI is at least 18.5 kg/m^2 but less than 25 kg/m^2. The cut-offs for low BMI (i.e., underweight) and high BMI (i.e., overweight) for a woman are BMI< 18.5 kg/m^2 and BMI≥25 kg/m^2, respectively.

BMI provides a crucial numeric measure to assess a woman's current health. Both low and high BMI have health consequences. A low adult BMI, which typically results from undernutrition, suggests inadequate access to adequate nutritious food. It may also indicate disease, leading to an increased likelihood of morbidity and mortality, birthing difficulties (as well as poor health in delivered infants), and reduced ability to work productively (see, e.g., Molini et al., 2010). Alternatively, a high adult BMI may also be detrimental to a woman's well-being, especially a BMI over 30 kg/m^2 (i.e., obesity). A woman with such a high BMI is likely to have a myriad of associated diseases such as high blood pressure, type 2 diabetes, coronary artery disease, asthma, gallbladder disease, and some cancers (see WHO, 2015). In addition, a National Institute of Health (NIH) study finds that being overweight, especially obese, increases the risk of death.[32]

The association of both low and high BMIs with a decline in well-being is an issue when calculating inequality indicators, as these measures assume a monotonic

---

[32] See http://www.nih.gov/news/health/dec2010/nci-01.htm, last accessed January 15, 2016.

ranking of the well-being measure among individuals. We ignore this concern for two reasons. First, as we shall see in section 4.3.2, 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. Second, as discussed by, for instance, Molini et al. (2010), it is unlikely that these obese BMI observations will severely skew the BMI distribution, in contrast to high incomes in that distribution.

Some studies that consider inequality in BMIs for developing countries include Araar et al. (2009), Sahn and Younger (2009), and Molini et al. (2010). Using the 2003/2004 Namibia Household Income and Expenditures Survey data, Araar et al. (2009) report an estimated Gini of 0.14 for Namibia, with no noticeable difference between BMI inequalities among males and females, but a higher inequality results among people with an urban residence compared to a rural residence. Sahn and Younger (2009) examine BMI inequalities at inter-country and intra-household levels, measured by Theil's mean log deviation, across 36 developing countries, based on DHSs and Living Standards Measurement Studies, with data from several surveys for each country. They find about half of total BMI inequality at the country level is within households, and suggest that the standard measures of inequality using household-level data may understate the true inequality. Molini et al. (2010) measure health and nutritional inequalities between and within households using BMI data in Vietnam by applying macro and micro analyses. For the 1990s, they show that the overall inequality increased with the improvement in individual socioeconomic status at the beginning of the period, and then decreased when the remaining part of the population achieved progress in health status, although at the micro level it seemed that men benefited more from economic improvements than did women.

This section of this chapter is organized as follows. Subsection 4.3.2 provides details on BMI status among Bangladeshi women, 15-49 years of age, focusing on the prevalence of thinness and obesity. In addition, we report summary statistics and the distributions of BMIs in various categories. In subsection 4.3.3, the Gini coefficient estimates, and their sampling variances for BMI are considered, and in subsection 4.3.4, we provide CI estimates, along with results from hypotheses tests, for inferences about the inequality in BMI.

4.3.2 BMI Status in Bangladeshi Ever-Married Women of Age 15-49 Years

We report the percentages of women for each sample that fall within traditional WHO classifications of BMI in Table 11. For the three BDHS surveys (2004; 2007; and 2011), about six in ten ever-married women (57.29%; 58.42%; and 59.41%, respectively) have BMIs within the normal range (18.50 - < 25.00). The percentage of women with normal BMIs increases marginally from one survey to the next. This trend in normal BMI prevalence over time is statistically significantly different from zero at the 0.05 level of significance.[33]

A BMI < 18.5 is used to define underweight and it represents chronic energy deficiency or acute undernutrition. The prevalence of underweight women dropped significantly from 32.25% in 2004 to 22.92% in 2011. We also observe proportionally fewer women in each of the underweight categories, indicating that some gains may have been made in reducing malnutrition. This suggests that Bangladesh has made progress

---

[33] A chi-square statistic is used to test for the null hypothesis, where $\chi^2_{(1)}$=12.69 and p-value< 0.01.

towards its MDG 1, particularly Target 1C.[34] Nevertheless, the fact that 1 in 5 women were underweight in 2011 is still troubling, as noted in a report on Bangladesh's progress in reaching the MDGs (BPC, 2012).

Table 11 Sample sizes and percentages of ever-married Bangladeshi women across BMI categories for BDHS 2011, 2007 and 2004 surveys.

	Survey Year		
	2004	2007	2011
Number of observations	11,166	10,836	17,309
Underweight (<18.50)	32.25%	27.79%	22.92%
Severe thinness (<16.00)	5.27%	4.26%	3.49%
Moderate thinness (16.00-<17.00)	7.80%	6.57%	5.54%
Mild thinness (17.00-<18.50)	19.18%	16.96%	13.89%
Normal weight (18.50-24.99)	57.29%	58.42%	59.41%
Overweight (≥25.00)	10.46%	13.79%	17.67%
Pre-obese (25.00-<30.00)	8.71%	11.54%	14.52%
Obese (≥ 30.00)	1.75%	2.25%	3.15%
Obese class I (30.00-<35.00)	1.60%	1.92%	2.69%
Obese class II (35.00-<40.00)	0.10%	0.27%	0.36%
Obese class III (≥40.00)	0.10%	0.06%	0.09%

On the other hand, the proportion of being overweight has significantly increased over the same period: from 10.46% in 2004 to 17.67% in 2011. In addition, even though only a small percentage of women are obese, the rate of overweight women has nearly doubled over the time span. In 2011, the percentage of overweight women (17.67%) was nearly the same as the percentage of underweight women (22.92%). This increase in overweight women suggests that Bangladesh faces growing issues associated with being overweight, along with difficulties in decreasing chronic energy deficiency for a large proportion of women. This feature of simultaneously observing a significant percentage of the population being underweight and overweight coincides with evidence from other developing countries; for example, for Namibia (Arrar et al. 2009) and for Kenya (Jayne et al., 2011) (also see Razak et al., 2013; Prentice, 2006; and Caballero, 2005). Although

---

[34] 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" (see BPC, 2012).

an investigation of the specific causes of this trend for Bangladeshi women is beyond the scope of our study, we anticipate that growing urbanization, more sedentary lifestyles, and higher consumption of energy-dense foods, forms part of the reason, as has been hypothesized for other developing countries (e.g., Martorell et al., 2004). The double burden of women who are either underweight or overweight is a dilemma for the health care system.

Further information on the BMI status for Bangladeshi women is presented in Tables 12-14, where the sample data are divided by place of residence, wealth category and educational attainment, respectively. We detail the percentages in three broad BMI categories: underweight, normal weight, and overweight, for the three surveys, and show hypotheses test results for differences in proportions and trends over time.

Place of residence is an important indicator of living environment that may influence the health of a woman. Urban residency is often positively associated with a higher level of socioeconomic status (see, e.g., Neuman et al., 2013), and women residing in urban areas are more likely to have adequate dietary intake, access to health care services and basic sanitation facilities, compared to those living in rural areas (see, e.g., Borbor et al., 2014; UNICEF, 2009; Osório et al., 2001). As reported in Table 12, the prevalence of being underweight, an indicator of chronic energy deficiency, is more widespread for rural women than for women residing in urban regions. Although the trend in the percentage of underweight women over time has been declining, e.g., at a rate of 4.50% in each successive survey period, that is similar for both rural and urban women, of concern is the fact that one in four rural women is still underweight.

Table 12 Percentages of ever-married Bangladeshi women across BMI categories by place of residence and BDHS 2004, 2007 and 2011 surveys and hypothesis tests.

Category	Survey Year	Number of Observations	BMI Status		
			Underweight (<18.5)	Normal (18.5-<25)	Overweight (≥25)
Urban	2004	3,816	24.00%	56.30%	19.70%
	2007	4,118	19.50%	57.00%	23.50%
	2011	6,024	15.00%	56.90%	28.20%
Rural	2004	7,350	35.40%	58.80%	5.80%
	2007	6,718	32.00%	60.10%	8.00%
	2011	11,285	26.50%	61.30%	12.20%
Hypothesis tests					
$H_0: P_{2004}^{\text{urban}} - P_{2004}^{\text{rural}} = 0$					
$\chi^2(1)$			151.197	6.512	515.063
p-value			0.000	0.011	0.000
$H_0: P_{2007}^{\text{urban}} - P_{2007}^{\text{rural}} = 0$					
χ^2 (1)			201.338	10.225	513.722
p-value			0.000	0.001	0.000
$H_0: P_{2011}^{\text{urban}} - P_{2011}^{\text{rural}} = 0$					
χ^2 (1)			297.718	31.578	688.199
p-value			0.000	0.000	0.000
H_0: Trend in BMI prevalence in urban women over 2004-2011=0					
Slope			-0.045	0.003	0.043
χ^2_{trend}			125.846	0.305	93.978
p-value			0.000	0.581	0.000
H_0: Trend in BMI prevalence in rural women over 2004-2011=0					
Slope			-0.045	0.004	0.033
χ^2_{trend}			172.774	11.700	228.818
p-value			0.000	0.000	0.000

Notes: P is the population proportion or percentage. A chi-square statistic is used to test for the null hypothesis that the trend in proportion over the survey span is equal to zero.[35]

On the other hand, being overweight is significantly more prevalent among women living in urban areas, although the proportion of overweight women in rural areas is rising at a faster rate than that of urban areas. For example, during 2004 and 2011, the overweight prevalence among women in urban areas increased by 43%, whereas in rural areas it increased by over 110%. We also see that more rural women have normal BMIs, compared to urban women, and the trend for normal BMIs among rural women over time

---

[35] See https://ideas.repec.org/c/boc/bocode/s426101.html, for discussion on the Stata module developed by Patrick Royston in 2002, for trend analysis when using proportions; last accessed June 10, 2015

is positive. For urban women, no significant change was seen in normal BMIs over the studies period. Such features suggest that our research for Bangladesh is consistent with those of Corsi et al. (2011, p.637), who note a “distinct nutritional transition [may be occurring] between urban and rural”.

The proportions of women in the BMI classifications across wealth categories are presented in Table 13 for the three surveys. For the BDHS surveys, wealth, which is measured with an asset-based index, is used to indicate the relative socioeconomic status of the household. 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 the wealth index, constructed using principal components analysis, which combines weighted values for indicator variables that reflect economic status, including household assets (e.g., television, refrigerator, telephone, 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), the number of persons per sleeping room, ownership of agricultural land, and whether the household has one or more domestic servants (see, e.g., Rutstein and Johnson, 2004). Many studies link household wealth to health outcomes, such as prevalence of diseases, or availability and access to health services and mortality, via different channels such as nutrition, sanitation, education, information (see, e.g., Ahmmed, 2013; Corsi et al., 2011; Khan et al., 2011; Musgrove, 1996).

Table 13 Percentages of ever-married Bangladeshi women across BMI categories by wealth categories and BDHS 2011, 2007 and 2004 surveys, and hypothesis tests.

Category	Survey Year	Number of Observations	BMI Status		
			Underweight (<18.5)	Normal (18.5-<25)	Overweight (≥25)
Poorest	2004	1,979	45.30%	25.70%	2.00%
	2007	1,751	41.70%	54.90%	3.40%
	2011	3,013	38.30%	56.70%	5.00%
Poorer	2004	1,997	39.50%	57.80%	2.80%
	2007	1,964	34.30%	61.50%	4.20%
	2011	3,235	28.30%	64.50%	7.20%
Middle	2004	2,092	34.60%	60.40%	5.00%
	2007	2,070	30.60%	62.40%	7.00%
	2011	3,328	24.00%	64.00%	12.00%
Richer	2004	2,230	29.80%	61.00%	9.20%
	2007	2,173	25.20%	62.90%	11.9%
	2011	3,670	18.70%	61.10%	20.20%
Richest	2004	2,868	15.60%	57.50%	27.00%
	2007	2,878	12.80%	54.00%	33.20%
	2011	4,063	8.40%	53.60%	38.00%
Hypothesis tests					
H_0: There is no relationship between BMI prevalence and wealth in 2004					
Slope			-0.071	0.061	0.061
χ^2 (p-value)			546.345(*)	346.751(*)	900.437(*)
H_0: There is no relationship between BMI prevalence and wealth in 2007					
Slope			-0.069	-0.005	0.074
χ^2 (p-value)			528.328(*)	2.099(0.147)	1003.385(*)
H_0: There is no relationship between BMI prevalence and wealth in 2011					
Slope			-0.070	-0.012	0.082
χ^2 (p-value)			969.041(*)	20.938(*)	1596.298(*)
H_0: Trend in BMI prevalence in poorest women over 2004-2011 =0					
Slope			-0.035	0.147	0.015
χ^2_{trend} (p-value)			24.500(*)	420.850(*)	30.707(*)
H_0: Trend in BMI prevalence in poorer women over 2004-2011 =0					
Slope			-0.056	0.033	0.023
χ^2_{trend} (p-value)			71.824(*)	23.530(*)	52.035(*)
H_0: Trend in BMI prevalence in middle wealth group women over 2004-2011 =0					
Slope			-0.054	0.018	0.036
χ^2_{trend} (p-value)			73.980(*)	7.110(0.008)	85.327(*)
H_0: Trend in BMI prevalence in richer women over 2004-2011 =0					
Slope			-0.056	-0.001	0.057
χ^2_{trend} (p-value)			111.296(*)	0.014(0.905)	143.926(*)
H_0: Trend in BMI prevalence in richest women over 2004-2011 =0					
Slope			-0.036	-0.019	0.055
χ^2_{trend} (p-value)			86.474(*)	9.581(0.002)	90.926(*)

Notes: * indicates a p-value <=0.001; other p-values are reported in the table. A chi-square statistic is used to test for the null hypothesis that the trend in proportion over the survey span is equal to zero.

The percentages of women in the BMI classifications across wealth categories and the survey years appear to vary substantially. Close to one third of *poor* women are underweight, with only modest declines reported 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-year period covered by the three surveys. Although a woman in the *middle* wealth group is less likely to be underweight, in contrast to her poorer counterpart, chronic energy deficiency is still substantially prevalent for the *middle* wealth category of women. In contrast, *rich* women are now equally likely to be overweight or 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. This implies that a woman from the richest quintile is seven times more likely to be overweight and four times less likely to be undernourished compared to a woman from the lowest wealth quintile. We also note that more than half of the women in all wealth categories have BMIs within the normal category. This suggests that malnutrition (underweight or overweight) is of concern for just under half of the women.

The hypothesis test results in Table 13 shows that the prevalence of being underweight (overweight) and the wealth category are significantly negatively (positively) related in each survey year. No such dominant pattern is seen for women of normal weight. In recent years, the consequences of the decreasing proportions in both underweight and normal-weight, in relation to wealth level, is causing an increased incidence of being overweight across wealth categories. On the other hand, the trend in the percentage of underweight women in each wealth group is declining over 2004-2011, whereas, for overweight women, the trend is positive for each wealth category. In addition, the significant decrease in prevalence of being normal weight among the richest women indicates the increasing burden of overweight in the wealthiest class in the country for our studied time frame.

In exploring the BMI data, we classify women according to their educational attainment, where primary and secondary completions are defined as completing grade 5 and grade 10, respectively. The positive interaction between education and a woman's health is well recognized (see, e.g., Maddah et al., 2003; Neuman, 2013; Nour, 2008; Ross and Wu, 1995). As such, a negative correlation may also exist between education level and the prevalence of being underweight or overweight for Bangladeshi women.

As shown by the outcomes reported in Table 14, more than half of the women have normal BMIs, irrespective of educational attainment, and no noticeable change occurs in the percentage of normal BMI over the span of the surveys. The occurrence of being underweight is highest among women with no education, whereas the proportion of those who are overweight or obese is highest among women with secondary or higher education. In general, roughly, a woman with secondary education complete or a higher level of education is three times less likely to be underweight, and three times more likely to be overweight, when compared to a woman with no education.

We see that the percentage of undernourished women declined over time (2004-2011), irrespective of educational attainment, whereas the trend is positive for overweight women over the period. Although nearly three times as many uneducated women are underweight than overweight women, the rate of the increasing proportion of overweight women in the uneducated category is faster than the decreasing proportion of underweight women, which raises some concerns about their health.

Table 14 Percentages of ever-married Bangladeshi women across BMI categories by educational attainment and BDHS 2011, 2007 and 2004 surveys, and hypothesis tests.

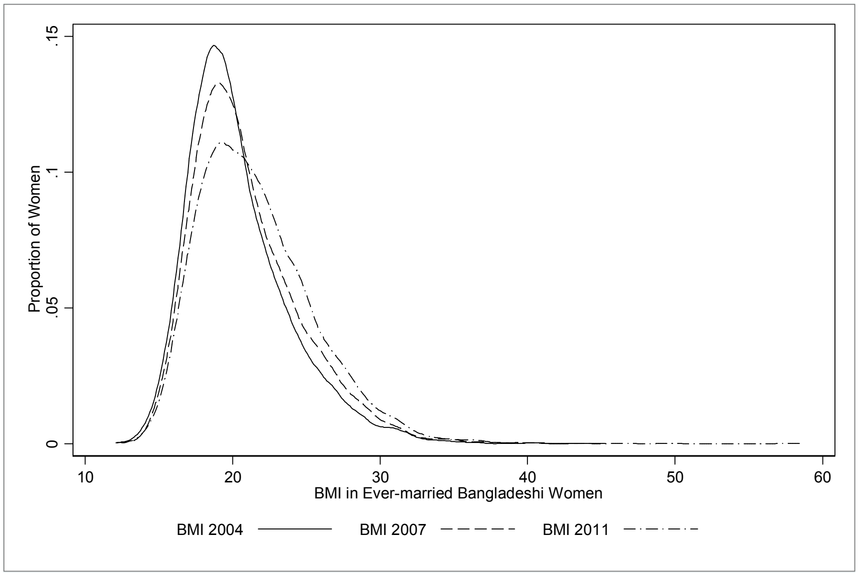
Category	Survey Year	Number of Observations	BMI Status		
			Underweight (<18.5)	Normal (18.5-<25)	Overweight (≥25)
No education	2004	4,333	39.42%	55.46%	5.12%
	2007	3,466	36.70%	56.03%	7.27%
	2011	4,532	29.57%	58.80%	11.61%
Primary incomplete	2004	2,268	34.57%	57.67%	7.76%
	2007	2,262	30.77%	59.86%	9.37%
	2011	3,133	26.08%	60.96%	12.96%
Primary complete	2004	1,019	29.93%	59.28%	10.79%
	2007	950	26.63%	59.68%	13.69%
	2011	2,038	24.79%	59.42%	15.79%
Secondary incomplete	2004	2,566	25.80%	60.09%	14.11%
	2007	2,608	22.01%	61.31%	16.68%
	2011	5,345	20.64%	60.60%	18.76%
Secondary or Higher	2004	980	14.39%	55.31%	30.30%
	2007	1,528	13.81%	55.96%	30.23%
	2011	2,261	8.98%	55.68%	35.34%
Hypothesis tests					
H_0: There is no relationship between BMI prevalence and education in 2004					
Slope			-0.053	0.008	0.045
χ^2 (p-value)			289.285(*)	6.188(0.013)	482.279(*)
H_0: There is no relationship between BMI prevalence and education in 2007					
Slope			-0.053	0.006	0.048
χ^2 (p-value)			335.015(*)	3.079(0.079)	453.152(*)
H_0: There is no relationship between BMI prevalence and education in 2011					
Slope			-0.041	-0.003	0.044
χ^2 (p-value)			338.243(*)	1.134 (0.287)	469.802(*)
H_0: Trend in BMI prevalence in women with no education over 2004-2011 =0					
Slope			-0.049	0.017	0.033
χ^2_{trend} (p-value)			94.973(*)	10.159(*)	125.753(*)
H_0: Trend in BMI prevalence in women with primary incomp educ over 2004-2011 =0					
Slope			-0.043	0.016	0.026
χ^2_{trend} (p-value)			45.988(*)	5.774(0.016)	40.010(*)
H_0: Trend in BMI prevalence in women with primary comp educ over 2004-2011 =0					
Slope			-0.025	0.000	0.025
χ^2_{trend} (p-value)			8.970 (0.003)	0.002(0.964)	14.108(*)
H_0: Trend in BMI prevalence in women with secondary incom educ over 2004-2011 =0					
Slope			-0.025	0.002	0.023
χ^2_{trend} (p-value)			25.193(*)	0.079(0.778)	26.845(*)
H_0: Trend in BMI prevalence in women with secondary com/higher over 2004-2011 =0					
Slope			-0.030	0.001	0.029
χ^2_{trend} (p-value)			25.418(*)	0.017(0.897)	10.953(*)

Notes: * indicates a p-value <=0.001; other p-values are reported in the table. A chi-square statistic is used to test for the null hypothesis that the trend in proportion over the survey span is equal to zero.

For the 2004 and 2007 surveys, the percentages of women with normal BMIs statistically increase with educational attainment, at least at the 10% significance level. For 2011, however, we fail to reject the null hypothesis of a difference across the education groups at any of the usual levels of significance. On the other hand, at the 5% nominal significance level, with a one-sided test, the proportion of being underweight (overweight) decreases (increases) significantly with the higher educational attainment, for each of the surveys. These results may suggest that education improves a woman's knowledge of nutrition for improving health status or realizing the consequences of being underweight or undernourished, but education does not seem to help women to be concerned about being overweight. Similar findings have been reported for Bangladeshi women in previous studies (e.g., Hossain et al., 2012; Corsi et al., 2011).

We now discuss some mean statistics and empirical distributions for BMI. The mean BMIs for all women are 20.26, 20.73 and 21.44 kg/m^2 in the 2004, 2007 and 2011 survey years, respectively. Evidently, the mean BMI has increased over time with slightly more variation, as indicated by the increased standard deviation (see Table D.1 in Appendix D). In addition, when BMI data is plotted (Figure 9), we can see that each distribution is positively skewed with higher excess kurtosis (>3.00) than would be associated with a normally distributed random variable. As noted in Table 11, the presence of some Class III obese women ($BMI \geq 40$) causes the distributions of BMI to have long right tails. Indeed, from the so-called Jarque-Bera test (JB) (Bera and Jarque, 1981), the null hypothesis is strongly rejected that BMI is normally distributed, irrespective of survey.

Figure 9 BMI distributions in Bangladeshi women age 15-49: BDHS 2004, 2007 and 2011



Notes: Hypothesis tests suggest that the distributions are positively skewed.[36]

When women are classified according to their place of residence, the mean BMIs for urban women are 21.59, 22.29 and 23.03 for the 2004, 2007 and 2011 surveys, respectively, which are statistically significantly higher than for rural women by 8.6%, 9.9%, and 10.2%, respectively. Typically, most rural women are involved in physical activities such as manual and agricultural work, and they may have a high level of psychological and physiological stress (e.g., Hossain et al., 2012; Subramanian and Smith, 2006). These factors along with inadequate access to food and health facilities may cause a general lowering of mean BMI for rural women, compared to urban woman. Nevertheless, some overweight women are present in both urban and rural areas, as demonstrated by the positively skewed empirical distributions of the BMIs (see Figure D.1 in Appendix D). The higher variability in the urban women distributions suggests the presence of more overweight women in urban areas, compared to rural areas.

---

[36] Test statistics with p-values and skewness and kurtosis are: JB=4434.933 (0.000), skewness= 1.143, kurtosis=5.075 (for 2004), JB=3327.605 (0.000), skewness= 1.059, kurtosis=4.698 (for 2007) and JB=5639.735 (0.000), skewness= 0.952, kurtosis=5.047 (for 2011).

When the sample data are divided by wealth category, we see that the mean BMIs for the different categories of wealth are significantly different, and the mean BMI increases across wealth categories (see Table D.2 in Appendix D). Since a woman could be underweight in a *rich* household or overweight in a *poor* household, nutritional status is determined by more than just wealth status. Nevertheless, these findings suggest that a strong positive correlation does exist between wealth and BMI, and that malnutrition does not occur randomly across the population. Moreover, the empirical distributions of the BMI data for the five wealth categories for three surveys are positively skewed, when presented by the boxplots (see, Figure D.2 in Appendix D).[37] A longer upper whisker and a large number of outliers on the upper side for each box suggest that in each wealth category there is some incidence of obesity, which is independent of the household's wealth level.

Lastly, we observe similar results, when the BMI data are sub-grouped by educational attainment to those when the data are classified by wealth – mean BMI and level of education are positively related in every survey (see Table D.3 in Appendix D). The mean BMIs, across the five levels of educational attainment, are significantly different from each other, irrespective of the survey, and the trends in mean BMIs over the survey period are statistically significantly upward for each level of educational attainment. On the other hand, the empirical distributions of BMI for each education level are positively skewed, with the variability in distribution being positively associated with education – more educated women have a larger spread in BMI (see Figure D.3 in Appendix D). Thus, the variability in weight increases with education.

---

[37] Boxplots illustrate the value spread, and whether or not unusual observations occur in the sample data, based on the median. The boxplot technique is often used to identify the shape of the distribution without hypothesis testing for normality, and is useful when the size of the sample data is large or when two or more data sets are being compared.

4.3.4 Gini Coefficient Estimates and Sampling Variances

In the previous subsections, we discussed health status in terms of the percentages of women with BMIs above or below a specified threshold. Such statistics, while useful, only provide some indication of the variation in the distribution of BMI. We estimate the Gini coefficient for different categories to measure the disparity in the distributions. Recall, when $G = 0$, perfect equality occurs, that is, all women have exactly the same BMI. In contrast, when $G = 1$, the unrealistic case of perfect inequality occurs; i.e., one woman has all the BMI. Our interest is in summarizing how health status is distributed among women, and G is a way to summarize how equally the total BMI health is distributed.

In addition, to undertake statistical inferences, we provide sampling variances for $\hat{G}$ using expressions (4.2), (4.3) and (4.4), which correspond to those obtained by Binder and Kovačević (1995), Bhattacharya (2007), and the bootstrap variance, respectively. We also report the breakdown of the variance estimates (for only Bhattacharya's formula using expression (4.3)) into three components: variance arising from naively assuming a simple random sample (SRS) design with weights, the effect of clustering, and the effect of stratification on the variance. The design effect arising from the complex survey design is estimated as the ratio of the left-side and the first component on the right-side of expression (4.3). The Wald statistics, and associated asymptotic chi-square p-values, for testing the equality of 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 the weights variance formula[38] are also provided. We first present and discuss the results for the full BMI sample (all women),

---

[38] We assume that the samples across the surveys are independent, a reasonable assumption given that clusters are randomly selected in each survey.

and then consider the outcomes when women are grouped by place of residence, wealth, and education attainment.

4.3.4.1 Gini Coefficient Estimates and Sampling Variances: All Women

The Gini coefficient estimates for all women for each survey and the estimated variances using expressions (4.2), (4.3), and (4.4), and the associated hypothesis test results are presented in Table 15. The table also provides the breakdown of Bhattacharya's variance, as given in expression (4.3), along with design effects. Hypothesis tests are undertaken using standard errors from three variance estimates: Binder and Kovačević's expression $\left(\widetilde{Var_{BK}}(\hat{G})\right)$, Bhattacharya's expression $\left(\widetilde{Var_{BT}}(\hat{G})\right)$, and the simple random sample variance $\left(\widetilde{Var_{SRS}}(\hat{G})\right)$. In some cases, we also provide hypothesis tests formed from bootstrap standard errors. This enables us to ascertain the impact that the variance estimate has on the outcome of the hypothesis test.

At the 5% nominal significance level, the increase in the Gini estimates over the three surveys is statistically significant. As variations in women's BMI will always occur, even in a healthy population with normal weight, we calculated Gini estimates for those women whose standardized BMIs were in the *normal* range (18.50-<25.00) to ascertain how our Gini estimates differ from such a sample. For the three survey years, these Gini estimates are, respectively, 0.048, 0.047, and 0.046; we see that the observed estimates for the whole sample are approximately double these values. Malnutrition (under and over) results in significant increases in inequality. How do these estimates compare with other countries? Interestingly, the absolute magnitudes of the Gini estimates for Bangladesh are approximately 30% lower than the Gini estimate reported by Araar et al. (2009) for Namibian females. These indices are also lower than those reported by

Contoyannis and Wildman (2007) for females from two developed countries, Canada and England.

Table 15 Gini coefficient estimates and variances, design effects and hypothesis tests for BMI: ever-married Bangladeshi women aged 15-49 - BDHS 2004, 2007 and 2011

Survey year	$\hat{G}$	$\widehat{Var_{BK}}(\hat{G})$	$\widetilde{Var_{BT}}(\hat{G})$	$\widehat{V_{BOOT}}(\hat{G})$
2004	0.088	0.880	0.806	0.763
2007	0.092	1.114	1.056	1.068
2011	0.097	0.607	0.589	0.500

$\widetilde{Var_{BT}}(\hat{G})$ *Breakdowns and Design Effects*

Survey year	$\widetilde{Var_{SRS}}(\hat{G})$	Stratum Effect	Cluster effect	Design Effect
2004	0.529	0.644	0.922	1.524
2007	0.591	0.471	0.936	1.787
2011	0.369	0.170	0.390	1.596

Hypothesis Tests

Null Hypothesis	Change in $\hat{G}$	Using $\widehat{Var_{BK}}(\hat{G})$		Using $\widetilde{Var_{BT}}(\hat{G})$		Using $\widetilde{Var_{SRS}}(\hat{G})$	
		Wald statistic	p-value	Wald statistic	p-value	Wald statistic	p-value
$H_0: G_{2007} = G_{2004}$	0.004	7.921	0.005	8.593	0.003	14.286	0.000
$H_0: G_{2011} = G_{2007}$	0.005	14.310	0.000	15.198	0.000	26.042	0.000
$H_0: G_{2011} = G_{2007} = G_{2004}$	n.a.	55.645	0.000	59.096	0.000	92.297	0.000

Notes: Variance estimates and associated components have been scaled by 10^6. $\widehat{Var_{BK}}(\hat{G})$ is Binder and Kovačević's (1995) estimator, see expression (4.2), $\widetilde{Var_{BT}}(\hat{G})$ is Bhattacharya's (2007) estimator, see expression (4.3) and $\widehat{V_{BOOT}}(\hat{G})$ is the bootstrap variance, see expression (4.4). $\widetilde{Var_{SRS}}(\hat{G})$ 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 $\widetilde{Var_{BT}}(\hat{G})$ 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 using Bhattacharya's (2007) and Binder and Kovačević's (1995) formulae, with $\widehat{Var_{BK}}(\hat{G})$ being between 3% and 10% higher than $\widetilde{Var_{BT}}(\hat{G})$. As some strata have few clusters, the correction factor $(n_h/(n_h - 1))$, discussed in Chapter 2, can markedly differ from unity for these strata, implying that for samples that have relatively few clusters in some of the strata, the asymptotic variance, $\widetilde{Var_{BT}}\ (\hat{G})$, may understate the variance. Then, in practice, it may be preferable to adopt $\widehat{Var_{BK}}(\hat{G})$. On the other hand, the bootstrap variances,

$\widehat{V_{BOOT}}(\hat{G})$, computed from 799 replications, are 4% to 17% smaller than $\widehat{Var_{BK}}(\hat{G})$, but they are not always smaller than $\widehat{Var_{BT}}\ (\hat{G})$. This suggests that some gains in precision may be realized by using a bootstrap technique. Nevertheless, in Chapter 3, we saw that when the number of clusters in the sample increases, these three variance estimators become similar to each other.

Allowing for the survey design[39] 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 (see NIPORT 2013, 2009, 2005). These results show the importance of accounting for the survey design when estimating variances. We see the markedly 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 we compare the variances, and the design components. Over 50% more women were surveyed for the 2011 survey than in earlier surveys and the number of clusters sampled nearly doubled. A 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. In contrast, sampling more clusters likely adds useful information that is reflected in a smaller cluster effect for the 2011 survey. As expected, the SRS variance estimate falls with a larger number of women being sampled.

Apart from design components, as seen in our MC simulations in Chapter 3, the intracluster correlation coefficient (ICC), ρ_h, affects the design effect. The larger the ICC values across strata, the larger the design effect. Clustering reduces the survey cost but

---

[39] In some studies, the design effect is relative to the SRS variance without weights; i.e., a sample that assumes that 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 $\widehat{Var_{SRS}}(\hat{G})$. Such an approach also enables a straightforward decomposition of the overall variance.

provides less precise estimates by increasing the correlation between observations within clusters, which is measured by ρ_h. For the DHS surveys, when the sampling design is implemented, precautions are taken to keep ICCs across strata as small as possible (see, e.g., Aliaga and Ren, 2006; UN, 2005), but we still see the impact showing up here.

The estimated ICCs, $\hat{\rho}_h$, in the three BMI samples along with the within-cluster and the between-cluster variance components vary substantially from strata to strata (see, Tables D.4-D.6 in Appendix D). For example, in the 2004 survey, the ICC estimates range from 0.000 to 0.233. Similar variations in ICC values are observed for the 2007 and 2011 surveys. These components inform us of how similar women's BMIs are within clusters and between clusters in a stratum.

An ICC of 0 suggests that women's BMIs within clusters are no more similar to each other than BMIs from different clusters (i.e., no between-cluster variability).[40] When BMIs among women within clusters are more similar to each other compared to that between clusters within a stratum, the ICC estimates are positive; and the homogeneity in women's BMIs rises within clusters for larger values of ICCs across strata. We see that the ICC estimates are large for city or metropolitan city strata. A larger ICC value may imply that households living in a city block (cluster) have similar socioeconomic status, and hence the BMIs among women, but they are different from other city blocks (clusters) within the city.

In our study, taking account of the survey design on the variances of $\hat{G}$ does not alter the quantitative outcomes of the hypothesis tests. Here, we reject the hypothesis that the Gini coefficients are equal from one survey to the next and that they are jointly equal across the three surveys. A significant increase in BMI inequality is seen over the three

---

[40] In the ICC calculation, when the mean squares within is greater than the mean squares between clusters (i.e., $MSB_h < MSW_h$), the between-cluster variance is truncated at zero in the STATA software.

surveys. This finding is in contrast with changes in Gini coefficient estimates for inequality in income over the same period, which shows a small decline.[41]

4.3.4.2 Gini Coefficient Estimates and Sampling Variances: Place of Residence

Gini coefficient estimates and their estimated variances when the women are divided by their place of residence (urban; rural) are reported in Table 16. The table also provides estimated variance components, design effects, and hypothesis tests for equality of G across survey years and across place of residence for a given survey year. For the former tests we provide outcomes using $\widehat{Var_{BK}}(\hat{G})$, $\widetilde{Var_{BT}}(\hat{G})$, and $\widetilde{Var_{SRS}}(\hat{G})$, but for the latter tests we only report findings using $\widetilde{Var_{BT}}(\hat{G})$ - qualitative outcomes do not change when using the other variance estimates.

For urban women, the nutritional inequalities are significantly higher than for rural women. Inequality has increased considerably among women living in rural regions across the surveys, whereas it remained relatively stable among urban women, even declining marginally from 2007 to 2011. Socioeconomic development of rural women over the period may have played a role in shifting the patterns of the BMI distribution (see, e.g., Corsi et al., 2011; Subramanian and Smith, 2006) resulting in an increase in the inequality in BMI for rural women.

Variance estimates are substantially higher for the urban Gini coefficient estimates than for 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 in each survey. In addition, design effects are relatively

---

[41] As reported by The World Bank; see http://data.worldbank.org/indicator/SI.POV.GINI/countries?display=default, last accessed 27 January 2016.

larger for rural areas than urban ones, suggesting, ceteris paribus, that a higher positive ICC value exists for rural women than for urban women with respect to BMI.[42]

Table 16 Gini coefficient estimates and variances, design effects and hypothesis tests for BMI among ever-married Bangladeshi women age 15-49: by place of residence, BDHS 2004, 2007 and 2011

Survey year	Category	$\hat{G}$	$\widehat{Var}_{BK}(\hat{G})$	$\widetilde{Var}_{BT}(\hat{G})$	$\widetilde{V}_{BOOT}(\hat{G})$
2004	Urban	0.104	2.422	1.981	1.810
	Rural	0.081	1.214	1.188	1.182
2007	Urban	0.104	2.402	2.169	2.184
	Rural	0.085	1.733	1.689	1.739
2011	Urban	0.101	1.814	1.736	1.714
	Rural	0.092	0.750	0.738	7.507

$\widetilde{Var}_{BT}(\hat{G})$ breakdowns and design effects

Survey year	Category	$\widetilde{Var}_{SRS}(\hat{G})$	Stratum effect	Cluster effect	Design effect
2004	Urban	1.806	0.659	0.744	1.097
	Rural	0.691	0.064	0.561	1.719
2007	Urban	1.659	0.302	0.821	1.307
	Rural	0.843	0.006	0.853	2.004
2011	Urban	1.307	0.243	0.672	1.328
	Rural	0.470	0.030	0.298	1.570

Hypothesis tests

Equality of Gini across surveys

Null Hypothesis	Category	Using $\widehat{Var}_{BK}(\hat{G})$		Using $\widetilde{Var}_{BT}(\hat{G})$		Using $\widetilde{Var}_{SRS}(\hat{G})$	
		Wald statistic	p-value	Wald statistic	p-value	Wald statistic	p-value
$H_0: G_{2007} = G_{2004}$	Urban	0.087	0.768	0.101	0.751	0.121	0.728
	Rural	5.429	0.020	5.561	0.018	10.432	0.001
$H_0: G_{2011} = G_{2007}$	Urban	2.577	0.108	2.783	0.095	8.314	0.004
	Rural	19.734	0.000	19.943	0.000	37.319	0.000
$H_0: G_{2011} = G_{2007} = G_{2004}$	Urban	3.016	0.221	3.259	0.196	4.229	0.121
	Rural	65.177	0.000	65.196	0.000	103.998	0.000

Equality of Gini across place of residence

	2004		2007		2011	
	Wald statistic	p-value	Wald statistic	p-value	Wald statistic	p-value
$H_0: G_{\text{urban}} = G_{\text{rural}}$	161.886	0.000	96.514	0.000	32.348	0.000

Notes: Variance estimates have been scaled by 10^6. $\widehat{Var}_{BK}(\hat{G})$ is Binder & Kovačević's (1995) estimator (expression (4.2)), $\widehat{Var}_{BT}(\hat{G})$ is Bhattacharya's (2007) estimator (expression (4.3)); $\widehat{Var}_{SRS}(\hat{G})$ is the first term of expression (4.4), the variance estimator under an assumption of SRS with weights. The "design effect" provides the ratio of $\widehat{Var}_{BT}(\hat{G})$ to $\widehat{Var}_{SRS}(\hat{G})$. For $H_0: G_{\text{urban}} = G_{\text{rural}}$, the statistics are formed using $\widehat{Var}_{BT}(\hat{G})$.

---

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

Two-sided hypothesis tests for urban women suggest equality of G between 2004 and 2007, but with the change in inequality between 2007 and 2011 being significantly different from zero, at least at the nominal 10% significance level. However, the inequality change between 2007 and 2011 is highly significant when the test is formed using $\widetilde{Var_{SRS}}(\hat{G})$, where the p-value <0.01, which implies that the cluster effect is larger than the stratum effect so that the simple random sample variance is far smaller than the design based variance. Overall, we conclude that the urban G are jointly equal across the three surveys. In contrast, the changes in G between surveys for rural women are statistically significantly different from zero. In addition, the urban and rural G are unequal for each survey. These findings suggest that BMI inequality among rural women is significantly increasing over time, and differs from that for urban women.

4.3.4.3 Gini Coefficient Estimates and Sampling Variances: Wealth Categories

Here, we examine BMI inequality among women when they are grouped by wealth category: poorest; poorer; middle; richer; and richest. Estimation in this subsection is different from previous subsections – results are reported after correcting for the survey design. For reasons discussed below, we ignore that the data are from clusters, and instead estimate G and associated variances assuming only a stratified sampling design.

When the average cluster size in the sample is unusually small, estimates, especially variance estimates, under the complex survey design assumption may not be reliable (see, e.g., Clarke and Wheaton, 2007; Mass and Hox, 2005; UN, 2005, p. 51). We face this situation when the BMI data are divided into the five wealth categories. The average number of women in a cluster ranges from 6 to 11, and about 6 to 16% of the clusters contain just one woman. Of the clusters, 20 to 35% consist of one to three women, and only 25 to 41% have at least 10 women (see Table D.7 in Appendix D).

Despite the lack of firm recommendations on cluster sizes needed for reliable estimation of a parameter, several rules are available, with suggestions ranging from 15 to 30 observations per cluster. For instance, Bell et al. (2010) propose a minimum cluster size of 30 observations to examine the within-cluster variations; Hox (1998) suggests a minimum cluster size of 20 to examine interactions among observations in a sample consisting of 50 clusters. Using MC simulations, Clarke and Wheaton (2007) show that estimators are subatintially biased in samples with a large number of clusters having a single observation or the average cluster size is less than three. They also find that true population values for the variance components cannot be replicated with cluster sizes of 2 or less.

Maas and Hox (2004, 2005) ascetain that cluster sizes of 20 to 100 are necessary to obtain less biased variance estimators from clustered samples. Indeed, for the DHS surveys, cluster sizes of 20 observations is considered an optimal size and keeps the design effect at a tolerable level, while also minimizing the survey costs; e.g., see Aliaga and Ren (2006). Clearly, our samples, when divided by wealth categories, do not have cluster sizes anywhere near these recommendations.

When estimated using the expression in (4.3), the contribution of clustering to the overall variance of $\widehat{G}$ in a number of cases is found to be negative in the wealth category. For example, in the 2004 poorest sample, the estimated variance for $\widehat{G}$ is 2.17×10^{-06}, where the cluster effect is -1.25×10^{-06}; and in the 2011 poorer sample, the variance estimate is 1.46×10^{-06} with the cluster effect contribution being -2.413×10^{-08}. Although, a negative cluster effect is not ruled out in theory, it is almost always positive in survey applications (see, e.g., Kalton et al., 2005, p.106; Skinner et al., 1989, p. 37).

We ascertain that such situations probably arise due to the limited number of women in many of the clusters when subgroups are explored. As such, we ignore the clustering from our sampling design. Accordingly, we estimate inequalities along with their sampling variances for BMIs for all five wealth categories assuming stratified samples with weights.

When estimating G and their variances, ignoring clustering but including stratification and weights, we anticipate smaller sampling variances for $\hat{G}$ than that obtained from simple random samples; hence, the estimated design effects are less than one. In Table 17, we report $\hat{G}$ with associated variances and design effects under the stratified sampling design assumption. Irrespective of survey, inequality in women's BMI increases with socioeconomic status, as measured by wealth asset status (see Figure D.4 in Appendix D). This is not unexpected as an increasing mean BMI in relation to wealth status is associated with higher variability in the associated distributions, as measured by the standard deviations (see Table D.2 in Appendix D). Other factors such as place of residence may influence the sampling variability. For example, the mean BMI of urban women in the *richest* category is significantly higher than that for rural women.

When estimated from stratified samples, the variance estimate using $\widetilde{Var_{BT}}(\hat{G})$ is slightly larger than that using the $\widehat{Var_{BK}}(\hat{G})$ formula due to the correction factor associated with the former. However, we know that this diffference will vanish with an increasing sample size, measured by the number of women in the stratum. Alternatively, the bootstrap variance estimates are not always smaller than the two types of variance estimates; indeed, the bootstrap variance is up to 20% larger in some cases.

Table 17 BMI Gini coefficient estimates and variances by wealth category: ever-married Bangladeshi women aged 15-49 - BDHS 2004, 2007, and 2011 surveys.

Survey year	Wealth category	$\hat{G}$	$\widehat{Var}_{BK}(\hat{G})$	$\widetilde{Var}_{BT}(\hat{G})$	$\widehat{Var}_{BOOT}(\hat{G})$	$\widetilde{Var}_{BT}(\hat{G})$ breakdown		Design effect
						$\widetilde{Var}_{SRS}(\hat{G})$	Stratum effect	
2004	Poorest	0.070	2.419	2.372	2.206	2.416	0.045	0.982
	Poorer	0.073	2.379	2.366	2.514	2.387	0.021	0.991
	Middle	0.078	2.250	2.236	1.824	2.248	0.012	0.995
	Richer	0.088	2.451	2.432	2.842	2.458	0.026	0.989
	Richest	0.100	1.880	1.868	1.846	1.900	0.032	0.983
2007	Poorest	0.073	2.880	2.862	2.584	2.881	0.019	0.993
	Poorer	0.075	2.205	2.193	2.179	2.206	0.014	0.994
	Middle	0.081	3.179	3.162	3.232	3.173	0.011	0.997
	Richer	0.089	2.485	2.468	2.467	2.487	0.019	0.992
	Richest	0.102	2.239	2.223	2.172	2.245	0.022	0.990
2011	Poorest	0.081	1.535	1.528	1.556	1.535	0.007	0.995
	Poorer	0.081	1.515	1.509	1.588	1.514	0.005	0.997
	Middle	0.088	1.436	1.430	1.297	1.435	0.005	0.997
	Richer	0.095	1.458	1.452	1.361	1.468	0.017	0.989
	Richest	0.098	1.688	1.681	1.669	1.687	0.006	0.996

Notes: Variance estimates have been scaled by 10^6. $\widehat{Var}_{BK}(\hat{G})$ $\left(\widetilde{Var}_{BT}(\hat{G})\right)$ is Binder & Kovačević's (1995) (Bhattacharya's (2007)) estimator and $\widehat{Var}_{BOOT}(\hat{G})$ is the bootstrap variance estimator using expression (4.4) for 799 replications. $\widetilde{Var}_{SRS}(\hat{G})$ 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 $\widetilde{Var}_{BT}(\hat{G})$ to $\widetilde{Var}_{SRS}(\hat{G})$.

The breakdowns of $\widetilde{Var_{BT}}(\hat{G})$ into its components and associated design effects show that the contribution of stratification to variance estimation is relatively small, with the design effects being less than, but very close to, one. For a large complex survey sample with only a few strata, the precision gain from stratification is small, suggesting women's BMIs within a stratum are not more similar than those in other stratum.

We also report Wald statistics, and corresponding asymptotic chi-square p-values to examine for equality of Gini coefficients across the three surveys by wealth category using the three different variances in Table 18. Hypothesis test results show that the change in inequality in women's BMIs between 2004 and 2007 is not statistically significantly different from zero at any level of wealth. Changes in inequalities between 2007 and 2011, however, are significantly different from zero for all wealth categories.[43] In addition, at the nominal 5% significance level, using a two-sided test, except for the richest women, we reject the null hypothesis that G are the same across the surveys for each wealth category. The inequality among richest women fell significantly in 2011 from that of 2007. Although the BMI inequality in this category is still higher than in other categories, a falling G indicates that more women in the richest category are converging into the overweight group (see Table 13) and the variability among their weights is decreasing. We notice the same qualitative outcome in hypothesis testing irrespective of the variance estimate being adopted.

---

[43] At the nominal 5% significance level. Note that for some of the cases, we would conclude a statistical significance in the change in G if we had we examined an appropriate one-sided alternative hypothesis.

Table 18 Hypothesis tests for BMI inequality by wealth category for ever-married Bangladeshi women aged 15-49 -BDHS 2004, 2007 and 2011 surveys

Null Hypothesis	Variance used	Test statistic	Wealth category				
			Poorest	Poorer	Middle	Richer	Richest
$H_0: G_{2007} = G_{2004}$	$\widehat{Var_{BK}}(\hat{G})$	Wald statistic	1.720	0.873	1.658	0.203	0.971
		p-value	0.190	0.350	0.197	0.653	0.324
	$\widetilde{Var_{BT}}(\hat{G})$	Wald statistic	1.698	0.877	1.662	0.204	0.978
		p-value	0.192	0.349	0.197	0.651	0.323
	$\widetilde{Var_{SRS}}(\hat{G})$	Wald statistic	1.879	0.871	1.660	0.202	0.965
		p-value	0.170	0.351	0.198	0.653	0.326
$H_0: G_{2011} = G_{2007}$	$\widehat{Var_{BK}}(\hat{G})$	Wald statistic	14.496	9.677	10.618	9.130	4.484
		p-value	0.000	0.002	0.001	0.003	0.034
	$\widetilde{Var_{BT}}(\hat{G})$	Wald statistic	14.579	9.724	10.657	9.184	4.098
		p-value	0.000	0.002	0.001	0.002	0.043
	$\widetilde{Var_{SRS}}(\hat{G})$	Wald statistic	15.459	9.677	10.634	9.102	4.069
		p-value	0.000	0.002	0.001	0.003	0.044
$H_0: G_{2011} = G_{2007} = G_{2004}$	$\widehat{Var_{BK}}(\hat{G})$	Wald statistic	34.244	19.099	29.506	15.914	4.102
		p-value	0.000	0.000	0.000	0.000	0.128
	$\widetilde{Var_{BT}}(\hat{G})$	Wald statistic	34.618	19.099	29.659	16.001	4.126
		p-value	0.000	0.000	0.000	0.000	0.127
	$\widetilde{Var_{SRS}}(\hat{G})$	Wald statistic	35.569	19.082	29.530	15.840	4.097
		p-value	0.000	0.000	0.000	0.000	0.129

On the other hand, inequalities in BMIs across the wealth categories are significantly different from each other.[44] Moreover, using linear regression analysis, we find that the inequality in BMIs increases significantly with the improvement in wealth status.[45] The estimated slope coefficient of 0.007 for 2007, for example, implies that the average rate of increase in inequality when a women's wealth status improves by one level, e.g., moving to the *poorer* from the *poorest* category is 0.007.

4.3.4.4 Gini Coefficient Estimates and Sampling Variances: Educational Attainment

We again observe the issue of small numbers of women in many of the clusters (see Table D.8 in Appendix D), when the BMI sample is classified according to women's educational attainment (no education; primary incomplete; primary complete; secondary incomplete, and secondary complete or higher level of education) to examine the impact of education on BMI inequality among ever-married women, 15-49 years of age. Therefore, we ignore the clustering and assume that the obtained samples are under a stratified design. As discussed in the previous subsection, the implication of this assumption is that we could understate the variances. Estimated Gini coefficients and their sampling variances using the three variance formulae, stratum effects, and design effects by educational attainment are presented in Table 19.[46]

---

[44] We tested the null hypothesis that Gini coefficients across the five wealth levels are the same. Tests are formed using $\widetilde{Var_{BT}}(\hat{G})$, and the Wald statistics for 2004, 2007 and 2011 surveys are 1916.379 (p-value 0.000), 237.177 (p-value 0.000) and 157.047 (p-value 0.000), respectively.

[45] The estimated slope coefficients of the linear fitted lines are 0.008 (se 0.001), 0.007 (se 0.001) and 0.005 (se 0.001) for the 2004, 2007, and 2011 samples, respectively.

[46] Additional hypothesis test results are reported in Table D.9 in Appendix D.

Table 19 Gini coefficient estimates and their variances for BMI among ever-married Bangladeshi women aged 15-49 - BDHS 2004, 2007 and 2011: by educational attainment

Survey year	Educational attainment	$\hat{G}$	$\widehat{Var}_{BK}(\hat{G})$	$\widehat{Var}_{BT}(\hat{G})$	$\widehat{Var}_{BOOT}(\hat{G})$	$\widehat{Var}_{BT}(\hat{G})$ breakdown		Design effect
						$\widehat{Var}_{SRS}(\hat{G})$	Stratum effect	
2004	No education	0.080	1.099	1.093	1.084	1.111	0.018	0.984
	Primary incomplete	0.086	3.135	3.104	3.031	3.189	0.085	0.973
	Primary complete	0.088	4.733	4.603	4.848	4.824	0.221	0.954
	Secondary incomplete	0.090	2.177	2.153	2.067	2.268	0.115	0.949
	Secondary or higher	0.102	5.096	4.991	5.159	5.064	0.073	0.986
2007	No education	0.085	2.009	2.002	2.011	2.023	0.021	0.990
	Primary incomplete	0.087	2.538	2.523	2.483	2.571	0.048	0.981
	Primary complete	0.094	6.326	6.171	6.148	6.600	0.429	0.935
	Secondary incomplete	0.091	2.083	2.070	2.223	2.146	0.076	0.965
	Secondary or higher	0.102	3.972	3.928	3.523	3.967	0.039	0.990
2011	No education	0.094	1.419	1.413	1.329	1.446	0.033	0.977
	Primary incomplete	0.092	1.913	1.902	1.856	1.943	0.041	0.979
	Primary complete	0.096	2.827	2.802	2.760	2.847	0.045	0.984
	Secondary incomplete	0.097	1.125	1.122	1.135	1.139	0.017	0.985
	Secondary or higher	0.096	2.784	2.764	2.765	2.782	0.018	0.994

Notes: Variance estimates have been scaled by 10^6. $\widehat{Var}_{BK}(\hat{G})$ $\left(\widehat{Var}_{BT}(\hat{G})\right)$ is Binder & Kovačević's (1995) (Bhattacharya's (2007)) estimator and $\widehat{Var}_{BOOT}(\hat{G})$ is the bootstrap variance estimator using expression (4.4) for 799 replications. $\widehat{Var}_{SRS}(\hat{G})$ 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 $\widehat{Var}_{BT}(\hat{G})$ to $\widehat{Var}_{SRS}(\hat{G})$.

We find that, except for the most recent survey (2011), the inequality in women's BMIs increases with the level of educational attainment.[47] At the nominal 5% significance level, the change in Gini coefficients between 2004 and 2007 at each level of education (except for the no education group) is not significantly different from zero, regardless of the variance used to construct the test. Between 2007 and 2011, the differences in inequalities are statistically significantly different from zero, except for the subgroup of women who had completed primary education. For the women who had completed neither primary nor secondary education, inequality increases, after showing no change between 2004 and 2007, and we find that for women who have completed a secondary or higher level of education, BMI inequality falls between 2007 and 2011. These results suggest that nutritional diversity in women with less than secondary complete education increased over time.

A higher value of $\hat{G}$ for highly educated women likely arises from the larger number of overweight women in this category compared to those from the other subgroups. Nevertheless, even though overweight prevalence increased over time, a faster falling rate of undernourished women with no significant change in normal weight prevalence (see Table 14) reduces the BMI variation in the distribution. A declining BMI inequality overtime accompanied by an increasing overweight prevalence for educated women suggests that overcoming thinness or malnutrition is prioritized over being overweight as an achievement of better health.

---

[47] At the nominal 5% significance level, we reject the null hypothesis that inequalities for the five educational attainment levels are the same. The test is formed using $\widetilde{Var_{BT}}(\hat{G})$, and the Wald statistics for 2004, 2007 and 2011 surveys, are 92.530 (p-value 0.000), 55.651 (p-value 0.000), and 9.707 (p-value 0.008), respectively. The estimated equation by regression technique for inequality in women's BMIs against educational attainment leads to slope estimates for 2004, 2007 and 2011 of 0.001 (z-statistic 4.768, p-value 0.036), 0.001 (z- statistic 3.549, p-value 0.076), and 0.0005 (z- statistic 1.754, p-value=0.356), respectively.

As expected, the differences in the variance estimates for $\hat{G}$ across various levels of education calculated by expressions (4.2) and (4.3), are quite similar to those obtained in the previous subsection when we explored inequality across wealth categories, with $\widehat{Var_{BK}}(\hat{G})$ being between 0.26% and 2.51% higher than $\widetilde{Var_{BT}}(\hat{G})$. Again, as expected, bootstrap variances for small cluster samples are often smaller than $\widehat{Var_{BK}}(\hat{G})$ or $\widetilde{Var_{BT}}(\hat{G})$. When $\widetilde{Var_{BT}}(\hat{G})$ is disaggregated into the SRS variance with weights and the stratification effect, we observe that the stratification effect increases with the number of strata in the sample, ranging from 0.6% to 6.5% and leading to design effects being less than one.

Finally, although women's mean BMIs and education are positively associated, the inequality in BMI in the 2011 sample is independent of education level. These findings have policy implications. We suggest that policies be directed at all levels of education to improve the equality in women's well-being. In addition, as the inequality of the BMI distribution for women with higher education is decreasing, or at least, is not increasing over time, women should be motivated and supported to receive more health education.

4.3.5 Interval Estimates for the Gini Coefficient

We now report estimated 95% CIs for G as discussed in subsection 4.2.2.5. Recall, two of the CIs are SN approximation interval estimators formed using the asymptotic analytical variances, given in expressions (4.6) and (4.7). One of them is a bootstrap percentile CI and the other one is a percentile-*t* bootstrap CI, as provided in expressions (4.8) and (4.9), respectively. Estimated CIs for G of BMI among all women and when women are grouped by place of residence are reported in Table 20.

These interval estimates are constructed when variances for $\hat{G}$ are obtained based on the stratified two-stage cluster sampling design. The results suggest that both $\hat{G}$ and the entire range of the interval in each case differs from a Gini coefficient associated with perfect equality in the BMI distribution, $G = 0$, implying the presence of significant inequality in women's health when measured by BMI.

We see that the two 95% SN approximation CIs for G using standard errors from the Binder and Kovačević (1995) and Bhattacharya (2007) variance expressions are similar in almost all cases, when reported up to three decimal places, irrespective of the survey year. Variance estimates are fairly small; a slightly larger Binder and Kovačević's variance estimate due to the associated correction factor, compared to Bhattacharya's estimates, does not seem to affect the range or position of an interval estimate. This finding is in accord with our MC simulation results in Chapter 3; the two SN approximation intervals and their widths are similar in finite samples.

Table 20 95% interval estimates for the Gini coefficient for BMI among all women and women by place of residence: BDHS 2004, 2007 and 2011.

	Survey year	$\hat{G}$	95% CI for the Gini coefficient			
			se from $\widehat{Var}_{BK}(\hat{G})^*$	*se* from $\widetilde{Var}_{BT}(\hat{G})^*$	Percentile Bootstrap	Percentile-*t* Bootstrap
All women	2004	0.088	[0.086; 0.090]	[0.086; 0.090]	[0.087; 0.090]	[0.087; 0.089]
	2007	0.092	[0.090; 0.094]	[0.090; 0.094]	[0.090; 0.095]	[0.090; 0.093]
	2011	0.097	[0.096; 0.099]	[0.096; 0.099]	[0.096; 0.099]	[0.096; 0.098]
By place of residence						
Urban	2004	0.104	[0.101; 0.107]	[0.101; 0.107]	[0.102; 0.107]	[0.102; 0.106]
	2007	0.104	[0.101; 0.107]	[0.101; 0.107]	[0.102; 0.109]	[0.101; 0.106]
	2011	0.101	[0.098; 0.104]	[0.098; 0.104]	[0.099; 0.104]	[0.099; 0.102]
Rural	2004	0.081	[0.079; 0.083]	[0.079; 0.083]	[0.079; 0.083]	[0.080; 0.083]
	2007	0.085	[0.082; 0.088]	[0.083; 0.088]	[0.082; 0.087]	[0.083; 0.087]
	2011	0.092	[0.090; 0.094]	[0.090; 0.094]	[0.090; 0.094]	[0.091; 0.093]

Notes: All standard errors in this table are calculated based on the full complex survey design, i.e., stratified two-stage cluster sampling. The percentile-*t* confidence interval is calculated using standard error from expression (4.2) $\left(\widehat{Var}_{BK}(\hat{G})\right)$. A * indicates that the critical value for the interval estimate is obtained from a standard normal distribution.

Except for a few cases, the percentile bootstrap CIs are similar to the SN approximation intervals, suggesting that it is reasonable to use the asymptotic approximations when reporting interval estimates, rather than having to form the more tedious bootstrap percentile intervals. In contrast, the percentile-*t* bootstrap CIs are generally slightly narrower and often right-shifted in comparison to the other three intervals. Our findings are consistent with those reported by Peng (2011). A narrower CI is more precise, and a centered interval may perform better, as seen in Chapter 3. Nevertheless, we believe the gain in performance from the percentile-*t* bootstrap interval estimator, compared to the two SN approximation interval estimators, may not amount too much, when the computational burden involved with the method is accounted for.

In Table 21, we report 95% CIs for G, when women are sorted by wealth category and educational attainment. Standard errors are from variance estimates obtained under the stratified sampling design assumption. Again, the estimated Gini coefficient and interval estimates of G for every category of women is nonzero, implying the presence of some degree of inequality in BMI at all levels of socioeconomic status. The sampling variances of $\hat{G}$ could have been underestimated, resulting in narrower widths of the 95% CIs for G, under the stratified sampling design, compared to those under the stratified two-stage cluster sampling design discussed in subsection 4.3.4.3.

We see that under the stratified sampling design, the two 95% SN approximation CIs are also similar to each other in terms of position and width. It also provides reassurance that the two asymptotic variance estimates, $\widehat{Var_{BK}}(\hat{G})$ and$\widetilde{Var_{BT}(\hat{G})}$, used in forming the interval estimators are similar, as shown in

Chapter 2. The percentile bootstrap and percentile-*t* bootstrap intervals behave as before; the former is comparable with the two SN approximation intervals, and the latter is generally slightly narrower than the other three asymptotic interval estimates.

Table 21 95% interval estimates for the Gini coefficient for BMI among women by wealth category and educational attainment for BDHS 2004, 2007, and 2011.

Category	Survey year	$\hat{G}$	95% CI for the Gini coefficient			
			se from $\widehat{Var_{BK}}(\hat{G})^*$	*se* from $\widetilde{Var_{BT}}(\hat{G})^*$	Percentile Bootstrap	Percentile-*t* bootstrap
By wealth category						
Poorest	2004	0.070	[0.067; 0.073]	[0.067; 0.073]	[0.068; 0.074]	[0.068; 0.073]
	2007	0.073	[0.071; 0.078]	[0.071; 0.078]	[0.072; 0.078]	[0.072; 0.076]
	2011	0.081	[0.079; 0.084]	[0.079; 0.084]	[0.079; 0.084]	[0.079; 0.083]
Poorer	2004	0.073	[0.070; 0.076]	[0.070; 0.076]	[0.071; 0.077]	[0.071; 0.075]
	2007	0.075	[0.071; 0.077]	[0.071; 0.077]	[0.073; 0.079]	[0.071; 0.075]
	2011	0.081	[0.072; 0.077]	[0.072; 0.077]	[0.079; 0.084]	[0.067; 0.070]
Middle	2004	0.078	[0.075; 0.081]	[0.075; 0.081]	[0.076; 0.081]	[0.076; 0.080]
	2007	0.081	[0.078; 0.085]	[0.078; 0.085]	[0.079; 0.086]	[0.079; 0.084]
	2011	0.088	[0.086; 0.090]	[0.086; 0.090]	[0.086; 0.091]	[0.086; 0.089]
Richer	2004	0.088	[0.085; 0.091]	[0.085; 0.091]	[0.085; 0.093]	[0.086; 0.091]
	2007	0.089	[0.087; 0.093]	[0.087; 0.093]	[0.087; 0.094]	[0.087; 0.092]
	2011	0.095	[0.093; 0.098]	[0.093; 0.098]	[0.093; 0.098]	[0.093; 0.097]
Richest	2004	0.100	[0.097; 0.103]	[0.097; 0.103]	[0.098; 0.103]	[0.098; 0.102]
	2007	0.102	[0.099; 0.105]	[0.099; 0.105]	[0.099; 0.105]	[0.099; 0.104]
	2011	0.098	[0.096; 0.101]	[0.096; 0.101]	[0.096; 0.101]	[0.096; 0.999]
By educational attainment						
No education	2004	0.080	[0.078; 0.082]	[0.078; 0.082]	[0.078; 0.082]	[0.078; 0.081]
	2007	0.085	[0.083; 0.088]	[0.083; 0.088]	[0.083; 0.089]	[0.083; 0.087]
	2011	0.094	[0.091; 0.096]	[0.091; 0.096]	[0.092; 0.096]	[0.092; 0.095]
Primary incomplete	2004	0.086	[0.082; 0.089]	[0.082; 0.089]	[0.083; 0.090]	[0.083; 0.088]
	2007	0.087	[0.084; 0.091]	[0.084; 0.091]	[0.085; 0.091]	[0.085; 0.089]
	2011	0.092	[0.090; 0.095]	[0.090; 0.095]	[0.090; 0.096]	[0.090; 0.094]
Primary complete	2004	0.088	[0.083; 0.091]	[0.083; 0.091]	[0.085; 0.094]	[0.082; 0.088]
	2007	0.094	[0.090; 0.099]	[0.090; 0.099]	[0.091; 0.101]	[0.091; 0.097]
	2011	0.096	[0.093; 0.099]	[0.093; 0.099]	[0.093; 0.100]	[0.093; 0.098]
Secondary incomplete	2004	0.090	[0.087; 0.093]	[0.087; 0.093]	[0.088; 0.093]	[0.088; 0.092]
	2007	0.091	[0.088; 0.094]	[0.088; 0.094]	[0.088; 0.094]	[0.089; 0.093]
	2011	0.097	[0.095; 0.099]	[0.095; 0.099]	[0.095; 0.099]	[0.095; 0.098]
Secondary complete or higher	2004	0.102	[0.098; 0.107]	[0.098; 0.107]	[0.099; 0.108]	[0.098; 0.104]
	2007	0.102	[0.098; 0.106]	[0.098; 0.106]	[0.099; 0.106]	[0.099; 0.104]
	2011	0.096	[0.093; 0.099]	[0.093; 0.099]	[0.093; 0.100]	[0.094; 0.098]

Notes: All standard errors in this table are calculated based stratified survey design. The percentile-*t* confidence interval is calculated using standard error from expression (4.2) $\left(\widehat{Var_{BK}}(\hat{G})\right)$. A * indicates that the critical value for the interval estimate is obtained from a standard normal distribution.

4.4 APPLICATION TWO: HEALTH INEQUALITY OF BANGLADESHI CHILDREN AGED 6-59 MONTHS USING BLOOD HEMOGLOBIN LEVEL

4.4.1 Introduction

This section examines the health status of Bangladeshi children, 6-59 months of age, using hemoglobin level. Hemoglobin (Hb), measured in grams per deciliter (g/dl), is the iron-containing oxygen-transport metalloprotein in the red blood cells of all vertebrates that carries oxygen from the respiratory organs (lungs) to the rest of the body. Oxygen helps burn nutrients and provides energy to power the functions of the organism. Major nutrients involved in the synthesis of Hb are iron, folic acid, and vitamin B_{12}. The Hb level in the blood is primarily used to define whether or not an individual is suffering from anemia, which is a global health problem affecting about two billion people or over 30% of the world's population (see, e.g., Kotecha, 2011). The prevalence of anemia is highest among preschool-age children (<5 years) (over 47%), followed by pregnant women (over 41%) (see, e.g., Benoist et al., 2008).

According to the WHO, a person is considered to be anemic, when his or her Hb level is below two standard deviations (-2SD) of the Hb distribution mean of a normal population of people of the same gender and age, and living at the same altitude. In a normal population, 2.5% of individuals would be expected to be below the 2SD threshold. When this proportion exceeds 5% of the population, anemia is considered a public health problem (e.g., WHO, 2001, p. 4). In addition, the WHO established cut-off points of Hb levels, for different age groups at sea level, below which individuals are said to be anemic (see Table 22).

Table 22 Anemia cut-offs for different age groups at sea level, measured by Hb, g/dl.

Population by age group	No Anemia Hb (g/dl)	Anemia Mild Hb g/dl	 Moderate Hb g/dl	 Severe Hb g/dl
Children 6-59 months of age	≥11.00	10.00-10.90	7.00 - 9.90	<7.00
Children 5-11 years of age	≥11.50	11.00-11.40	8.00-10.90	<8.00
Children 12-14 years of age	≥12.00	11.00-11.90	8.00-10.90	<8.00
Non-pregnant women (≥15 yrs)	≥12.00	11.00-11.90	8.00-10.90	<8.00
Pregnant women	≥11.00	10.00-10.90	7.00 - 9.90	<7.00
Men (≥15 years)	≥13.00	11.00-12.90	8.00-10.90	<8.00

Notes: Hemoglobin is measured in grams per decilitre, g/dl. Sources: WHO (2001, p. 33) and WHO (2011, p. 3). Adjustments have to be made to diagnose anemia in case of individuals who either live at higher altitudes or who smoke.

Because the Hb level for a child under six months of age is determined by maternal iron provision, no cut-off is provided. Most newborn infants up to four months old have abundant iron in their blood from their mothers (see, e.g., Kotecha 2011). In contrast, after about four months of age, a gradual decline occurs in the iron reserve due to continued rapid growth in infant body size, and from six months of age iron is derived mostly from external nutritional sources. While any deviation from the optimal Hb level can result in detrimental health consequences for any age group, children, 6-59 months old, are often regarded as being the most vulnerable (e.g., Singh and Patra, 2014; WHO, 2001). Thus, our attention is on children of age 6-59 months.

Anemia can be caused by many factors,[48] however, iron deficiency (ID), [49] which reduces the erythropoiesis and consequently the Hb level, primarily due to inadequate nutritional intake, is considered to be either the main cause or a major

---

[48] For example, nutritional deficiency (e.g., low intakes of folic acid and vitamins A, or B_{12}), acute and chronic inflammation, parasitic infections, sickle cell disease, malaria, and inherited or acquired disorders that affect hemoglobin synthesis (WHO, 2011).

[49] Iron deficiency is a condition in which the supply of mobilizable iron stores in the blood of an individual diminishes to zero and iron supply to the transport protein apotransferrin is compromised (e.g., WHO, 2001, p.3).

contributing factor of anemia globally, accounting for over 50% of all cases (e.g., Benoist et al., 2008; WHO 2001, p. 3); this particular form of anemia is known as iron-deficiency anemia (IDA). The terms anemia and IDA are used interchangeably throughout this section.

Anemia has long-term deleterious consequences for these children, as well as for the well-being of the entire nation. Anemic children typically have poor defense against infections, lower physical growth, and poor motor and mental development (see, e.g., Zhao et al., 2012; Baker et al., 2010; Benoist et al., 2008; WHO, 2001). The economic implications of anemia are large, and include the costs incurred for prevention, reduced productivity due to physical and mental deterioration and long-term effects on mental development and human capital formation (e.g., Singh and Patra, 2014; WHO, 2001, p. 11-13). The current economic cost of poor Hb status in Bangladesh is estimated to be 7.9% of GDP (e.g., see USAID et al., 2011).

As discussed previously, our choice for using Hb as a measure of health status among children is part of our work on exploring inequality using non-traditional well-being measures. The aforementioned consequences of low Hb levels motivated us to adopt this measure to examine the health status of young children in Bangladesh.

A number of studies detail the prevalence and impacts of anemia among pre-school age children in developing countries (e.g., Singh and Patra 2014; Ayoya et al., 2013; Zhao et al., 2012; Kotecha, 2011; Chang et al., 2011; Lozoff et al., 2006; Osório et al., 2001; Khusun et al., 1999; Cornet et al., 1998). Nevertheless, we are unaware of any study that considers the Gini coefficient to measure children's health inequality using Hb level[50] as the measure of health status. Our work on children's health inequality using BDHS data for Hb level contributes to the literature on the inequality of health among Bangladeshi children.

This section is organized as follows. Subsection 4.4.2 presents a discussion on the prevalence of anemia among Bangladeshi children of age 6-59 months. In subsection 4.4.3, we estimate survey adjusted means for Hb levels for all children, as well as when the children are categorized by several background features; we also report results from a number of hypothesis tests. Our key inequality measures are provided in subsection 4.4.4, along with the estimated sampling variances. Interval estimates for the Gini coefficient are reported in subsection 4.4.5.

4.4.2 Anemia Prevalence in Bangladeshi Children, 6-59 Months Old

Percentages of Bangladeshi children aged 6-59 months with different levels of anemia is presented in Figure 10 for the survey year 2011. We see that more than half the children suffer from some degree of anemia (<11.0 g/dl). This prevalence of anemia is well above the 40% benchmark rate considered as making a severe public health problem,[51] even though some interventions have already been implemented to address the issue.[52] By category, 29% of the children have mild anemia (Hb 10.0-10.9

---

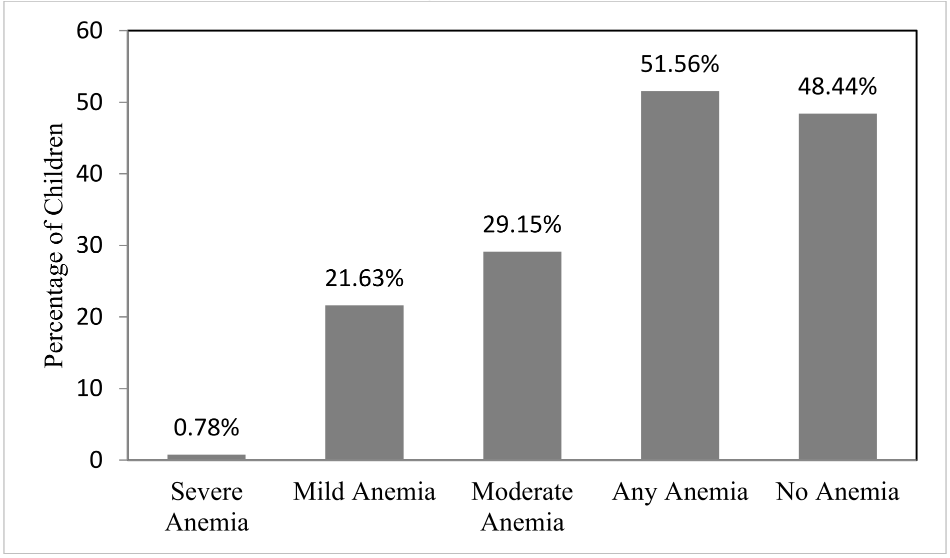
[50] Wang (2013) estimates an economic-related inequality measure, the Concentration Index, for anemia for Cambodian children, 6-59 months old. The Concentration Index and the Gini coefficient (G) are both calculated with reference to the Lorenz curve (LC). For the Concentration Index calculation, the data is arranged by a second variable, whereas G is calculated for a variable when sorted in ascending order. For example, Wang (2013) calculates the Concentration Index using the LC, where the x-axis represents the cumulative proportion of the sample, ranked by wealth status from poorest to richest categories; the y-axis represents the cumulative proportion of the Hb variable (for <11 g/dl level) corresponding to each wealth category. Using Cambodian DHS 2000, 2005 and 2010 data, he examines the anemia inequality among children under five years of age.

[51] WHO declares the situation as a severe public health problem when the anemia prevalence rate is ≥ 40%. For other cut-off points; see WHO (2001, p. 17).

[52] For instance, increased distribution of iron supplements and deworming medication to children, 1-5 years of age, every six months (see NIPORT et al., 2013, p. 178).

g/dl), over 21% have moderate anemia (Hb 7.0-9.9 g/dl), and less than 1% have severe anemia (Hb <7.0 g/dl).

Figure 10 Prevalence of anemia in Bangladeshi children, 6-59 months old, BDHS 2011.



Notes: Anemia cut-offs for children aged 6-59 months are: any anemia, Hb <11.00 g/dl; mild anemia, 10.00 g/dl ≤Hb<11.00 g/dl; moderate anemia, 7.00 g/dl≤Hb<10.00 g/dl and severe anemia, Hb <7.00 g/dl.

The proportions of children with anemia, when the children are classified by gender, age group, and place of residence are given in Table 23. In addition, we report the relative risk (*RR*) factor whenever the sample is divided by a category that has two groups (e.g., gender: male and female). To determine which group (e.g., male or female) is at higher risk, the *RR* is calculated as the ratio of the proportion or percentage of one group to another, given as:

$$RR = \frac{P_1}{P_2} = \frac{P_1(\text{anemic children}|\text{children in group1})}{P_2(\text{anemic children}|\text{children in group2})},$$

where P_1 and P_2 are probabilities or percentages of anemic children in group 1 and group 2 respectively. If RR >1, the numerator group (group 1) has a higher risk; $RR = 1$, both groups are at the same risk; and when RR <1, the denominator group is at higher risk of anemia.

Table 23 Percentages of Bangladeshi children aged 6-59 months with anemia by gender, age groups and place of residence, BDHS 2011.

	Anemia Status by Hb Level				
	Severe (<7.0 g/dl)	Moderate (7.0-9.9 g/dl)	Mild (10-10.9 g/dl)	Any anemia (<11 g/dl)	Number of Children
All Children	0.78%	21.63%	29.15%	**51.56%**	2432
Gender					
Male	0.80%	23.15%	28.70%	52.65%	1244
Female	0.76%	20.03%	29.63%	50.42%	1188
Relative Risk, RR	1.05	1.16	0.97	1.04	
$\chi^2(1)$ [a]	0.017	3.484	0.256	1.212	
p-value	0.449	0.062	0.307	0.136	
Age Group					
6-23 months	1.53%	37.55%	32.95%	72.03%	783
24-59 months	0.42%	14.07%	27.35%	41.84%	1649
Relative Risk, RR	3.64	2.67	1.21	1.72	
$\chi^2(1)$ [a]	8.409	172.658	8.062	193.707	
p-value	0.002	0.000	0.002	0.000	
Place of residence					
Rural	0.76%	22.21%	30.04%	53.01%	1711
Urban	0.83%	20.25%	27.05%	48.13%	721
Relative risk, RR	0.916	1.097	1.111	1.101	
$\chi^2(1)$ [a]	0.034	1.149	2.203	4.841	
p-value	0.427	0.142	0.069	0.014	

Notes: [a] H_0: Relative risk factor, RR = 1, or anemia prevalence between the two groups of children is the same and H_1: RR≠1.

We see that female children have a slightly lower prevalence of anemia than male children; calculated *RR* factors show that male children are at higher risk for all types of anemia except for mild anemia. At the 5% nominal significance level; however, we fail to reject the null hypothesis that the *RR* factor is equal to 1, suggesting that anemia prevalence across gender is the same. The idea that gender is not a significant issue in determining anemia prevalence among children under five years of age is consistent with recent suggestions that Bangladeshi girls are as well cared for as young boys within their households up to their fifth birthday (UNICEF, 2011). Similar findings have been reported for other developing countries; for example, Borbor et al. (2014) and Osório et al. (2001).

Iron deficiency typically starts to develop in a child after six months of age when complementary food supply, with or without breastfeeding, is inadequate for necessary iron synthesis. In Bangladesh, while nearly all children under two years of age are breastfed, only 21% of them are fed appropriately, following the recommended infant and young child feeding guidelines (see NIPORT et al., 2013, p. 176). Hence, young children, 6-23 months old, may be at the highest risk of anemia. To examine this possibility, we divided the children into those 6-23 months old and those 24-59 months old.

We find that the anemia prevalence is 72.03% in the younger group (6-23 months), which is significantly higher than that of the older group (24-59 months). Hypothesis test results (Table 23) show that children aged 6-23 months are at a higher risk of anemia compared to the older children, with the estimated *RR* factor significantly greater than 1 for each form of anemia (mild, moderate and severe). A higher rate of anemia for younger children has been reported in several other developing countries. For example, in Ghana, 85.1% of the younger group of children are anemic, which is 1.14 times higher than that of the older group (see Ewusie et al., 2014). In Haiti and southern Brazil these rates are 75% and 61.8%, which are 1.24 and 2 times higher than that of children, 24-59 months old, respectively (see Ayoya et al., 2013; Osório et al., 2001).

Among other causes, this high incidence of anemia among infants under two years of age is likely due to widespread poverty[53] and the infants' vulnerability to

---

[53] Poverty can be measured in two ways: the direct calorie intake (DCI) method, and the cost of basic needs (CBN) method. The DCI method estimates the incidence of poverty based on a threshold level of food calorie intake by a person. A person is said to be *absolute* poor if his or her daily calorie intake is less than 2,122 kilocalories, while a person with an intake of less than 1,805 kilocalories is considered to be *hard-core* or *extreme* poor. The CBN method specifies a consumption basket necessary for basic consumption needs and then estimates its cost. An upper poverty line and a lower poverty line are estimated based on a household's expenditure on basic need items. Persons living below these lines are considered to be poor.

diseases and infections, which further decreases their body's capacity to ingest and absorb iron (see, e.g., Ewusie et al., 2014; Osório et al., 2001). Poverty is assumed to be the main reason for the higher prevalence of anemia in younger children in developing countries (see, e.g. Hipgrave et al., 2104). Food insecurity is deeply rooted in poverty, which may end up with households being unable to provide young children with the necessary nutrients for synthesizing Hb. The same reason may apply to malnourished mothers with low levels of iron in their breast milk, which results in anemic children.

The incidence of anemia when children are classified by their place of residence suggests that anemia is more of an issue for rural children than for urban children. Almost three-quarters of the population of Bangladesh live in rural areas, and the incidence of poverty among people living in rural areas is higher than among those living in urban areas.[54] Thus, children living in rural areas tend to be less well-nourished than those living in urban areas (see, e.g., Borbor et al., 2014; Fosto, 2007; Smith et al., 2005) and are more likely to be anemic.

At first glance, the sample data is consistent with the above; rural children are more vulnerable to the risk of anemia, compared to urban children. The difference between proportions of anemic children in rural and urban areas is statistically significantly different from zero (χ^2 =4.84, p-value 0.014), and the RR factor for anemia is statistically greater than 1, at the nominal 5% significance level. For severe and moderate anemia, however, we fail to reject the hypothesis that the RR factor is 1, at least at a 10% level of significance. A similar prevalence of severe and moderate

---

[54] The proportions of rural and urban people in Bangladesh living under the upper poverty line are 35.2% and 21.3% respectively (BPC, 2014). The poor health outcomes of children related to poverty are higher in rural areas; e.g., 39% of rural children under 5 years old are underweight, compared to 28% of urban children (BDHS, 2011, p. 166)

anemia for rural and urban children could result from chronic malnutrition, which could be linked to the general level of extreme poverty in Bangladesh.[55]

To ascertain whether or not anemia prevalence varies for children from different wealth levels, we group children by the BDHS wealth categories. Recall, this variable indicates the relative socioeconomic status of the household, which is measured using an asset-based index. Children in wealthy households may receive better nutritional intake and avoid anemia, given the likely positive relationship between feeding practices and household wealth status (see, e.g., NIPORT et al. 2013, p. 176). As shown in Table 24, a child living in the poorest household is 32% more likely to be anemic than a child living in the richest household, which further supports our claim that poverty is likely a major cause of anemia in children.

We also note that the percentage of children with various levels of anemia (except for severe anemia) declines with increasing wealth, at the 5% nominal significance level. A small proportion of the children in each wealth category suffer from severe anemia, and the proportions of children with severe anemia across wealth categories are statistically the same (χ^2=2.842, p-value 0.585). This may be due to the fact that chronic anemia can be related to a range of factors, such as acute and chronic inflammation, parasitic infections, sickle cell disease, malaria, or may be the result of lack of identifying anemia, rather than simply nutritional deficiency due to the child's household economic standing.

---

[55] Over 17% of the population of the country lives under the *lower* poverty line, corresponding to the per capita expenditure level at which the household could just buy enough food with no money left over to buy anything else (Haughton and Khandker, 2009, p. 53). This statistic is used to measure the undernourishment status of a population as it does not include non-food items.

Table 24 Percentages of Bangladeshi children, 6-59 months old, with anemia, by division and wealth category

	Anemia Status by Hb Level				
	Severe (<7.0)	Moderate (7.0-9.9)	Mild (10-10.9)	Any anemia (<11)	Number of obs.
Wealth category					
Poorest	0.89%	25.40%	32.20%	58.60%	559
Poorer	1.26%	26.68%	29.41%	57.35%	476
Middle	0.45%	19.86%	31.60%	51.92%	443
Richer	0.43%	16.96%	27.17%	44.57%	460
Richest	0.81%	18.42%	25.10%	44.33%	494
$\chi^2(4)$ [b]	2.842	21.600	8.615	26.133	
p-value	0.585	0.000	0.071	0.000	
Anemia across Wealth category					
Slope	-0.001	-0.023	-0.016	-0.034	
z-score [b]	0.770	4.046	2.592	4.943	
p-value	0.221	0.000	0.005	0.000	
Division					
Barisal	1.07%	25.62%	31.67%	58.36%	281
Chittagong	0.66%	23.46%	27.41%	51.54%	456
Dhaka	0.99%	19.11%	27.09%	47.29%	406
Khulna	0.73%	19.78%	32.97%	53.48%	273
Rajshahi	0.34%	19.39%	28.91%	48.64%	294
Rangpur	0.00%	24.15%	32.51%	56.66%	323
Sylhet	1.50%	20.05%	26.32%	47.87%	399
$\chi^2(6)$ [a]	6.581	95.691	7.613	15.112	
p-value	0.361	0.000	0.268	0.019	

Notes: Hemoglobin is measured in g/dl. [a] H_0: Percentage of anemic children is the same for all age groups and H_1: H_0 is not true. [c] H_0: Association between anemia prevalence and wealth category is zero and H_1: Association between anemia prevalence and wealth is negative.

When children are grouped by the division in which they reside, we see that the prevalence of overall anemia and moderate anemia differs across divisions, but this is not the case with mild or severe anemia (Table 24). With overall anemia, several divisions (Barisal, Khulna and Rangpur) have much higher rates of some forms of anemia, compared to other divisions. Again, poverty differences across regions likely leads to such variations. Some of these regions are economically weaker with higher poverty levels. For instance, Ferdousi and Dehai (2014) report headcount rates of poverty (using the upper poverty line) of 42.3% for Rangpur, 39.4% for Barisal and 33% for Khulna, which are much higher than for the four other divisions (Chittagong, Dhaka, Rajshahi, and Sylhet). Regarding severe anemia, our findings suggest that this is predominantly a national concern, with it being a larger issue in some divisions.

Our investigation of anemia incidence in children aged 6-59 months by various background characteristics suggests that children under two are more likely to suffer from anemia with the pervasiveness decreasing with increased age and better household economic conditions. Because our key objective is to estimate inequality in Hb distribution using the Gini coefficient, we now turn to additional summary statistics of the well-being variable.

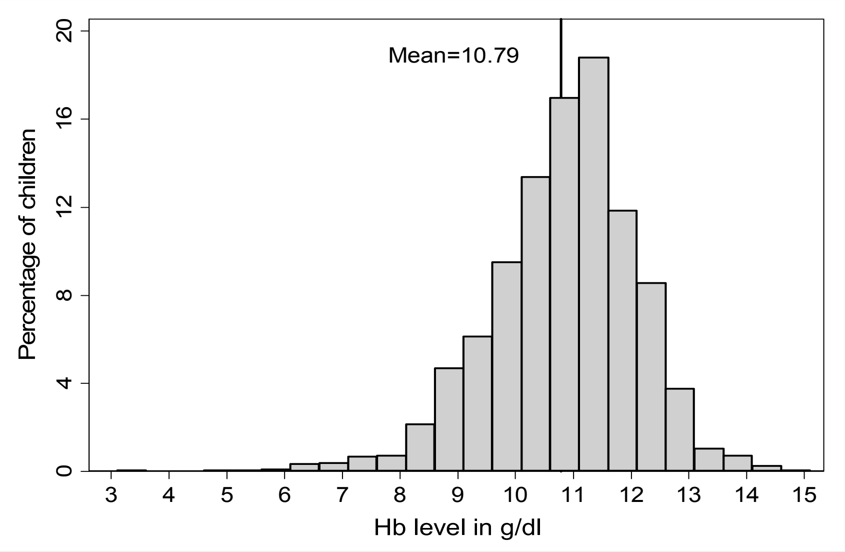
4.4.3 Mean Statistics and Empirical Distributions of Hb

Recall, the Hb sample of children, 6-59 months old, is a subsample of the BDHS 2011 survey, with observations on 2,432 children, drawn from 583 clusters across 20 strata. For this sample, the average size of a cluster is just over four children; ranging from 1 to 16 children in a cluster. More than 10% of the clusters consist of just one child, almost 43% of the clusters contain 1 to 3 children, and less than 3% of the clusters have more than 10 children. The issue associated with small numbers of children in clusters deteriorates further when children are divided by gender, age group, place of residence, and division or wealth category (see Table D.10 in Appendix D).

That we have a large number of clusters containing one child, irrespective of which subgrouping we examine, and for the full sample, makes cluster analysis almost impossible. Accordingly, we discount the clustering feature from our sampling design, and proceed as if the data are obtained under a stratified sampling design. We discussed the possible consequences of estimating parameters of interest under the full complex survey design, i.e., stratified two-stage cluster sampling design, in subsection 4.3.4.3, when the average size of a cluster is small. However, as mentioned previously, estimates, especially sampling variance, are likely somewhat underestimated under the stratified sampling design.

The mean Hb level in the *all children* sample is 10.79 g/dl with a standard deviation of 1.248, which is lower than the anemia threshold level (11.00 g/dl); recall that 51.6% of children in this sample are anemic. While exploring the distribution of the Hb data, we see that the empirical distribution of the Hb in the *all children* sample is slightly negatively skewed (see Figure 11), and likely not normally distributed.[56] The shape of the distribution implies the presence of some severely anemic children, with very low levels of Hb; in some cases, as low as 3.1 g/dl.

Figure 11 Empirical distribution of Hb for Bangladeshi children aged 6-59 months, BDHS 2011.



Notes: Skewness and kurtosis estimates of the distribution are -0.593 and 4.399 respectively.

---

[56] A Jarque-Bera test (JB) (Bera and Jarque, 1981) strongly rejects the null hypothesis that the Hb level is normally distributed (JB=340.9, p-value 0.000).

The common perception that male children are preferred to female children in many developing countries, with male children consequently believed to receive better care, is unsubstantiated in our analysis of Bangladeshi children under five, at least in terms of anemia. If this were indeed the case, then male children might receive better nutrition and have higher levels of Hb, at least on average, whereas, we find that the difference in mean levels of Hb between male and female children, even under place of residence or age group classifications, is not statistically significantly different from zero, at least at the nominal 10% level of significance.

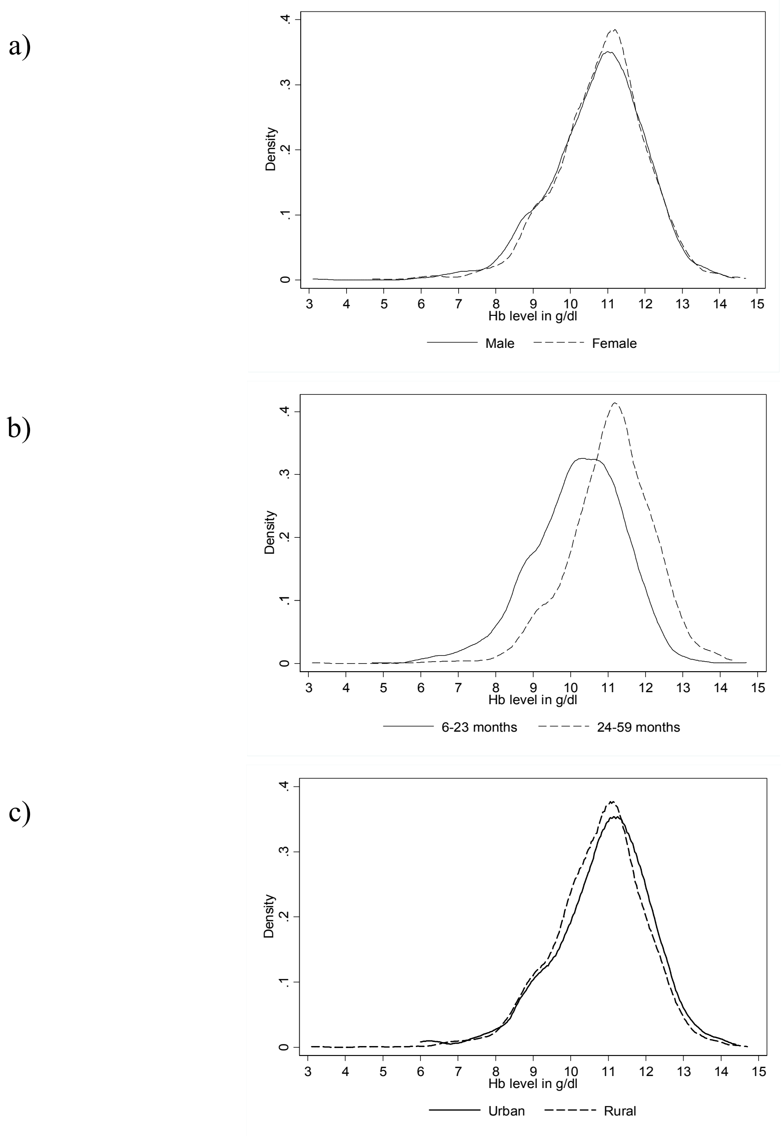
However, as expected, the difference between the mean levels of Hb for the 6-23 month old and 24-59 month old children is significantly different from zero (at the nominal 5% significance level; see Table D.10 in Appendix D), with the older group having, on average, a 0.9 g/dl higher level of Hb than the younger group. These additional results affirm our previous finding that children under two years of age are more likely to be anemic than older children. On the other hand, the difference between the mean Hb levels in rural and urban children is statistically different from zero. With rural children having on average 1.5% less Hb than urban children, this suggests that the place of residence is another important aspect of children's Hb levels, at least on average.

Empirical distributions of Hb data, when classified by gender, age group and place of residence, are presented in Figure 12. We observe each distribution having a slightly longer left tail. The distributions of male and female children (Figure 12a) seem to be similar to each other, even when compared by sample skewness and kurtosis statistics. These results essentially reflect the previous discussion on the mean Hb level by gender of a child, in that gender has a minor role in determining the Hb level in Bangladeshi children, suggesting that males do not receive better (or more) nutrients than female children, at least in terms of anemia for children, 6-59 months old.

A flatter and left-shifted empirical Hb distribution for children aged 6-23 months, compared to that for children, 24-59 months old (Figure 12b) implies the greater vulnerability of younger group children to anemia, with a mean level of Hb equal to 10.761 g/dl, which is about 3% lower than that of the older group. For Brazilian children, Osório et al. (2001) reported Hb empirical distributions for children, 6-23 months old and 24-59 months old that were similar to our empirical distributions. Alternatively, although the empirical distributions of Hb for rural and urban children are quite similar, the distribution for the rural children has a marginally longer left tail, compared to distributions for urban children (Figure 12c). In rural areas some children have extremely low levels of Hb, as low as 3.1 g/dl.

We sort the Hb data by wealth categories to understand whether or not the socioeconomic status of a child's household can help explain the variation in Hb levels across children. We see that the mean Hb levels for children across wealth categories vary markedly, with the lowest (10.56 g/dl) and highest (10.99 g/dl) mean Hb levels for children being from the *poorer* and *richest* wealth category households, respectively. It makes sense that a child from a wealthier household is more likely to receive adequate dietary intake of nutrients necessary for the synthesis of Hb, compared to a child from a less wealthy household. Lastly, the variation in mean Hb levels across divisions is statistically significant,[57] which is in accordance with our findings on the regional anemia prevalence from subsection 4.4.2.

Figure 12 Empirical distributions of Hb level for Bangladeshi children, 6-59 months old, BDHS 2011: by gender, age group, and place of residence.



In addition, we see that the mean Hb levels between children, 6-23 months old and children, 24-59 months old, are statistically significantly different (at the nominal

---

[57] At the 5% nominal significance level, we reject the hypothesis that the mean Hb levels across the seven divisions are the same; see Table D.14 in Appendix D for hypothesis test results.

5% significance level) for each category of wealth and each division (see Table D.13-14 in Appendix D). These findings imply that children under two, irrespective of their households' economic background and region of residence, are more vulnerable in terms of Hb synthesis compared to older children. This is important public health information. To reduce the prevalence of anemia, necessary interventions, e.g., perhaps distribution of iron supplements and deworming medication, should focus heavily on children under two.

4.4.4 Gini Coefficient Estimates and Sampling Variances

Assuming a stratified sampling design, we report estimated Gini coefficients and three sampling variances (the asymptotic variances given in expressions (4.2) and (4.3) and the bootstrap variance for 799 replications from expression (4.4)) for Hb data for all children and by gender, age group, place of residence, division, and wealth categories (Table 25). In addition, the table shows the stratum effect on variance estimates using Bhattacharya's (2007) variance formula in (4.3).

The Gini coefficient estimate of Hb in the *all children* sample is 0.064 for the BDHS 2011 survey; a certain level of health inequality amongst children, 6-59 months old, is present. Some variation will always exist in children's Hb levels, even in a healthy population. As such, we estimated G for our nonanemic children to ascertain a base G for comparison with the Gini for the full sample of children. It is 0.031, which is less than half of what we observe for the full sample. This suggests that an inequality exists in the levels of anemia, well above the natural variation in Hb levels we would observe if all the children were healthy.

Table 25 Gini coefficient estimates, sampling variances, and design effects for Hb level for Bangladeshi children aged 6-59 months and children by gender, age group, place of residence, wealth category, and division, BDHS 2011.

Category	n	$\hat{G}$	$\widehat{Var}_{BK}(\hat{G})$	$\widetilde{Var}_{BT}(\hat{G})$	$\widehat{Var}_{BOOT}(\hat{G})$	$\widetilde{Var}_{BT}(\hat{G})$ breakdown		Design effect
						$\widetilde{Var}_{SRS}(\hat{G})$	Stratum effect	
All children	2432	0.064	1.609	1.598	1.611	1.629	0.032	0.981
Gender								
Male	1244	0.066	2.928	2.890	2.928	2.942	0.052	0.982
Female	1188	0.063	3.542	3.492	3.345	3.610	0.187	0.967
Age Group								
6-23 months	783	0.070	5.816	5.702	5.836	5.837	0.134	0.977
24-59 months	1649	0.057	1.719	1.700	1.695	1.753	0.053	0.970
Place of residence								
Rural	1711	0.063	2.001	1.994	2.015	2.000	0.007	0.997
Urban	721	0.068	8.257	8.115	8.033	8.732	0.616	0.929
Wealth Category								
Poorest	556	0.065	5.700	5.620	5.687	5.746	0.126	0.978
Poorer	474	0.067	10.120	9.924	9.988	10.086	0.162	0.984
Middle	443	0.065	7.431	7.220	7.671	7.670	0.450	0.941
Richer	457	0.060	5.861	5.705	5.804	5.909	0.204	0.965
Richest	494	0.070	11.510	11.180	10.410	12.458	1.281	0.897
Division/Region								
Barisal	281	0.069	11.190	11.120	10.940	11.381	0.256	0.978
Chittagong	456	0.067	6.617	6.586	6.348	6.652	0.065	0.990
Dhaka	406	0.066	8.465	8.394	7.909	8.744	0.350	0.960
Khulna	273	0.068	12.230	12.150	11.030	12.391	0.238	0.981
Rajshahi	294	0.064	10.320	10.206	10.140	10.542	0.286	0.973
Rangpur	323	0.064	7.040	7.008	6.934	7.187	0.179	0.975
Sylhet	399	0.068	12.970	12.910	12.510	13.001	0.087	0.993

Notes: Variances have been scaled by 10^6. All coefficients are estimated based on stratified sampling design. $\widehat{Var}_{BK}(\hat{G})$ $\left(\widetilde{Var}_{BT}(\hat{G})\right)$ is Binder & Kovačević's (1995) (Bhattacharya's (2007)) estimator and $\widehat{Var}_{BOOT}(\hat{G})$ is the bootstrap variance estimator using expression (4.4) for 799 replications. $\widetilde{Var}_{SRS}(\hat{G})$ 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 $\widetilde{Var}_{BT}(\hat{G})$ to $\widetilde{Var}_{SRS}(\hat{G})$.

We assess the precision of $\hat{G}$ for all children by estimating sampling variances in three ways: using Binder and Kovačević's (1995) and Bhattacharya's (2007) asymptotic formulae, and with the traditional bootstrap technique. For a finite sample, due to the correction factor associated with the former variance formula, which now accounts for the number of observations in a stratum, it is expected to be slightly larger than Bhattacharya's variance estimate using expression (4.3). We know that this difference between the variance formulae diminishes as the number of children in the strata increases.[58] Consistent with our observations with the BMI samples, we see that the bootstrap variance estimates are not always more efficient than the analytical asymptotic variance estimates; the bootstrap technique sometimes performs better in small samples.

For the *all children* Hb level sample, we find that the three variance estimates are quite similar to each other, with $\widetilde{Var_{BT}}(\hat{G})$ being less than 1% smaller than $\widehat{Var_{BK}}(\hat{G})$, and $\widehat{V_{BOOT}}(\hat{G})$ being almost the same as $\widehat{Var_{BK}}(\hat{G})$. This suggests that there is little to be gained in undertaking a bootstrap to calculate the variance. In addition, when the variance estimate $\widetilde{Var_{BT}}(\hat{G})$ is broken down, we see a 2% gain from stratification over the variance calculated from SRS with sampling weights. This results in the calculated design effect being less than 1, suggesting that the stratified sampling variance estimate is more efficient than the SRS variance estimate, though the stratification results in only a small gain.

When children are divided by gender, the inequality among male children is 4.76% higher than among female children, however, we fail to reject the hypothesis that the inequalities are the same (at the nominal 5% significance level), no matter which standard error is used to calculate the test statistic (Table 26). Combining this result with the findings from Table 23 and the mean statistic analysis, it can be said that gender is not a

---

[58] For instance, the average size of a stratum in the urban sample is 55, and the estimated difference between these two variances is 0.142; whereas, for the rural sample, it is 0.007 with an average stratum size of 244.

statistically significant factor in determining a child's Hb level. As variation in Hb levels among children, especially anemia, often arises from nutritional deficiencies, this suggests that gender of a child under five is not linked to nutritional intake.

Table 26 Hypothesis tests for Gini coefficients for Hb level for Bangladeshi children aged 6-59 months, BDHS 2011.

Null Hypothesis	Change in $\hat{G}$	*se* from $\widetilde{Var_{BK}}(\hat{G})$		*se* from $\widetilde{Var_{BT}}(\hat{G})$		*se* from $\widetilde{Var_{SRS}}(\hat{G})$	
		Wald statistic	p-value	Wald statistic	p-value	Wald statistic	p-value
H_0: $G_{\text{male}} = G_{\text{female}}$	0.003	1.391	0.238	1.410	0.235	1.374	0.241
H_0: $G_{6-23} = G_{24-59}$	0.013	22.429	0.000	22.832	0.000	22.266	0.000
H_0: $G_{\text{rural}} = G_{\text{urban}}$	-0.005	2.437	0.118	2.473	0.116	2.329	0.127
H_0: G's across five wealth categories are the same	n.a.	6.759	0.149	7.169	0.127	6.541	0.162
H_0: G's across seven divisions are the same	n.a.	2.234	0.897	2.471	0.872	2.416	0.878

While the gender of a child does not seem to have a significant impact on the Hb level inequality, the difference in inequalities between the two age groups is practically and statistically different from zero. The estimated Gini coefficient of the Hb level among the younger children is 22.81% higher than that of the older group. Our findings suggest that in addition to higher anemia prevalence among children under two, more variability also exists in the Hb level distribution for this group, as measured by their standard deviations, and the inequality statistics. As mentioned earlier, children under two grow quickly and need adequate iron-rich food to keep their Hb levels within the normal range. However, the economic condition of households and imbalanced child feeding practices may affect their nutritional needs more severely, compared to children, 24-59 months old, leading to higher inequality in the Hb level.

We saw that the place of residence (rural; urban) has an impact on children's Hb levels. The rural children are at higher risk of anemia with a significantly lower mean Hb level, compared to urban children. The inequality estimates of rural and urban children Hb distributions are also different. The estimated G for rural children is 7.35% lower than that for urban children. However, the rural sampling variance is approximately one quarter of the urban estimated variance. This difference in variance estimates of $\hat{G}$ reflects sample size effects – the rural sample is almost 2.5 times larger than the urban sample – and more rural children are anemic and centered around the mean value of the distribution.

We observe that the bootstrap variance is slightly smaller than $\widetilde{Var_{BK}}(\hat{G})$ or $\widetilde{Var_{BT}}(\hat{G})$ for the urban sample, but not for the rural sample. Again, this may be related to sample size effects as the bootstrap estimate is likely to be preferred in smaller samples. The stratum effect is considerably larger when calculating the variance for the urban children than the variance generated with the rural children. This is likely because the urban sample has more strata with a smaller average number of children per stratum (13 strata; average size 55.5) compared to the rural sample, which consists of only seven strata with an average size of 244 children. The smaller size of a stratum implies a higher homogeneity of the observations within the strata. Despite these differences, we fail to reject the hypothesis that the Hb level G for rural and urban children are significantly different, no matter which standard error is used to construct the test statistic (see Table 26). Unlike the anemia prevalence and mean Hb level, this suggests that inequality in anemia is unaffected by place of residence of a child.

Estimated Gini coefficients for Hb levels of children, and variance estimates, when the children are sorted by their socioeconomic status and region of residence (division) are reported in Table 26. The results reveal that while a marginally positive relationship exists between the mean Hb level and the wealth category, the inequality in Hb level does not seem to follow a similar relationship with the wealth category. Although the inequality of the Hb level is the highest from the *richest* households, it is followed by inequality among children from the *poorer* households – not by the children in *richer* or *middle* category households. The least inequality is among children from households in the *richer* category. This results in failing to reject our null hypothesis that G across five wealth categories are the same, at the 10% level of significance. Even though the mean level of Hb in children increases with improvement in their household wealth status, in each wealth category, some children have all forms of anemia,[59] leading to more even distributions across the wealth categories.

Variance estimates obtained using the three formulae for $\hat{G}$ for each wealth category are reasonably close to each other. When subdividing the sample by the wealth category, the stratum effect on variance estimates is relatively large, up to 10% of the SRS variance for the richest category. This arises because each wealth subgroup has a large number of strata relative to sample size – from 18 to 20 strata.

Finally, when G coefficients are estimated for children, subgrouped by division, we do not observe any noticeable differences between the Hb inequality statistics. The highest inequality prevails for children from the Barisal division, with the estimated Gini coefficient being 0.069, which is 7.21% higher than the lowest Gini coefficient estimates for children from the Rajshahi and Rangpur divisions. Statistically, however, we cannot reject the null hypothesis that the Gini coefficients across the seven divisions are the same.

The above discussion on the empirical Gini coefficients from the Hb data suggests that health inequality among children under five years of age is predominantly a national concern in Bangladesh. The age of a child plays an important role, with children under two

---

[59] Earlier, we related the causes of severe to moderate types of anemia among children from *rich* households to various diseases and infections, rather than to just nutritional consequences linked to the household's wealth status.

showing the highest inequality. Other factors, like place of residence, wealth category, and division are important considerations in explaining anemia prevalence, but do not statistically significantly affect the inequality in Hb among Bangladeshi children of age 6-59 months.

4.4.5 Interval Estimates for the Gini Coefficient

We report four 95% CIs for G of Hb level among all children, 6-59 months old, and include their background characteristics (gender, age group, place of residence, wealth category, and division) in Table 28. Two of the intervals are standard normal (SN) approximation CIs using standard errors from expressions (4.2) and (4.3), given in expression (4.6) and (4.7), respectively. The two other CIs are bootstrap percentile CI and percentile-t bootstrap CI, given in expressions (4.8) and (4.9), respectively. For children's Hb level data, we construct these interval estimates when estimation is performed under the stratified sampling design assumption. children's Hb level data, we construct these interval estimates when estimation is performed under the stratified sampling design assumption.

The interval estimates for G using standard errors from the Binder and Kovačević's (1995) and Bhattacharya's (2007) variance expressions are very similar, at least to the third decimal place, which is in accordance with our MC simulation results in Chapter 3 and the BMI example.

Table 27 Empirical 95% confidence intervals (CIs) for the Gini coefficient for Hb level for children and children subgrouped by background features for BDHS 2011.

	$\hat{G}$	*se* from $\widehat{Var}_{BK}(\hat{G})$		*se* from $\widetilde{Var}_{BT}(\hat{G})$		Bootstrap Percentile		Bootstrap *t*-Percentile	
		95% CI	Width	95% CI	Width	95% CI	Width	95% CI	Width
All children	0.064	[0.061; 0.066]	0.005	[0.061; 0.066]	0.005	[0.062; 0.067]	0.005	[0.062; 0.065]	0.003
Gender									
Male	0.066	[0.062; 0.069]	0.007	[0.062; 0.069]	0.007	[0.063; 0.070]	0.007	[0.063; 0.067]	0.004
Female	0.063	[0.059; 0.067]	0.008	[0.059; 0.067]	0.008	[0.060; 0.068]	0.008	[0.060; 0.065]	0.005
Age Group									
6-23 months	0.070	[0.065; 0.074]	0.009	[0.065; 0.074]	0.009	[0.066; 0.076]	0.010	[0.066; 0.073]	0.007
24-59 months	0.057	[0.054; 0.059]	0.005	[0.054; 0.059]	0.005	[0.055; 0.060]	0.005	[0.055; 0.058]	0.003
Place of residence									
Rural	0.063	[0.060; 0.066]	0.006	[0.060; 0.066]	0.006	[0.061; 0.067]	0.006	[0.061; 0.065]	0.004
Urban	0.068	[0.062; 0.074]	0.012	[0.062; 0.074]	0.012	[0.065; 0.076]	0.011	[0.063; 0.071]	0.008
Wealth category									
Poorest	0.065	[0.061; 0.070]	0.009	[0.061; 0.070]	0.009	[0.062; 0.072]	0.010	[0.061; 0.068]	0.007
Poorer	0.067	[0.061; 0.073]	0.012	[0.061; 0.073]	0.012	[0.063; 0.076]	0.013	[0.062; 0.070]	0.008
Middle	0.065	[0.059; 0.070]	0.011	[0.060; 0.070]	0.010	[0.062; 0.073]	0.011	[0.060; 0.067]	0.008
Richer	0.060	[0.055; 0.065]	0.010	[0.055; 0.064]	0.009	[0.058; 0.067]	0.009	[0.055; 0.061]	0.006
Richest	0.070	[0.063; 0.076]	0.013	[0.063; 0.076]	0.013	[0.066; 0.079]	0.013	[0.064; 0.073]	0.009
Division/Region									
Barisal	0.069	[0.062; 0.076]	0.010	[0.062; 0.076]	0.010	[0.066; 0.079]	0.013	[0.063; 0.072]	0.009
Chittagong	0.067	[0.062; 0.072]	0.010	[0.062; 0.072]	0.010	[0.064; 0.075]	0.011	[0.062; 0.069]	0.007
Dhaka	0.066	[0.060; 0.072]	0.012	[0.060; 0.072]	0.012	[0.063; 0.074]	0.011	[0.061; 0.069]	0.008
Khulna	0.068	[0.061; 0.075]	0.014	[0.061; 0.075]	0.014	[0.065; 0.078]	0.013	[0.062; 0.071]	0.009
Rajshahi	0.064	[0.058; 0.070]	0.012	[0.058; 0.070]	0.012	[0.061; 0.074]	0.013	[0.058; 0.066]	0.008
Rangpur	0.064	[0.059; 0.069]	0.010	[0.059; 0.069]	0.010	[0.062; 0.072]	0.010	[0.059; 0.066]	0.007
Sylhet	0.068	[0.061; 0.075]	0.014	[0.061; 0.075]	0.014	[0.064; 0.078]	0.013	[0.062; 0.072]	0.010

Notes: All standard errors in this table are calculated based the stratified sampling design. The percentile-*t* confidence interval is calculated using standard error from expression (4.2) $\left(\widehat{Var}_{BK}(\hat{G})\right)$.

The bootstrap percentile intervals, formed using the distribution of bootstrap $\hat{G}$, are comparable with the SN approximation intervals in terms of widths of intervals. In some cases, especially for smaller samples, the interval is slightly wider than the SN approximation intervals, implying that the bootstrap methods are likely preferable for samples with smaller number of children. On the other hand, the percentile-*t* bootstrap CIs are generally narrower, compared to the other three interval estimates. Although a narrower interval might seem more "precise", often such an interval implies a lower CP in repeated sampling. However, as one of our asymptotic standard errors $\left(se_{BK} = \sqrt{\widehat{Var_{BK}}(\hat{G})}\right)$ is used in forming the percentile-*t* CIs, we expect this computationally intensive CI to achieve a higher order accuracy (e.g., MacKinnon 2006).

4.5 Concluding Remarks

In this chapter, we assess health inequality among women, 15-49 years of age, and children, 6-59 months old, in Bangladesh via Gini coefficients and BDHS data on women's BMI and children's Hb levels. We also provide a number of useful descriptive statistics for the well-being variables to explore the health status of women and children. Our analysis suggests that health inequalities in both women and children are above the naturally arising inequality in a healthy population.

We find that the prevalence of overweight women increased at a faster rate than the fall in underweight prevalence over the survey span (2004-2011), which resulted in an increasing trend in women's health inequality. The place of residence, wealth and education are important factors in explaining this inequality in BMI of women. BMI inequality among urban women is higher than that for rural women, but the inequality among rural women is increasing over time. On the other hand, while the inequality of BMI among women in the upper socioeconomic and educational subgroups is higher than in the lower subgroups, it is not increasing over time. Rather, we observe increasing inequalities of BMI among women in the lower wealth and educational attainment categories during the survey years, 2004 and 2011, that are statistically significant.

Although place of residence and socioeconomic condition of a child's household are significant reasons for explaining anemia among children under five, neither of these factors nor the gender of a child shows any impact on the inequality of Hb levels among children. The age of a child appears to be the only statistically significant factor for the inequality in Hb levels assessed by the Gini coefficient. Thus, children's health inequality seems to be predominantly a national problem, with children under two being more vulnerable. To improve children's Hb health status in Bangladesh, health policies should be focused on children under two.

We conducted hypothesis tests and constructed CI estimates using asymptotic standard errors from Binder and Kovačević's (1995) and Bhattacharya (2007) formulae to make inferences about the Gini coefficients for BMI and Hb level. In small samples, the variance estimates using Binder and Kovačević's (1995) formula are slightly larger than those of Bhattacharya's (2007) due to the correction factor associated with the former, but there is a little difference between them in samples with relatively more women/children and/or more clusters. On the other hand, the conventional bootstrap variance estimates are sometimes larger in samples with fewer clusters/observations per stratum, compared to the asymptotic variance estimates. As the magnitude of the sampling variances of the Gini coefficient estimates are similar, the hypothesis test results for inequality are qualitatively the same, regardless of the standard error used in constructing the test statistic.

The 95% CIs for G using the standard errors from Binder and Kovačević's and Bhattacharya's variance expressions (4.2) and (4.3), respectively, are similar to each other to three decimal places. The bootstrap percentile and the percentile-t interval estimates are slightly right-shifted, compared to the other two SN approximation intervals, as they are formed from the resampling distributions of the Gini coefficient estimates. Although the percentile-t interval estimates are slightly narrower than the bootstrap percentile intervals, the widths of the bootstrap percentile intervals are highly comparable to those of the two normal approximation interval estimates that use asymptotic variances.

These findings are useful to applied researchers. We suggest that the SN approximation interval estimator using asymptotic variances could be more convenient tools for making inferences about the inequality of a well-being variable since they are easier and require less calculation time, compared to bootstrap interval estimates. Although the percentile-t interval estimates, which are formed using the standard error from Binder and Kovačević's (1995) variance formula, are sometimes slightly narrower (we know from our previous work in Chapter 3 that they may perform better in samples with fewer clusters/observations), the applied work in this chapter illustrates that the gain may not warrant the computational effort.

CHAPTER FIVE: SUMMARY

5.1 Conclusion

This dissertation examines analytical and resampling techniques to estimate the variance of a plug-in estimator for the Gini coefficient when sample data are obtained from a complex survey. Our key contribution for applied researchers is to have shown that, for the sake of computational burden, reporting a standard error, along with a sample Gini coefficient in applied work is not an issue, as is often claimed, even when sample data are from a complex survey. We have detailed several ways that applied researchers can calculate the standard error, demonstrating that none of the methods is computationally prohibitive.

In addition, we have algebraically shown that two asymptotic variance estimators for $\hat{G}$ under a complex survey design are equivalent. The first estimator is obtained by Binder and Kovačević (1995) in the statistics literature and the second estimator is derived by Bhattacharya (2007) in the econometrics literature - a result seemingly unrecognized by Bhattacharya. We also indicated that Davidson's (2009) proposed variance estimator for $\hat{G}$ is a special case of Binder and Kovačević's formula under an *iid* sample case. This finding serves two key purposes. First, the seeming lack of awareness of contributions from other disciplines has led to a broad literature devoted to variance estimation of the Gini coefficient estimator using a variety of approaches. Our attempt to unite some of this literature contributes to the works that have highlighted this situation, especially Langel and Tillé (2013). Uniting branches of research is helpful to theoreticians as well as applied researchers. As much of the sample data are obtained from complex surveys, our detailed mathematical equivalency presentation of the two variance estimator formulae contributes to unifying understanding and solution of the problem under a complex survey design. Second, despite claims by some that such asymptotic variance estimators are difficult to code, we have indicated how these estimators can be readily coded. Furthermore, to aid applied researchers, we also illustrated how straightforward auxiliary regressions can be employed to generate the plug-in estimator for G and these asymptotic variance estimators, regardless of the sampling design.

Given that $\hat{G}$ is a nonlinear function of the sample data, obtaining its variance estimator with a complex sampling design is not straightforward compared to an estimator based on an *iid* assumption. To estimate an asymptotic variance estimator for $\hat{G}$, Binder and Kovačević (1995) and Bhattacharya (2007) derive approximations for $(\hat{G} - G)$ based on the estimating equations and generalized method of moments approaches, respectively, under the complex survey framework. Mathematically, we showed that Bhattacharya's approximation for $\left(\hat{G} - G\right)$ is equivalent to that proposed by Binder and Kovačević (1995) first. Thereafter, via a theorem, we proved that their variance estimators for $\hat{G}$ are equal, at least asymptotically.

Using MC simulation experiments with 36 DGPs under four distributional assumptions (beta, lognormal, chi-square and the Pareto distributions) with a complex survey design, we examine two finite sample properties of the estimator of G: the empirical bias of $\hat{G}$ and the empirical coverage probability of 95% confidence intervals for G. The results suggest that both features are highly sensitive to the number of strata and the underlying distribution of the population data. When there are few strata or the distribution of the population data is not highly skewed, the bias of $\hat{G}$ is negligible, with the ECP of the interval estimator is close to the 95% nominal level. In contrast, the estimator can be quite biased, and the ECP of the interval estimator is far from the desired nominal level, when there are many strata or the distribution of the population data is considerably skewed. Irrespective, the coverage probabilities are closer to the nominal level when the proportion of the population clusters in the sample is increased.

Our simulations examined the ECPs for three SN approximation interval estimators (two use asymptotic variance estimators and one uses a bootstrap variance estimator), and a bootstrap MC percentile estimator for G. Overall, at least for the DGPs that we examined, ECPs of the three SN approximation interval estimators were highly comparable, often intervals using analytical asymptotic variance estimators with more clusters in the sample (e.g., over 10% of the population clusters) had higher coverages. When a relatively small number of clusters were drawn from the population (e.g., less than 10% of the population clusters) or the population data arose from a highly skewed parent distribution, forming interval estimators using the bootstrap percentile method seemed preferable. However, we saw that the SN approximation interval estimators performed better than the bootstrap percentile interval estimators for samples with larger numbers of clusters (over 18% and 24% of the population clusters for DGPs with two strata and five strata, respectively) drawn from population data under the lognormal and chi-square distributions. These SN approximation interval estimators even worked reasonably well for cases from the highly skewed Pareto distribution when samples contain over 20% of the population clusters and there is only relatively few strata, say three.

In addition, we also found that ECPs of the interval estimators formed using the warp-speed method were often much higher (up to 19%, for the sample with 20% of the population clusters from the DGP with three strata under the Pareto distribution) than those of the bootstrap percentile interval estimators across samples and DGPs. In some cases, (e.g., for samples containing more clusters than 20% of the population clusters from typical skewed distribution, or when sample were drawn from DGPs with two or five strata under the Pareto distribution) the coverages of the two percentile interval estimators were comparable. Overall, we do not believe that Giacomini et al.'s (2013) warp-speed method could successfully match the performance of our bootstrap MC percentile interval estimator (based on a double bootstrap method) under the complex sampling design. These findings suggest that their enthusiasm for the benefit of the warp-speed approach in MC studies may need rethinking with complex survey data or when applied more widely than the examples they considered

Our findings have practical implications. When the researcher suspects that the population data are from a typical skewed distribution, as is often the case with distributions of well-being variables, we recommend that applied researchers can use analytical asymptotic variances for undertaking inferences about G in many situations. In the event of samples with a small number of clusters from a heavy-tailed underlying distribution (as may be the case with some income type variables), it may be preferable for applied practitioners to adopt the bootstrap percentile approach to form interval estimates. However, whenever samples with more clusters are available (as is usual with many survey samples, e.g., DHS samples), our recommendation is that even when samples are taken from highly skewed population distributions, employing SN approximation interval estimates using analytical asymptotic variance estimators may be a reasonable way to proceed to assess the Gini coefficient. Calculating sample analytical asymptotic variances, using, for instance, Binder and Kovačević's (1995) or Bhattacharya's (2007) formula, and constructing interval estimators using these variance estimators for the Gini coefficient are far easier and computationally timesaving compared to those using the bootstrap methods.

Finally, in Chapter 4, we applied analytical and resampling techniques investigated in Chapter 2 and 3 to explore inequality in some health measures among women and children in Bangladesh by employing the Gini coefficient. Using the Bangladesh DHS data on two vital health indicators, BMI among women of age 15-49 years and Hb (anemia) in children aged 6-59 months, we estimated the Gini coefficients, along with associated sampling variances, under stratified two-stage cluster sampling and stratified sampling designs, and formed confidence intervals and hypothesis tests about the Gini coefficients. The latter sampling design was considered when the average cluster size was small and/or there was a large number of singleton clusters in a stratum of the sample, which arose when we classified women/children by their background characteristics, e.g., place of residence, age group, wealth category. In addition, a number of useful descriptive statistics for both well-being variables were reported to further understand the health status of those two quotients for women and children. Amongst other findings, our results showed that health inequalities in these well-being variables for both women and children were well above (more than double) the naturally arising inequality that might be expected for a corresponding "healthy" populations. Furthermore, we found that urban residency, higher level of education and wealth, and the age of a child (under two years of age) were significant factors contributing to worsening the women's BMI and children's Hb inequalities, respectively.

We saw that the sampling variances for $\hat{G}$ using Binder and Kovačević's (1995) and Bhattacharya's (2007) formulae were similar to each other to three decimal places, resulting in almost identical 95% CI estimators and qualitative hypothesis test outcomes, irrespective of assumed samples and sampling designs. For our cases, we observed that the estimated variances for $\hat{G}$ using the bootstrap were no smaller than those using the analytical asymptotic formulae, except for some samples with fewer number of women or children in the clusters or strata, where the bootstrap variance estimates were slightly smaller. Despite this, as the magnitudes of variance estimates were small, interval estimates using bootstrap variances were comparable to those formed using analytical variance estimates. In the small sample cases, we found that the other 95% CI estimates using the bootstrap percentile-*t* method were slightly narrower and right-shifted compared to the three SN approximation interval estimates. However, such differences did not alter any qualitative outcomes of our inequality analysis when we examined hypothesis tests at the usually adopted nominal significance levels. In light of the higher computational issues associated with bootstrapping, these findings suggest that, for samples with similar characteristics to ours, any gain for variance estimators or inference about the Gini coefficient based on the bootstrap methods was trivial compared with using analytical asymptotic variance estimation techniques.

5.2 Future Research

The analysis in the dissertation suggests several avenues for future research. The Gini coefficient, a popular measure in assessing and analyzing inequality in well-being variables, is often criticized for not being able to be readily decomposed into within and between components. Given the importance and relevance of the decomposability property of a standard inequality measure (see, e.g., Bourguignon, 1979), a number of studies, including Araar (2006), Lopez-Feldman (2006), and Dikhanov (1996), have considered decomposition of the Gini coefficient into within- and between- groups under certain conditions and the *iid* sample framework. A topic for future research would be to decompose the Gini coefficient across relevant groupings (e.g., rural/urban, gender), for instance as considered by Araar (2006), determining between- and within-group inequality and associated sampling variances under the complex survey design. Such analyses may aid in our understanding of sources of inequality.

In our simulation experiments in Chapter 3, for ease of demonstration, we have made several assumptions in the data generating process and sample selection technique.

For example, we assume that the random effects for clusters are normally distributed, we only allow for two or five strata in the population, and we assume that a fixed number of households are sampled from each cluster. Expanding each of these assumptions offers further research opportunity, to ascertain the sensitivity of our findings to such assumptions. In addition, considering an alternative approach to generate the finite population can be a potential future research venue. For example, generating the well-being variable under a distributional assumption first, and then dividing it into various strata and clusters using some parametric models.[60] In our empirical illustrations in Chapter 4, we performed multiple hypothesis tests using a single sample in several occasions believing that a null hypothesis may not always be true (see, e.g., Gelman et al., 2012). However, testing a set of hypotheses simultaneously for a specific level of significance is a concern in classical statistics, as it typically increases the probability of observing that some of the results are significant, even all the tests are actually not significant. As such, a multiple testing adjustment, e.g., the Bonferroni correction method (see, e.g., Noble, 2009), can be considered in future research.

Finally, although we too echo Langel and Tillé's (2013) comment that the variance estimation problem for a Gini coefficient estimator has mostly been resolved, and the analytical asymptotic variance estimators for the coefficient are reliable and easily obtainable under a complex survey design, it may be interesting to extend our simulations to cover other resampling methods. One obvious extension is employing the jackknife, as considered by, for example, Karagiannis and Kovačević (2000), Yitzhaki (1991), and Sandström et al. (1985, 1988) for the Gini coefficient. Arnab et al. (2015), Krewski and Rao (1981), and Berger (2007) use the jackknife approach to estimate the variance for an

---

[60] I appreciate Dr Brian Krauth, Department of Economics, Simon Fraser University, BC, for suggesting this potential approach.

estimator of interest (e.g., total, mean or ratio) under the survey sampling design. Key issues associated with using the jackknife with a complex survey sample are size of the sample, and sampling features, such as the number of cluster, sampling techniques for obtaining clusters. For instance, if the sample size is large, as is the case of survey samples, computing variance estimates for the Gini coefficient estimator using the traditional jackknife method (see, e.g., Wolter, 1985, p. 182) can be computationally highly intensive, which may not justify the accuracy gained regarding the estimator (see, e.g., Karagiannis and Kovačević, 2000). Further, if the jackknife variance estimates for the Gini coefficient are similar to those obtained by the analytical asymptotic technique under the survey sampling design, as found by Langel and Tillé (2013) with an *iid* sample, then we would have further affirmation for the latter approach.

Bibliography

Ahmmed, M. (2013). Impact of wealth inequality on child nutrition in Bangladesh. *Paediatrica Indonesia* 53, 299-304.

Aliaga, A. and R. Ren (2006). Optimal sample sizes for two-stage cluster sampling in Demographic and Health Surveys. *DHS Working Papers.* No. 30. Demographic and Health Research Division, USA: ORC Macro

Anand, S. (1983). *Inequality and Poverty in Malaysia: Measurement and Decomposition*. New York: Oxford University Press.

Andrews, D. W. K. (1999). Estimation when a parameter is on a boundary. *Econometrica* 6, 1341-1383

Antal, E. and Y. Tillé (2014). A new resampling method for sampling designs without replacement: the doubled half bootstrap. *Computational Statistics* 29, 1345-1363.

Antal, E., M. Langel and Y. Tillé (2011). Variance estimation of inequality indices in complex sampling designs. Presented at the 58th Congress of the International Statistical Institute, Dublin, Ireland. Available from http://www.2011.isiproceedings.org/papers /450008.pdf, last accessed June 4, 2016.

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, December 14, 2009. Available from www.the-eis.com/data/literature/Namibia_nutrition_2009.pdf, last accessed June 4, 2016.

Araar, A. (2006). On the decomposition of the Gini coefficient: an exact approach, with an illustration using Cameroonian data. *Working Paper 06-02*. Department of Economics, Laval University, Quebec, Canada.

Arasan, J. and M. B. Adam (2014). Double bootstrap confidence interval estimates with censored and truncated data. *Journal of Modern Applied Statistical Methods* 13, 399-419.

Arnab, R., T. Zewotir and D. North (2015). Variance estimation from complex survey designs: a case study of household income and expenditure survey design 2002/03, Botswana. C*ommunications in Statistics—Theory and Methods 44*, 63–79.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory* 2, 244-263.

Ayoya, M. A., I. Ngnie-Teta, M. N. Séraphin, A. Mamadoultaibou, E. Boldon, J. E. Saint-Fleur, L. Koo and S. Bernard (2013). Prevalence and Risk Factors of Anemia among Children 6–59 Months Old in Haiti. Hindawi Publishing Corporation, Volume 2013. Available from http://www.hindawi.com/journals/anemia/2013/502968/, last accessed November 12, 2015.

Bahadur, R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics* 37, 577-580.

Baker, R., F. Greer, and the Committee on Nutrition (2010). Diagnosis and prevention of iron deficiency and iron-deficiency anemia in infants and young children (0-3 years of age). *Pediatrics* 126, 1040-1050.

Bangladesh Demographic and Health Survey (BDHS) (2011). *Policy Briefs.* The US Agency for International Development, Dhaka, Bangladesh (January 2013).

Bangladesh Planning Commission (BPC) (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.

Bangladesh Planning Commission (BPC) (2014). *The Millennium Development Goals: Bangladesh Progress Report 2013.* General Economics Division, Bangladesh Planning Commission, Government of the People's Republic of Bangladesh, August 2014.

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

Bell, B. A., G. B. Morgan, J. D. Kromrey, and J. M. Ferron (2010). The impact of small cluster size on multilevel models: a monte carlo examination of two-level models with binary and continuous predictors. *JSM Proceedings, Survey Research Methods Section*, Alexandria, VA: American Statistical Association, 4057-4067.

Benoist, B. D., E. McLean, I. Egli and M. Cogswell (2008). *Worldwide Prevalence of Anaemia 1993–2005: WHO Global Database on Anaemia*. Geneva, Switzerland: World Health Organization.

Bera, A. and C. Jarque (1981). Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals: monte carlo evidence, *Economics Letters* 7, 313-318.

Berger, Y. G. (2008). A note on the asymptotic equivalence of jackknife and linearization variance estimation for the Gini coefficient. *Journal of Official Statistics* 24, 541-555.

Berger, Y. G. (2007). A jackknife variance estimator for unistage stratified samples with unequal probabilities. *Biometrika 94*, 953-964.

Bhattacharya, D. (2005). Asymptotic inference from multistage samples. *Journal of Econometrics* 126, 145-171.

Bhattacharya, D. (2007). Inference on inequality from household survey data. *Journal of Econometrics* 137, 674-707.

Biewen, M. and S. Jenkins (2006). Variance estimation for generalized entropy and Atkinson inequality indices: the complex survey data case. *Oxford Bulletin of Economics and Statistics* 68, 371-383.

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

Binder, D. and M. S. Kovačević (1995). Estimating some measures of income inequality from survey data: an application of the estimating equations approach. *Survey Methodology* 21, 137-145.

Binder, D. and Z. Patak (1994). Use of estimating functions for estimation from complex surveys. *Journal of the American Statistical Association* 89, 1035-1043.

Borbor, F. M., A. K. Kumi -Kyereme, E. Yendaw and A. Adu-Opong (2014). A study on the determinants of anemia among under-five children in Ghana. *International Journal of Development Research* 4, 858-867

Bourguignon, F. (1979). Decomposable income inequality measures. *Econometrica* 47, 901-920.

Brehm, J. and S. Gates (1993). Donut shops and speed traps: evaluating models of supervision on police behavior. *American Journal of Political Science* 37, 555–581

Bruch. C., R. Münnich, and S. Zins (2011). Deliverable 3.1: Variance estimation for complex surveys. *Research project report FP7-SSh-2007-21322 AMELI.* URL http://ameli.surveystatistics.net/

Caballero, B. (2005). A nutrition paradox – underweight and obesity in developing countries. *New England Journal of Medicine* 352, 1514-1516.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*, New York: Cambridge University Press.

Chang, J. and P. Hall (2015). Double-bootstrap methods that use a single double-bootstrap simulation. *Biometrika* 102, 203-214.

Chang, S., L. Wang, Y. Wang, I.D. Brouwen, F.J. Kok, B. Lozoff, and C. Chen (2011). Iron-deficiency anemia in infancy and social emotional development in preschool-aged Chinese children. *Pediatrics* 127, e927-e933.

Chipman, J. S. (1985). The theory and measurement of income distribution. In R. Basmann and G. Rhodes (Ed.), *Advances in Econometrics* (Vol. 4, p. 135-166). Greenwich, CT: JAI Press Inc.

Clarke, J. A. and N. Roy (2012). On statistical inference for inequality measures calculated from complex survey data. *Empirical Economics* 43*,* 499-524.

Clarke, P. and B. Wheaton (2007). Addressing data sparseness in contextual population research using cluster analysis to create synthetic neighborhoods. *Sociological Methods & Research* 35, 311-351.

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, 929-944.

Cornet, M., J. Y. Le Hesran, N. Fievet, M. Cot, P. Personne, R. Gounoue, M. Beyeme and P. Deloron (1998). Prevalence of and risk factors for anemia in young children in Southern Cameron. *American Journal of Tropical Medicine and Hygiene* 58, 606-611.

Corsi, D. J., H. H. Kyu and S. V. Subramanian (2011). Socioeconomic and geographical patterning of under- and over-nutrition among women in Bangladesh. *The Journal of Nutrition* 141, 631-638.

Cowell, F. A. (1977). *Measuring Inequality*. Oxford, UK: Phillip Allan Publishers Limited.

Cowell, F. A. (1999). Estimation of inequality indices. In Silber, J. (Ed.), *Handbook on Income Inequality Measurement* (p. 269-289). Norwell, MT: Kluwer Academic Publishers

Cowell, F. A. (1989). Sampling variance and decomposable inequality measures. *Journal of Econometrics* 42, 27-41.

Cowell, F. A. and M.-P. Victoria-Feser (2003). Distribution-free inference for welfare indices under complete and incomplete information. *Journal of Economics Inequality* 1, 191–219.

Davidson, R. (2009). Reliable inference for the Gini index. *Journal of Econometrics* 150, 30-40.

Davidson, R. and J. G. MacKinnon (2000). Bootstrap tests: how many bootstraps? *Econometric Reviews* 19, 55–68.

Davidson, R. and J. G. MacKinnon (2004). *Econometric Theory and Methods.* New York: Oxford University Press.

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, 660-667.

Deaton, A. (1997). *Analysis of Household Survey: A Microeconometric Approach to Development Policy.* Baltimore, MD: The Johns Hopkins University Press.

Deville, J.-C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology* 25, 193–204.

Deltas, G. (2003). The small sample bias of the Gini coefficient results and implications for empirical research. *The Review of Economics and Statistics* 85, 226-234.

DiCiccio, T. J. and B. Efron (1996). Bootstrap confidence interval. *Statistical Science* 11, 189-228.

Dikhanov, Y. (1996). Decomposition of inequality based on incomplete information. Paper presented at the IARIW 24th General Conference (August 18-24, 1996), Lillehammer, Norway. Available from http://siteresources.worldbank.org/INTPGI/Resources/13279_Decomposition_of_inequality_based.doc, last accessed June 4, 2016.

Dikhanov, Y. (2005). Trends in global income distribution, 1970-2000, and scenarios for 2015. *Human Development Report 2005,* Human Development Report office, UNDP.

Dippo, C. S. (1984). Variance estimation for nonlinear estimators based upon stratified samples from finite populations. Unpublished PhD dissertation, Department of Statistics, George Washington University.

Dixon, P. M., J. Weiner, T. Mitchell-Olds and R. Woodley (1987). Bootstrapping the Gini coefficient of inequality. *Ecology* 68, 1548-1551.

Eckerstoreer, P., J. Halak, J. Kapeller, B. Schütz, F. Springholz and R. Wildauer (2014). Correcting wealth survey data for the missing rich: The case of Austria. *Working Paper No. 1401,* Department of Economics Johannes Kepler University of Linz, Austria.

Efron, B. (1979). Bootstrap method: another look at the jackknife. *The Annals of Statistics* 7, 1-26.

Ewusie, J., C. Ahiadeke, J. Beyene and J. S. Hamid (2014). Prevalence of anemia among under-5 children in the Ghanaian population: estimates from the Ghana demographic and health survey. BMC Public Health 2014, 14:626. Available from http://www.biomedcentral.com/1471-2458/14/626, last accessed November 12, 2015.

Ferdousi, S. and W. Dehai (2014). Economic growth, poverty and inequality trend in Bangladesh. *Asian Journal of Social Sciences & Humanities* 3, 1-11.

Foster, J. E. and A. A. Shneyerov (1999). A general class of additively decomposable inequality measures*? Economic Theory* 14, 89–111.

Fotso, J. C. (2007). Urban-rural differentials in child malnutrition: Trends and socioeconomic correlates in Sub-Saharan Africa. *Health and Place* 13, 205- 223.

Francisco, C. A. and W.A. Fuller (1986). Estimation of the distribution function with a complex survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 37-45.

Gastwirth, J., R. Modarres and E. Bura (2005). The use of the Lorenz curve, Gini index and related measures of relative inequality and uniformity in securities law. *Metron* 63, 451-469.

Gastwirth, J. L. (1972). The estimation of the Lorenz curve and the Gini Index. *Review of Economics and Statistics* 54, 306-316.

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*, 229-243. Greenwich, CT: JAI Press Inc.

Gelman, A., J. Hill and M. Yajima (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness* 5, 189-211.

Giacomini, R., D. N. Politis and H. White (2013). A warp-speed method for conducting monte carlo experiments involving bootstrap estimators. *Econometric Theory* 29, 567-589.

Giles, D. E. A. (2004). Calculating a standard error for the Gini coefficient: some further results. *Oxford Bulletin of Economics and Statistics* 66, 425-433.

Gini, C. (1914). Sulla misura della concertrazione e della variabilitia 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* LX111, 3-38.

Giorgi, G. M. (1999). Income inequality measurement: the statistical approach. In Silber, J. (Ed.), *Handbook on Income Inequality Measurement* (p. 245-267). Norwell, MT: Kluwer Academic Publishers

Giorgi, G. M., P. Palmitesta and C. Provasi (2006). Asymptotic and bootstrap inference for the generalized Gini indices. *Metron* 64, 107-124.

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

Godambe, V. P. and M. E. Thompson (1984). Robust estimation through estimating equations, *Biometrika* 71, 115-125.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* 31, 1208-1211.

Godambe, V. P. (1976). Conditional likelihood and (unconditionally) optimum estimating equation. *Biometrika* 63, 277-284.

Godambe, V. P. and M. E. Thompson (1978). Some aspects of the theory of estimating equations. *Journal of Statistical Planning and Inference* 2, 95-104.

Greene, W. H. (2008). *Econometric Analysis.* 6th ed., New Jersey: Pearson Prentice Hall.

Hall, A. R. (2005). *Generalized Method of Moments.* Manchester, UK: Oxford University Press.

Haughton, J. and S. R. Khandker (2009). *Handbook on Poverty and Inequality,* Washington, DC: The World Bank.

Hendy, D. F. (1984). Monte Carlo experimentation in econometrics. In *Handbook of Econometrics,* Vol. 2, Z. Griliches and M.D. Intriligator, eds. Amsterdam: North-Holland, Chapter 16.

Hipgrave, D., X. Fu, H. Zhou, Y. Jin, X. Wang, S. Chang, R. Scherpbier, Y. Wang and S. Guo (2014). Poor complementary feeding practices and high anaemia prevalence among infants and young children in rural central and western China. *European Journal of Clinical Nutrition* 68, 916-924.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* 19, 293-325.

Hong, G. (2015). *Causality in a Social World: Moderation, Meditation and Spill-over.* West Sussex, UK: John Wiley & Sons, Ltd.

Hong, R., J. E. Banta and J. A. Betancourt (2006). Relationship between household wealth inequality and chronic childhood under-nutrition in Bangladesh. *International Journal for Equity in Health* 5(15). Available from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1702347/pdf/1475-9276-5-15.pdf, last accessed June 10, 2015.

Hossain, M. G., P. Bharath, S. Aik, P. E. Lestrel, A. Abeer, and T. Kamarul (2012). Body mass index of married Bangladeshi women: trends and association with socio-demographic factors. *Journal of Biosocial Science* 44, 385-399.

Hoque, A. A. and J. A. Clarke (2015). On variance estimation for a Gini coefficient estimator obtained from complex survey data. *Communications in Statistics: Case Studies, Data Analysis and Applications* 1, 39-58.

Hox, J. J. (1998). Multilevel modeling: when and why. In I. Balderjahn, R. Mathar and M. Schader (Eds.). *Classification, data analysis, and data highways* (p. 147-154). New York: Springer Verlag.

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, 5035-5052.

Jędrzejczak, A. (2008). Decomposition of the Gini index by sources of income. *Institute of Advanced Econometrics Research* 14, 441-447.

Kalton, G., J. M. Brick and T. Le (2005). Estimating components of design effects for use in sample design. In United Nation's Household Sample Surveys in Developing and Transition Countries. United Nations Publications, ST/ESA/STAT/SER.F/96. Available from http://www.hunstats.**un**.org/unsd/hh**surveys**/pdf/Chapter_6.pdf, last accessed June 4, 2016.

Karagiannis, E. and M. Kovačević (2000). A method to calculate the Jackknife variance estimator for the Gini coefficient Source. *Oxford Bulletin of Economics and Statistics* 62, 119 -122.

Karoly, L. A. (1992). Changes in the distribution of individual earnings in the United States 1967-1986. *The Review of Economics and Statistics* 74, 107-115.

Kendall, M. G. and A. Stuart (1977). *The Advanced Theory of Statistics*, (4th ed., Vol. 1). London: Charles Griffin.

Ketz, Philipp (2014). Testing near or at the Boundary of the Parameter Space. Available from http://fakultaetsseminar.uni-mannheim.de/material/JMP_PhilippKetz.pdf, last accessed September 21, 2016.

Khan, M. M., A. Krämer, A. Khandoker, L. Prüfer-Krämer and A. Islam (2011). Trends in sociodemographic and health-related indicators in Bangladesh, 1993–2007: will inequities persist? *World Health Organization Bulletin* 89*,* 583–593.

Khusun, H., R. Yip, W. Schultink and D.H. Dillon (1999). World Health Organization hemoglobin cut-off points for the detection of anemia are valid for an Indonesian population. *The Journal of Nutrition* 129, 1669-1674.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons, Inc.

Kish, L. and M.R. Frankel (1974). Inference from complex samples. *Journal of the Royal Statistical Society* 36, 1-37.

Kotecha, P. (2011). Nutritional anemia in young children with focus on Asia and India. *Indian Journal of Community Medicine* 36, 8–16.

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.

Kozak, A., R. A. Kozak, C. L. Staudhammer and S. B.Watts (2008). *Introductory Probability and Statistics: Applications for Forestry and Natural Sciences.* Wallingford: CAB International.

Kreuter, F. and R. Valliant (2007). A survey on survey statistics: what is done and can be done in Stata. *The Stata Journal* 7, 1-21.

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, 1010-1019.

Lake, J. K., C. Power and T. J. Cole (1997). Women's reproductive health: the role of body mass index in early and adult life. *International Journal of Obesity* 21, 432 -438.

Langel, M. and Y. Tillé (2011). Corrado Gini, a pioneer in balanced sampling and inequality theory. METRON - *International Journal of Statistics* LXIX, 45-65.

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.

Lehtonen, R. and E. J. Pahkinen (1995). *Practical Methods for Design and Analysis of Complex Survey.* West Sussex, UK: John Wiley & Sons Ltd.

Lerman, R. I. and S. Yitzhaki (1984). A note on the calculation and interpretation of the Gini index. *Economic Letters* 15, 363-368.

Lerman, R. I. and S. Yitzhaki (1989). Improving the accuracy of estimates of Gini coefficients. *Journal of Econometrics* 42, 43-47.

Litchfield, J. A. (1999). Inequality: methods and tools. Text from World Bank's Web Site on Inequality, Poverty, and Socio-economic Performance. Available from http://www.worldbank.org/poverty/inequal/index.htm, last accessed June 20, 2015.

Lomnicki, Z. A. (1952). The standard error of Gini's mean difference. *Annals of Mathematical Statistics* 23, 635-637.

López, R., V. Thomas and Y. Wang (1998). Addressing the education puzzle: the distribution of education and economic reform. *Policy Research Working Paper,* No. 2031, Washington, DC: The World Bank.

López-Feldman, A. (2006). Decomposing inequality and obtaining marginal effects. *The Stata Journal 6*, 106-111.

Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association* 9, 209-219.

Lozoff, B., F. Corapci, M. J. Burden, N. Kaciroti, R. Anqulo-Barroso, S. Sazawal and M. Black (2007). Preschool-aged children with iron deficiency anemia show altered affect and behavior. *The Journal of Nutrition* 137, 683-689.

Lumley, T. (2010). *Complex Surveys: A Guide to Analysis using R*, New York: John Wiley & Sons.

Luus, R., A. Neethling and T. de Wet (2012). Effectiveness of weighting and bootstrap in the estimation of welfare indices under complex sampling. *South African Statistical Journal* 46*,* 85-114.

Maas, C. J. M. and J. J. Hox (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica* 58, 127-137.

Maas, C. J. M. and J. J. Hox (2005). Sufficient sample sizes for multilevel modeling. *Methodology* 1, 86-92.

Maas, J. and C. Criel (1982). Distribution of primary school enrollments in Eastern Africa. *World Bank Staff Working Papers*, No. 511, Washington, DC: The World Bank.

MacKinnon, J. G. (2006). Bootstrap methods in econometrics. *The Economic Record* 82*,* S2-S18.

Maddah, M., M. R. Eshraghian, A. Djazayery and R Mirdamadi (2003). Association of body mass index with educational level in Iranian men and women. *European Journal of Clinical Nutrition* 57, 819–823.

Mahabub-ul-Anwar, M., U. ROB and M. N. Talukder (2008). Inequality in maternal health care utilization in rural Bangladesh. *International Quarterly of Community Health Education* 27, 281-297.

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, 247-252.

Mebane, W. R. (2000). Coordination, moderation, and institutional balancing in American presidential and house elections. *American Political Science Review* 94, 37–58.

Mehran, F. (1976). Linear measures of income inequality. *Econometrica* 44, 805-809.

Milanovic, B. (1997). A simple way to calculate the Gini coefficient, and some implications. *Economics Letters* 56, 45-49.

Milanovic, B. (1998). *Income, inequality, and poverty during the transition from planned to market economy*. Washington, DC: World Bank. Available from https://www.gc.cuny.edu/CUNY.../Milanovic/.../Income_ineq_poverty_book.pdf, last accessed June 4, 2016.

Milanovic, B. (1999). Explaining the increase in inequality during transition. *Economics of Transition* 7*,* 299–341.

Milanovic, B. (2002). True world income distribution, 1988 and 1993: first calculation based on household surveys alone. *The Economic Journal* 112, 51-92.

Milanovic, B. (2005). *Worlds Apart: Measuring International and Global Inequality*, Princeton: Princeton University Press.

Milanovic, B. (2006). Global income inequality: what it is and why it matters? *DESA Working Paper No. 26,* New York: United Nations. Available from www.un.org/esa/desa/papers/2006/wp26_2006.pdf, last accessed June 4, 2016.

Milanovic, B. (2008). Reform and inequality during the transition: an analysis using panel household survey data. *Policy Research Working Paper 4780,* Washington, D.C.: The World Bank. Available from http://www. http://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-4780, last accessed December 12, 2015.

Milanovic, B. (2009). Global inequality and the global inequality extraction ratio: The Story of the Past Two Centuries. *Policy Research Working Paper 5044,* Washington, DC: The World Bank. Available from http://www.elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-5044, last accessed June 4, 2016.

Mitra, P. and Yemtsov, R. (2006). Increasing inequality in transition economies: is there more to come? *World Bank Policy Research Working Paper*, No. 4007.

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, 385-390.

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, 1012-1023.

Morales, A. F and M. Costa (1998). Income inequality measures: Gini ratio and generalized entropy indexes decomposition. *International Advances in Economic Research* 4, 448-448.

Moran, T. P. (2005a). *Bootstrapping the LIS: Statistical Inference with the Gini Index and Patterns of Inequality in the Global North*. Research report, Department of Sociology, State University of New York – Stony Brook, USA.

Moran, T. P. (2005b). Statistical inference for measures of inequality with a cross-sectional bootstrap application. *Sociological Methods & Research* 34, 296-333.

Moskowitz, C. S., V. E. Seshan, E. R. Riedel and C. B. Begg (2008). Estimating the empirical Lorenz curve and Gini coefficient in the presence of error with nested data. *Statistics in Medicine* 27, 3191-3208.

Murrary, C. and D. Evans (2003). *Health Systems Performance Assessment: Debates, Methods and Empiricism.* Geneva: World Health Organization.

Musgrove, P. (1996). Public and private roles in health, World Bank Discussion Paper No. 339, Washington DC: World Bank. Available from http://www.siteresources.worldbank.org/HEALTHNUTRITIONANDPOPULATION/.../Musgrove-, last accessed June 4, 2016.

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., 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, 428-436.

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, 1348-1357.

Nho, E. (2006). Statistical tests on the regional income inequality in Korea. *The Korean Economic Review* 22, 341-365.

Noble, W. (2009). How does multiple testing correction work? *Nature Biotechnology* 27, 1135-1137.

Nour, N. M. (2008). An introduction to global women's health. *Review in Obstetrics & Gynecology* 1, 33-37.

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, 136-144.

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

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, 123-128.

Ogwang, T. (2004). Calculating a standard error for the Gini coefficient: some further results: reply. *Oxford Bulletin of Economics and Statistics* 66, 435–437.

Omar, A., K. Ibrahim and A. M. Razali (2012). Confidence interval estimation of the shape parameter of Pareto distribution using extreme order statistics. *Applied Mathematical Sciences* 6, 4627-4640.

Osório, M., P. I. Lira, M. Batista-Filho and A. Ashworth (2001). Prevalence of anemia in children 6-59 months old in the State of Pernambuco, Brazil. *Pan American Journal of Public Health* 10, 101-107.

Paolino, P. (2001). Maximum likelihood estimation of models with beta distributed dependent variables. *Political Analysis* 9, 325-346.

Park, I. and H. Lee (2004). Design effects for the weighted mean and total estimators under complex survey sampling, *Survey Methodology,* Catalogue no. 12-001-XIE, Ottawa: Statistics Canada.

Park, I., M. Winglee, J. Clark, K. Rust, A. Sedlak and D. Morganstein (2003). Design Effects and Survey Planning. *Proceedings of the American Statistical Association,* Section on Survey Research Methods [CDROM], Alexandra VA.

Paxton, P., P. Curran, K. Bollen, J. Kirby and F. Chen (2001). Monte carlo experiments: design and implementation. *Structural Equation Modeling* 8, 287-312.

Peng, L. (2011). Empirical likelihood methods for the Gini Index. *Australian and New Zealand Journal of Statistics* 53, 131-139.

Prentice, A. M. (2006). The emerging epidemic of obesity in developing countries. *International Journal of Epidemiology* 35, 93-99.

Pyatt, G., C. N. Chen and J. Fei (1980). The distribution of income by factor components. *Quarterly Journal of Economics* 95, 451-473.

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, 1429-1435.

Quantitative Micro Software (2013). *EViews 8,* CA: Irvine, URL http://www.eviews.com.

Raj, D. (1968). *Sampling Theory.* New York: McGraw-Hill Book Company.

Rao, J. 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, 620-630.

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 (1). Available from http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001367, last accessed January 12, 2015.

Ross, C. E. and C. Wu (1995). The links between education and health. *American Sociological Review* 60, 719-745.

Rutstein, S. and K. Johnson (2004). *The DHS Wealth Index*, Calverton, Maryland: ORC Macro.

Sahn, D. and S. Younger (2006). Changes in inequality and poverty in Latin America: looking beyond income to health and education. *Journal of Applied Economics* 9, 215–234.

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, 113-119.

Scheaffer, R. L., W. M. Ill and R. L. Ott (2006). *Elementary Survey Sampling,* Sixth Edition, Belmont, CA: Thomson Brooks/Cole.

Schechtman, E. and S. Yitzhaki (1987). A measure of association based on Gini's mean difference. *Communications in Statistics – Theory and Methods* 16, 207-231.

Scott, C. D. and J. A. Litchfield (1994). Inequality, mobility and the determinants of income among the rural poor in Chile, 1968-1986. London School of Economics, STICERD DERP DP, No. 53.

Sen, A. K. (1973). *On Economic Inequality*, Oxford: Clarendon Press.

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

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley & Sons, Inc.

Shalit, H. (1985). Calculating the Gini index of inequality for individual data, *Oxford Bulletin of Economics and Statistics* 47, 185-189.

Shao, J. (1994). L-statistics in complex survey problems. *Annals of Statistics* 22, 946-967.

Shorrocks, A. F. (1983). The impact of income components on the distribution of family income. *Quarterly Journal of Economics* 98, 311-326.

Singh, R. K. and S. Patra (2014). Extent of Anaemia among Preschool Children in EAG States, India: A Challenge to Policy Makers. *Anemia,* Available from http://www.hindawi.com/journals/anemia/2014/868752/, last accessed March 20, 2015.

Skinner, C., D. Holt and T. Smith (1989). *Analysis of Complex Surveys*. New York: John Wiley & Sons.

Slater, J., Green, C., Sevenhuyes, G., O'Neil, J., and Edginton, B. (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.

Smith, L., M. Ruel, and A. Ndiaye (2005). Why is child malnutrition lower in urban than in rural areas? Evidence from 36 developing countries. *World Development* 33, 1285-1305.

StataCorp (2007). *Stata Statistical Software: Release 9*. College Station, TX: StataCorp LP.

StataCorp (2009). *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.

Subramanian, S. V. and D. Smith (2006). Patterns, distribution and determinants of under and overnutrition: a population-based study of women in India. *American Journal of Clinical Nutrition* 84, 633-640.

Sudheesh, K. K. and I. Dewan (2013). A simple non-parametric estimator for the Gini index. *Isis/ms/2013/08* (August 2, 2013). Indian Statistical Institute, Delhi Centre, New Delhi 110016, India. Available from http://www.www.isid.ac.in/~statmath/2013/isid201308.pdf, last accessed June 4, 2016.

Summers, R. and A. Heston (1995). The Penn World Tables, Version 5.6, *NBER,* (http://www.nber.org/pub/pwt56/), MA: Cambridge.

Theil, H. (1967). *Economics and Information Theory*. Amsterdam: North-Holland.

Thomas, V., Y. Wang and X. Fan (2001). Measuring education inequality: Gini coefficients of education. *Policy Research Woking Paper* 2525, Washington, DC: The World Bank.

U.S. Agency for International Development (USAID), National Institute of Population Research and Training (NIPORT), and Mitra and Associates (2011). *Policy Briefs: Bangladesh Demographic and Health Survey 2011*, Dhaka: USAID.

Uddin, M. T., M. N. Islam, A. Kabir and D. C. Nath (2012). A new index for measuring aging inequality: an application to Asian countries. *Health* 4, 685-694.

UNDP (United Nations Development Programme) (2010). *Human Development Report 2010.* NY: UNDP.

UNDP (United Nations Development Programme) (2014). *Human Development Report 2014.* NY: UNDP.

United Nations (UN) (2005). *Designing Household Survey Samples: Practical* Guidelines, Series F No. 98, Department of Economic and Social Affairs, New York: UN.

United Nations Children’s Fund (UNICEF) (2009). *State of the World’s Children 2010: Celebrating 20 years of the Convention on the Right of the Child.* UNICEF: New York.

United Nations Children's Fund (UNICEF) (2011). *Bangladesh a perspective on gender equality in Bangladesh; from young girl to adolescent: what is lost in transition*? Dhaka: UNICEF.

Van Ourti, T. and P. Clarke (2011). A simple correction to remove the bias of the Gini coefficient due to grouping. *The Review of Economics & Statistics* 93, 982-984.

Wang, W. (2013). Assessing trends in inequalities in maternal and child health and health care in Cambodia. *DHS Further Analysis Report No.* 86, Maryland: ICF International.

WHO (2001). *Iron Deficiency Anemia: Assessment, Prevention and Control, A Guide for Programme Managers.* Geneva: WHO/UNICEF/UNU.

WHO (2008). *Global Burdon of Diseases 2004 update.* Geneva: The World Health Organization.

WHO (2011). *Haemoglobin concentrations for the diagnosis of anemia and assessment of severity.* Vitamin and Mineral Nutrition Information System, (WHO/NMH/NHD/MNM/11.1), Geneva: The World Health Organization.

WHO (2014). *World Health Statistics 2014.* Geneva: The World Health Organization.

Wold, H. (1935). A study of the mean difference, concentration curves and concentration ratios. *Metron* 12, 39-58.

Wolter. K. (2007). *Introduction to Variance Estimation.* Second ed., New York: Springer.

Woodruff, R. S. (1971). A simple method for approximating the variance of a complicated estimate. *Journal of the American Statistical Association* 66, 411- 414.

Wooldridge, J.M. (2001). Applications of generalized method of moments estimation. *Journal of Economic Perspectives* 15, 87-100.

World Health Organization (WHO). (2015). Obesity and overweight. *Fact sheet Number 311* (January 2015). Available from http://www.who.int/mediacentre/factsheets /fs311/en/, last accessed November 07, 2015.

Wu, C. and J. N. K. Rao (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *The Canadian Journal of Statistics* 34, 359-375.

Yitzhaki, S. (1991). Calculating jackknife variance estimators for parameters of the Gini method. *Journal of Business & Economic Statistics,* 9, 235-239.

Yitzhaki, S. and E. Schechtman (2013). *The Gini Methodology: A Primer on a Statistical Methodology.* New York: Springer Science+Business Media.

Zhao, A., Y. Zhang, Y. Peng, J. Li, T. Yang, Z. Liu, Y. Lv and P. Wang (2012). Prevalence of anemia and its risk factors among children 6-36 months old in Burma. *American Journal of Tropical Medicine and Hygiene* 87, 306-311.

Appendix A

The remarkable popularity and acceptance of the Gini coefficient is reflected in the efforts of modifying and rewriting its expression for various purposes by many scholars over the last century without any alteration of the fundamental message of the measure. In this appendix, we present some of the commonly used expressions of the measure.

Gini (1914) proposed the concentration ratio, R, as a measure of inequality in a random variable Y. Let the distribution function be discrete and $y \in [0, \infty)$ take n values that can be arranged in a non-decreasing order such that $y_1 \leq y_2 \leq \cdots \leq y_n$. These are commonly called the order statistics for a series. Let y_i $(i = 1, \dots, n)$ denote the kth element in the sequence. For two arbitrary values $\{k, l\}$ of i where $k < l$, it can be shown that $\frac{\sum_{i=1}^{k} y_i}{\sum_{i=1}^{l} y_i} \leq \frac{k}{l}$ and for $l = n$, the expression is $\frac{\sum_{i=1}^{k} y_i}{\sum_{i=1}^{n} y_i} \leq \frac{k}{n}$.

Now, assuming $\sum_{i=1}^{k} y_i = A_k$, $\sum_{i=1}^{n} y_i = A_n$, $\frac{A_k}{A_n} = q_k$ and $\frac{k}{n} = p_k$, where $(k < n)$, we have (with strict inequality) $p_k > q_k$. Then noting that the larger is $p_k - q_k$, the stronger the inequality in y, Gini defined the concentration ratio as $R = \frac{\sum_{k=1}^{n-1}(p_k - q_k)}{\sum_{k=1}^{n-1} p_k}$.

The height of a rectangular column with base $1/n$ that reaches at the line of equality is p_k and q_k is the height of the column with base $1/n$ that reaches the Lorenz curve. Substituting in the values for p_k and q_k, we have $R = 1 - \frac{2}{(n-1)A_n} \sum_{k=1}^{n-1} A_k$. Using that $\sum_{k=1}^{n-1} k = n(n-1)/2$ and as $\sum_{k=1}^{n-1} A_k = \sum_{k=1}^{n-1} (n-k) y_k$, we have

$$R = \frac{2 \sum_{k=1}^{n} (k-1) y_k}{(n-1)A_n} - 1. \qquad (A.1)$$

To extend this form to the commonly considered Gini coefficient, in terms of the Lorenz curve, for a continuous distribution, we undertake the following algebraic manipulation

$$R = 1 - \frac{2}{(n-1)\sum_{i=1}^{n} y_i} \sum_{k=1}^{n-1} \left[\sum_{i=1}^{k} y_i \right]$$

$$= 1 - \frac{2}{(n-1)} \sum_{k=1}^{n-1} \left[\frac{\sum_{i=1}^{k} y_i}{\sum_{i=1}^{n} y_i} \right] \qquad (A.2)$$

Let the pair (p_k, q_k)denote the proportion in the population of the k poorest persons and their share in the total of y respectively, with 0 percent of the population receiving 0 of y; i.e., $(p_0, q_0) = (0,0)$. The Lorenz function is then defined as the set of points (p_k, q_k), where $q_k = L(p_k) \equiv \frac{\sum_{i=1}^{k} y_i}{\sum_{i=1}^{n} y_i}, k = 0,1,\ldots,n.$ Then

$$R = 1 - \frac{2}{(n-1)} \sum_{k=1}^{n-1} [L(p_k)]$$

$$= 1 - \frac{2}{(n-1)} \sum_{k=1}^{n-1} \left[\frac{1}{\bar{y}} \frac{\sum_{i=1}^{n} y_i I\{y_i \leq y_k\}}{n} \right], \qquad (A.2)$$

where, $I\{y_i \leq y_k\}$ is an indicator function such that $I\{y_i \leq y_k\} = \begin{cases} 1 & \text{if } \; y_i \leq y_k \\ 0 & \text{if } \; y_i > y_k \end{cases}.$

Moving to a continuous variable, with a population, let $F(y)$ be the continuous distribution function of y and $(y) = \frac{1}{n}$. Then $R = 1 - \frac{2}{(n-1)} \sum_{k=1}^{n-1} \left[\frac{1}{\mu} \int_0^{\infty} y_i I\{y_i \leq y_k\} \, dF(y) \right]$. Let, $\xi_p = y_k$ such that the pth quantile of F is $F^{-1}(p) = \xi_p$, where $F^{-1}(p) = \inf\{y \mid F(y) \geq p\}$. Then

$$R = 1 - 2\int_0^{\infty} \left[\frac{1}{\mu} \int_0^{\xi_p} y \, dF(y) \right] dF(\xi_p) = 1 - 2\int_0^{\infty} \left[L\left(F(\xi_p) \right) \right] dF(\xi_p)$$

$$= 1 - 2\int_0^{1} L(p) dp, \qquad (A.3)$$

where $L(p)$ is the Lorenz curve. Expression $(A.3)$ is one version of the Gini coefficient commonly examined. However, over time, in remembrance of Corrado Gini, the concentration ratio, R, has been called the Gini coefficient, G. From this expression, the Gini coefficient can be also written as

$$\begin{aligned}
G &= 1 - 2\int_0^1 L(p)dp \\
&= 1 - 2\int_0^\infty \frac{1}{\mu}\int_0^{\xi_p} y\, dF(y)dF(\xi_p) = 1 - 2\frac{1}{\mu}\int_0^\infty \int_y^\infty y\, dF(\xi_p)dF(y) \\
&= 1 - 2\frac{1}{\mu}\int_0^\infty y[1 - F(y)]dF(y) && (A.4) \\
&= 1 - \left\{2\frac{1}{\mu}\int_0^\infty ydF(y) - 2\frac{1}{\mu}\int_0^\infty yF(y)dF(y)\right\} \\
&= \frac{2}{\mu}\int_0^\infty yF(y)dF(y) - 1 && (A.5) \\
&= \frac{2}{\mu}\left\{\int_0^\infty yF(y)dF(y) - \frac{1}{2}\int_0^\infty ydF(y)\right\} \\
&= \frac{2}{\mu}\left\{\int_0^\infty y\left(F(y) - \frac{1}{2}\right)dF(y)\right\} && (A.6)
\end{aligned}$$

The expression $(A.6)$ is used to derive the Gini coefficient in terms of the covariance between the cumulative density function and the individual observations; this is provided shortly. From $(A.4)$ we can derive the expression for Gini in terms of the Lorenz area

$$\begin{aligned}
G &= 1 - \frac{2}{\mu}\int_0^\infty y[1 - F(y)]dF(y) \\
&= 2 \times \frac{1}{2} - \frac{2}{\mu}\int_0^\infty y[1 - F(y)]dF(y) \\
&= 2\int_0^\infty d\left(F(\xi_p)\right)^2 - \frac{2}{\mu}\int_0^\infty \int_y^\infty y\, dF(\xi_p)dF(y)\text{; where } \int_0^\infty d\left(F(\xi_p)\right)^2 = \frac{1}{2} \\
&= 2\int_0^\infty F(\xi_p)dF(\xi_p) - \frac{2}{\mu}\int_0^\infty \int_0^{\xi_p} y\, dF(y)dF(\xi_p)
\end{aligned}$$

$$= 2\int_0^{\infty} \left(F(\xi_p) - L\left(F(\xi_p)\right)\right) dF(\xi_p)$$

$$= 2\int_0^{1} \left(p - L(p)\right) dp$$

$$= 2 \times Lorenz\ Area \qquad (A.7)$$

Following Gini (1914), Hoeffding (1948) uses the mean difference version of the Gini coefficient to calculate its variance. Kendall and Stuart (1958, p. 46-50) provide details in their *Advanced Theory of Statistics* book; see also Mehran (1976) and Chipman (1985). Therein, the Gini index is defined as one-half of Gini's relative mean difference.

The mean difference in the index is the expected distance between two observations of the variate, say Y. For y_i and y_j $\{i, j = 1, \dots, n\}$, non negative discrete variates from the same distribution, the mean difference without repetition is given by

$$\Delta = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|$$

For a continuous distribution the mean difference is $\Delta = \int_0^{\infty} \int_0^{\infty} |y_i - y_j|\, dF(y_i) dF(y_j)$.

The Gini coefficient is then given by

$$G = \frac{\Delta}{2\mu} = \frac{1}{2 \cdot \int_0^{\infty} y\, dF(y)} \left[\int_0^{\infty} \int_0^{\infty} |y_i - y_j|\, dF(y_i) dF(y_j)\right] \qquad (A.8)$$

Peng (2011) simplifies the expression (A.8) as follows

$$G = \frac{1}{\int_0^{\infty} y\, dF(y)} \left[\int_0^{\infty} \int_0^{y_j} (y_j - y_i)\, dF(y_i) dF(y_j)\right]$$

$$= \frac{1}{\int_0^{\infty} y\, dF(y)} \left[\int_0^{\infty} y_j\, F(y_j) dF(y_j) - \int_0^{\infty} \int_0^{y_j} y_i\, dF(y_i) dF(y_j)\right]$$

$$= \mu^{-1} \left[\int_0^{\infty} y_j\, F(y_j) dF(y_j) - \int_0^{\infty} \int_{y_i}^{\infty} y_i\, dF(y_j) dF(y_i)\right]$$

$$= 2\mu^{-1} \int_0^{\infty} y_j\, F(y_j) dF(y_j) - 1 \qquad (A.9)$$

For a discrete distribution, from Kendall and Stuart (1958, p. 49-50), the expression can be simplified as follows

$$\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i - y_j| = 2\sum_{i=1}^{n}\sum_{j=1}^{n}\max(0, y_i - y_j)$$

$$= 2\sum_{i \geq j}^{n}(y_i - y_j) = 2\sum_{k=1}^{n-1}k(n-k)(y_{k+1} - y_k)$$

The Gini coefficient is then given by

$$G = \frac{\Delta}{2\mu} = \frac{2\sum_{k=1}^{n-1}k(n-k)(y_{k+1} - y_k)}{2n(n-1)\mu}$$

$$= \frac{\sum_{k=1}^{n-1}k(n-k)(y_{k+1} - y_k)}{n(n-1)\mu} \qquad (A.10)$$

When the distribution function of y is $F(y)$, the Gini coefficient, in terms of the distribution function, is given by

$$G = \frac{\Delta}{2\mu} = \frac{2\sum_{k=1}^{n-1}F_k(y)\big(1 - F_k(y)\big)}{2\mu}$$

$$= \frac{\sum_{k=1}^{n-1}F_k(y)\big(1 - F_k(y)\big)}{\mu} \qquad (A.11)$$

Chipman (1985, p. 154) derived the mean difference function applying the discrete version of the Heaviside function $h_{ij} = \begin{cases} 0, & i < j \\ i, & i \geq j \end{cases}$. The $n \times n$ matrix $\boldsymbol{H} = [h_{ij}]$ is such that the sum of the row is $\sum_{i=1}^{n} h_{ij} = i$ and the sum of the column is $\sum_{i=1}^{n} h_{ij} = n - j + 1$. Therefore,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i - y_j| = 2\sum_{\substack{i=1 \\ i>j}}^{n}\sum_{j=1}^{n}(y_i - y_j) = 2\sum_{i=1}^{n}\sum_{j=1}^{n}h_{ij}\,(y_i - y_j)$$

$$= 2\left(\sum_{i=1}^{n}\sum_{j=1}^{n}h_{ij}\,y_i - \sum_{i=1}^{n}\sum_{j=1}^{n}h_{ji}y_i\right)$$

$$= 2\sum_{i=1}^{n}\left(\sum_{j=1}^{n}h_{ij} - \sum_{j=1}^{n}h_{ji}\right)y_i = 2\sum_{i=1}^{n}(2i - n - 1)y_i.$$

Hence, the Gini coefficient is given by

$$G = (2\mu)^{-1}\left[\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i - y_j\right|\right]$$

$$= \frac{\sum_{i=1}^{n}(2i-n-1)y_i}{n(n-1)\mu} = \frac{\sum_{i=1}^{n}(2i-n-1)y_i}{n(n-1)(\sum_{i=1}^{n} y_i)/n}$$

$$= \frac{\sum_{i=1}^{n}(2i-n-1)y_i}{(n-1)\sum_{i=1}^{n} y_i} \qquad (A.12)$$

If the mean difference with repetition is considered, the above expression for the Gini coefficient can be rewritten in a form that has been used by some authors, including Sendler (1979) and Nygård and Sandström (1989):

$$G = \frac{\sum_{i=1}^{n}(2i-n-1)y_i}{n\sum_{i=1}^{n} y_i}$$

$$= \frac{\sum_{i=1}^{n}\left(\frac{2i}{n}-\frac{1}{n}-1\right)y_i}{\sum_{i=1}^{n} y_i}. \qquad (A.13)$$

From the continuous mean difference Gini formula, Shalit (1985) used the identity $\left|y_i - y_j\right| = y_i + y_j - 2\min(x,y)$ to give that

$$G = \mu^{-1}\left[\int_{y_{max}}^{y_{max}}(1-F(y))dy - \int_{y_{max}}^{y_{max}}(1-F(y))^2 dy\right]$$

$$= \mu^{-1}\left[\int_{y_{max}}^{y_{max}} F(y)(1-F(y))dy\right], \qquad (A.14)$$

where $y_{max} < \infty$ and $y_{min} \geq 0$.

Expression $(A.14)$ has been used to derive the Gini coefficient in terms of the covariance between the variable of study and its distribution function, e.g., Lerman and Yitzhaki (1984), Shalit (1985), Ogwang (2000). Giles (2004) also adopted this to propose a way to estimate the standard error of the Gini coefficient by a regression approach.

To derive the covariance approach of the Gini coefficient we use the integration by part method. Assuming $F(y)[1-F(y)] = u$ and $y = v$, and replacing the limits as $F(y_{max}) = 1$ and $F(y_{min}) = 0$, the solution is

$$G = \mu^{-1}\left[\left[F(y)\big(1 - F(y)\big)y\right]\Big|_{y_{min}}^{y_{max}} \; - \int_{y_{min}}^{y_{max}} yd(F(y)[1 - F(y)])\right]$$

$$= \mu^{-1}\left[0 \; - \int_{y_{min}}^{y_{max}} y\{f(y) - 2f(y)F(y)\}dy\right]$$

$$= \mu^{-1}\left[2\int_{y_{min}}^{y_{max}} y\left\{F(y) - \frac{1}{2}\right\}f(y)dy\right]$$

$$= 2\mu^{-1}\int_{0}^{1} y(F)\left(F(y) - \frac{1}{2}\right)dF(y) \qquad (A.15)$$

where, $y(F)$ is the inverse function of $F(y)$. When F is uniformly distributed between [0,1], it has a mean of $\frac{1}{2}$. Then the Gini coefficient is given by

$$G = \frac{2}{\mu}Cov[y, F(y)] \qquad (A.16)$$

Sen (1973) proposed another formula for the Gini index using lower weights for higher values of y and higher weights for lower values of y, and showed that the rank based weights are inversely related to the size of y

$$G = \frac{n+1}{n} - \frac{2}{n^2\mu}\sum_{i=1}^{n}(n + 1 - i)y_i \qquad (A.17)$$

Milanovic (1997) has proposed a simple way to calculate G that contains only three components: the coefficient of variation, linear correlation coefficients with ranks of the variable and a constant. For a sufficiently large sample, G is given by

$$G = \frac{1}{\sqrt{3}}\frac{\sigma_y}{\bar{y}}\rho\big(y, r_y\big) \qquad (A.18)$$

where, σ_y is the standard deviation of y, $\rho\big(y, r_y\big)$ is the correlation coefficient between y and r_y, y is the variable of interest, and r_y is the ranks of all recipients of y according to the y value.

Appendix B

B.1 Deriving the expression for an estimator for the Gini coefficient using a complex survey sample.

A generic estimator of the Gini coefficient is given by

$$\hat{G} = 1 - 2\int_0^1 \hat{L}(p)\,dp,$$

for which an estimator for the Lorenz curve, $L(p) = \frac{1}{\mu}\left(\int_0^\infty yI\{y \leq \xi_p\}dF(y)\right)$, with a complex survey is given by

$$\hat{L}(p) = \frac{1}{\hat{\mu}}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\,y_{hci}I\{y_{hci} \leq \hat{\xi}_p\},$$

where $\hat{\mu} = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\,y_{hci}$ and $\hat{\xi}_p = \hat{F}^{-1}(p) = \inf\{y_{hci} \in s\colon \hat{F}(y_{hci}) \geq p\}$. Then

$$\hat{G} = 1 - \frac{2}{\hat{\mu}}\int_0^1 \left[\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\,y_{hci}I\{y_{hci} \leq \hat{\xi}_p\}\right]dp$$

$$= 1 - \frac{2}{\hat{\mu}}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\,y_{hci}\int_0^1 I\{y_{hci} \leq \hat{\xi}_p\}\,dp$$

$$= 1 - \frac{2}{\hat{\mu}}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\,y_{hci}\left[1 - \int_0^1 I\{y_{hci} \geq \hat{\xi}_p\}\,dp\right]$$

$$= 1 - \frac{2}{\hat{\mu}}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\,y_{hci} + \frac{2}{\hat{\mu}}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\,y_{hci}\left[\int_0^1 I\{y_{hci} \geq \hat{\xi}_p\}\,dp\right]$$

Using $\hat{\mu} = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{k} w_{hci}\,y_{hci}$ and $p = \hat{F}(\hat{\xi}_p)$, we obtain,

$$\hat{G} = 1 - \frac{2}{\hat{\mu}}\hat{\mu} + \frac{2}{\hat{\mu}}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\,y_{hci}\left[\int_0^\infty I\{\hat{\xi}_p \leq y_{hci}\}\,d\hat{F}(\hat{\xi}_p)\right]$$

$$= \frac{2}{\hat{\mu}}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\,y_{hci}\left[\int_0^\infty I\{\hat{\xi}_p \leq y_{hci}\}\,d\hat{F}(\hat{\xi}_p)\right] - 1$$

Substituting a plug-in estimator for $\int_0^\infty I\{\hat{\xi}_p \leq y_{hci}\} d\hat{F}(\hat{\xi}_p)$, which is given by $\sum_{r=1}^{H}\sum_{s=1}^{r_h}\sum_{t=1}^{m} w_{rst} I\{y_{rst} \leq y_{hci}\} = \hat{F}(y_{hci})$ in the above equation, we obtain the following plug-in estimator for the Gini coefficient with a complex survey:

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

B.2 We show that for an *iid* sample

$$u_j^* = \frac{(\hat{Z}_j - \bar{Z})}{\hat{\mu}}, \;\; j = 1, \dots, M$$

where $u_j^* = \frac{2}{\hat{\mu}}\left[y_j\left(\hat{F}(y_j) - \left(\frac{\hat{G}+1}{2}\right)\right) + B(y_j) - \frac{\hat{\mu}}{2}(\hat{G}+1)\right]$, $\hat{Z}_j = -(\hat{G}+1)y_j + 2\left(\hat{F}(y_j)y_j - \widehat{m}(y_j)\right)$, $B(y_j) = \hat{\mu} - \widehat{m}(y_j)$ and $\widehat{m}(y_j) = \frac{1}{M}\sum_{i=1}^{M} y_i I\{y_i \leq y_j\}$.

We have,

$$\begin{aligned}
u_j^* &= \frac{2}{\hat{\mu}}\left[y_j\left(\hat{F}(y_j) - \left(\frac{\hat{G}+1}{2}\right)\right) + B(y_j) - \frac{\hat{\mu}}{2}(\hat{G}+1)\right] \\
&= \frac{2}{\hat{\mu}}\left[\hat{F}(y_j)y_j - \frac{\hat{G}+1}{2}y_j + \hat{\mu} - \widehat{m}(y_j) - \frac{(\hat{G}+1)}{2}\hat{\mu}\right] \\
&= \frac{1}{\hat{\mu}}\left[-(\hat{G}+1)y_j + 2\hat{F}(y_j)y_j - 2\widehat{m}(y_j) + 2\hat{\mu} - (\hat{G}+1)\hat{\mu}\right].
\end{aligned}$$

Following Davidson (2009), let $\hat{Z}_j = -(\hat{G}+1)y_j + 2\hat{F}(y_j)y_j - 2\widehat{m}(y_j)$, so that

$$\begin{aligned}
\bar{Z} &= \frac{1}{M}\sum_{j=1}^{M}\hat{Z}_j = \frac{1}{M}\sum_{j=1}^{M}\left[-(\hat{G}+1)y_j + 2\hat{F}(y_j)y_j - 2\widehat{m}(y_j)\right] \\
&= -\frac{1}{M}\sum_{j=1}^{M}(\hat{G}+1)y_j + 2\frac{1}{M}\sum_{j=1}^{M}\hat{F}(y_j)y_j - 2\frac{1}{M}\sum_{j=1}^{M}\widehat{m}(y_j) \\
&= -(\hat{G}+1)\hat{\mu} + (\hat{G}+1)\hat{\mu} - 2\frac{1}{M}\sum_{j=1}^{M}\left[\frac{1}{M}\sum_{i=1}^{M} y_i I\{y_i \leq y_j\}\right]
\end{aligned}$$

$$
\begin{aligned}
&= -2\frac{1}{M}\sum_{j=1}^{M}\left[\frac{1}{M}\sum_{i=1}^{M} y_i I\{y_i \le y_j\}\right] = -2\frac{1}{M}\sum_{i=1}^{M} y_i \left(\frac{1}{M}\sum_{j=1}^{M} I\{y_i \le y_j\}\right) \\
&= -\frac{2}{M}\sum_{i=1}^{M} y_i \left[1-\hat{F}(y_i)\right] = -\frac{2}{M}\sum_{i=1}^{M} y_i + \frac{2}{M}\sum_{i=1}^{M} y_i \,\hat{F}(y_i) \\
&= -2\hat{\mu} + \left(\hat{G}+1\right)\hat{\mu}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
u_j^* &= \frac{1}{\hat{\mu}}\left[\left(-\left(\hat{G}+1\right)y_j + 2\hat{F}\left(y_j\right)y_j - 2\widehat{m}\left(y_j\right)\right) - \left(-2 + \left(\hat{G}+1\right)\hat{\mu}\right)\right] \\
&= \frac{1}{\hat{\mu}}\left[\hat{Z}_j - \bar{Z}\right].
\end{aligned}
$$

B.3 Proof of the equivalence of the approximations for $\left(\hat{G}-G\right)$ obtained by Bhattacharya (2007) and Binder and Kovačević (1995). Specifically, we show that the approximation provided by Bhattacharya (2007)

$$
\hat{G} - G \approx -2\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hic}\,\Phi_{hci}\,,
$$

and Binder and Kovačević (1995)

$$
\hat{G} - G \approx \sum_{h=1}^{H}\sum_{c=1}^{n_h} u_{hc}^*\,,
$$

are equivalent. To achieve this, we need to show that $-2\sum_{i=1}^{m}\Phi_{hci} = u_{hc}^*$. We start with Φ_{hci}, which is given by

$$
\Phi_{hci} = \int_0^1 \left(\frac{1}{\mu}\left[y_{hci} I\{y_{hci} \le \xi_p\} + \xi_p\left(p - I\{y_{hci} \le \xi_p\}\right) - \frac{\alpha(p)}{\mu} y_{hci}\right]\right) dp.
$$

Using that $p = F\left(\xi_p\right)$ and $\frac{\alpha(p)}{\mu} = L(p)$, and a change of variable, we have

$$
\Phi_{hci} = \frac{1}{\mu}\left[\begin{array}{c} \int_0^\infty y_{hci} I\{y_{hci} \le \xi_p\} dF\left(\xi_p\right) \\ + \int_0^\infty \xi_p F\left(\xi_p\right) dF\left(\xi_p\right) - \int_0^\infty \xi_p I\{y_{hci} \le \xi_p\} dF\left(\xi_p\right) - \int_0^1 y_{hci} L(p) dp \end{array}\right]. \quad (B1)
$$

We first rearrange the first component in expression (B1), $\int_0^\infty y_{hci} I\{y_{hci} \leq \xi_p\} dF(\xi_p)$, as

$$\int_0^\infty y_{hci} I\{y_{hci} \leq \xi_p\} dF(\xi_p) = y_{hci}\left[1 - \int_0^\infty I\{\xi_p \leq y_{hci}\} dF(\xi_p)\right]$$

$$= y_{hci}[1 - F(y_{hci})].$$

For the second component, we rearrange the Gini coefficient equation as

$$\begin{aligned}
G &= 1 - 2\int_0^1 L(p)\, dp \\
&= 1 - 2\int_0^\infty \left[\frac{\alpha(p)}{\mu}\right] dF(\xi_p) \\
&= 1 - \frac{2}{\mu}\int_0^\infty \int_0^{\xi_p} y_{hci}\, dF(y_{hci})\, dF(\xi_p), \qquad \text{where } \alpha(p) = \int_0^{\xi_p} y_{hci}\, dF(y_{hci}) \\
&= 1 - \frac{2}{\mu}\int_0^\infty \int_{y_{hci}}^\infty y_{hci}\, dF(\xi_p) dF(y_{hci}) \qquad \text{[a change in variable]} \\
&= 1 - \frac{2}{\mu}\int_0^\infty \int_{\xi_p}^\infty \xi_p\, dF(y_{hci}) dF(\xi_p) \qquad \text{[a change in variable]} \\
&= 1 - \frac{2}{\mu}\int_0^\infty \xi_p [1 - F(\xi_p)] dF(\xi_p) \\
&= 1 - \frac{2}{\mu}\int_0^\infty \xi_p dF(\xi_p) + \frac{2}{\mu}\int_0^\infty \xi_p F(\xi_p) dF(\xi_p) \\
&= \frac{2}{\mu}\int_0^\infty \xi_p F(\xi_p) dF(\xi_p) - \frac{2}{\mu}\mu - 1 \\
&= \frac{2}{\mu}\int_0^\infty \xi_p F(\xi_p) dF(\xi_p) - 1.
\end{aligned}$$

Rearranging this expression, we obtain

$$\int_0^\infty \xi_p F(\xi_p) dF(\xi_p) = \frac{G+1}{2}\mu,$$

which is the second component in equation (B1). The third component in expression (B1) is

$$\int_0^\infty \xi_p I\{y_{hci} \leq \xi_p\} dF(\xi_p) = \int_0^\infty \xi_p I\{\xi_p \geq y_{hci}\} dF(\xi_p).$$

The final component in expression (B1) is rearranged as follows. We have

$$\int_0^1 y_{hci} L(p)dp = y_{hci}\int_0^1 L(p)dp$$

$$= y_{hci}\left(\frac{1-G}{2}\right).$$

Substituting these parts into expression (B1), we obtain

$$\Phi_{hci} = \frac{1}{\mu}\left[\begin{array}{c} y_{hci} - F(y_{hci})y_{hci} \\ +\frac{G+1}{2}\mu - \int_0^\infty \xi_p I\{\xi_p \geq y_{hci}\}dF(\xi_p) - y_{hci}\left(\frac{1-G}{2}\right) \end{array}\right]$$

$$= \frac{1}{\mu}\left[\begin{array}{c} -F(y_{hci})y_{hci} \\ +\left(\frac{G+1}{2}\right)\mu - \int_0^\infty \xi_p I\{\xi_p \geq y_{hci}\}dF(\xi_p) + \left(\frac{G+1}{2}\right)y_{hci} \end{array}\right]. \qquad (B2)$$

A plug-in estimator for Φ_{hci} as given in expression (2C2)

$$\widehat{\Phi}_{hci} = \frac{1}{\hat{\mu}}\left[\begin{array}{c} -\hat{F}(y_{hci})y_{hci} \\ +\frac{\hat{G}+1}{2}\hat{\mu} - \sum_{r=1}^{H}\sum_{s=1}^{n_r}\sum_{t=1}^{m} w_{rst}y_{rst}I\{y_{rst} \geq y_{hci}\} + \frac{\hat{G}+1}{2}y_{hci} \end{array}\right].$$

Using this, we have $-2\sum_{i=1}^{m} w_{hci}\widehat{\Phi}_{hci} =$

$$\tfrac{2}{\hat{\mu}}\textstyle\sum_{i=1}^{m} w_{hci}\left[\begin{array}{c} \hat{F}(y_{hci})y_{hci} - (\hat{G}+1)\frac{\hat{\mu}}{2} \\ +\sum_{r=1}^{H}\sum_{s=1}^{n_r}\sum_{t=1}^{m} w_{rst}y_{rst}I\{y_{rst} \geq y_{hci}\} - \frac{\hat{G}+1}{2}y_{hci} \end{array}\right]$$

$$= \frac{2}{\hat{\mu}}\sum_{i=1}^{m} w_{hci}\left[\left(\hat{F}(y_{hci}) - \frac{\hat{G}+1}{2}\right)y_{hci} + B(y_{hci}) - (\hat{G}+1)\frac{\hat{\mu}}{2}\right]$$

$$= u_{hc}^{*}\,.$$

It directly follows then that the approximations for $(\hat{G}-G)$ given by Bhattacharya and Binder and Kovačević are equivalent. Hence, the subsequent variance estimators for $\hat{G}$ are also equivalent based on Theorem 1.

B.4 We show that for an *iid* sample

$$\bar{u}^{*} = M^{-1}\sum_{j=1}^{M} u_j^{*} = 0,$$

where $u_j^{*} = \frac{2}{\hat{\mu}}\left[y_j\left(\hat{F}(y_j) - \left(\frac{\hat{G}+1}{2}\right)\right) + B(y_j) - \frac{\hat{\mu}}{2}(\hat{G}+1)\right]$ and

$B(y_j) = \frac{1}{M}\sum_{i=1}^{M} y_i I\{y_i \geq y_j\}$. We also show that for a complex survey

$$\bar{u}^* = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}u_{hci}^* = 0,$$

where $u_{hci}^* = \frac{2}{\hat{\mu}}\left[y_{hci}\left(\hat{F}(y_{hci}) - \frac{(\hat{G}+1)}{2}\right) + B(y_{hci}) - \frac{\hat{\mu}}{2}(\hat{G}+1)\right]$, and

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

Proof: For the *iid* case, we have

$$\bar{u}^* = M^{-1}\sum_{j=1}^{M}\frac{2}{\hat{\mu}}\left[y_j\left(\hat{F}(y_j) - \left(\frac{\hat{G}+1}{2}\right)\right) + B(y_j) - \frac{\hat{\mu}}{2}(\hat{G}+1)\right]$$

$$= \frac{2}{\hat{\mu}}\left[\begin{array}{c}\frac{1}{M}\sum_{j=1}^{M} y_j\,\hat{F}(y_j) \\ -\frac{1}{M}\sum_{j=1}^{M} y_j\frac{\hat{G}+1}{2} + \frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{M}\sum_{i=1}^{M} y_i I\{y_i \geq y_j\}\right) - \frac{1}{M}\sum_{j=1}^{M}\frac{\hat{\mu}}{2}(\hat{G}+1)\end{array}\right]. (B3)$$

Now

- $\hat{G} = \frac{2}{\hat{\mu}}\frac{1}{M}\sum_{j=1}^{M} y_j\,\hat{F}(y_j) - 1$, so that

  $$\frac{1}{M}\sum_{j=1}^{M} y_j\,\hat{F}(y_j) = \frac{(\hat{G}+1)}{2}\hat{\mu}\,;$$

- $\frac{1}{M}\sum_{j=1}^{M} y_j\frac{\hat{G}+1}{2} = \frac{\hat{G}+1}{2}\hat{\mu}$;
- $\frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{M}\sum_{i=1}^{M} y_i I\{y_i \geq y_j\}\right) = \frac{1}{M}\sum_{i=1}^{M} y_i\left(\frac{1}{M}\sum_{j=1}^{M} I\{y_j \leq y_i\}\right)$

  $$= \frac{1}{M}\sum_{i=1}^{M} y_i\,\hat{F}(y_i) = \frac{(\hat{G}+1)}{2}\hat{\mu}\,;$$

- $\frac{1}{M}\sum_{j=1}^{M}\frac{\hat{\mu}}{2}(\hat{G}+1) = \frac{\hat{\mu}}{2}(\hat{G}+1)$.

Plugging these expressions into equation (B3), we obtain

$$\bar{u}^* = \frac{2}{\hat{\mu}}\left[\frac{(\hat{G}+1)}{2}\hat{\mu} - \frac{\hat{G}+1}{2}\hat{\mu} + \frac{(\hat{G}+1)}{2}\hat{\mu} - \frac{\hat{\mu}}{2}(\hat{G}+1)\right]$$

$$= \frac{2}{\hat{\mu}} \times 0 = 0.$$

Now, with complex survey sample data, we have:

$$\bar{u}^* = \sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci} u_{hci}^*$$

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

$$= \frac{2}{\hat{\mu}}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\, y_{hci}\widehat{F}(y_{hci}) - \frac{2}{\hat{\mu}}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\, y_{hci}\frac{(\hat{G}+1)}{2}$$

$$+ \frac{2}{\hat{\mu}}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\left(\sum_{r=1}^{H}\sum_{s=1}^{n_r}\sum_{t=1}^{m} w_{rst}y_{rst}I\{y_{rst} \geq y_{hci}\}\right)$$

$$- \frac{2}{\hat{\mu}}\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\frac{\hat{\mu}}{2}(\hat{G}+1)$$

$$= (\hat{G}+1) - \frac{2}{\hat{\mu}} \times \hat{\mu} \times \frac{(\hat{G}+1)}{2}$$

$$+ \frac{2}{\hat{\mu}}\left(\sum_{r=1}^{H}\sum_{s=1}^{n_r}\sum_{t=1}^{m} w_{rst}y_{rst}\right)\sum_{h=1}^{H}\sum_{c=1}^{n_h}\sum_{i=1}^{m} w_{hci}\, I\{y_{hic} \leq y_{rst}\} - \frac{2}{\hat{\mu}}$$

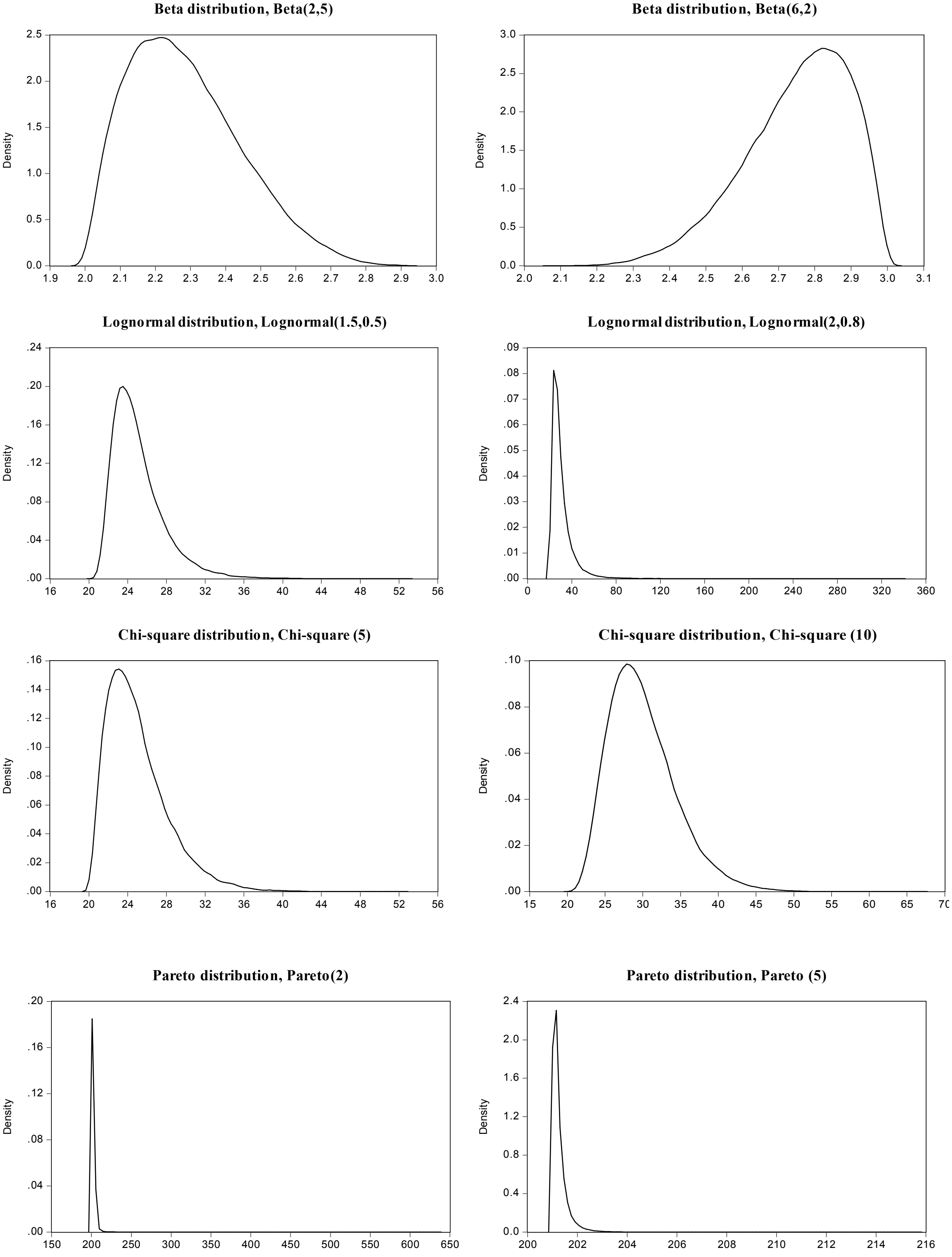
$$\times 1 \times \frac{\hat{\mu}}{2}(\hat{G}+1)$$

$$= \frac{2}{\hat{\mu}}\sum_{r=1}^{H}\sum_{s=1}^{n_r}\sum_{t=1}^{m} w_{rst}y_{rst}\,\widehat{F}(y_{rst}) - (\hat{G}+1)$$

$$= (\hat{G}+1) - (\hat{G}+1) = 0.$$

Appendix C

Figure C.1: Common shapes of beta, chi-square, lognormal and the Pareto distributions.



Beta distribution, Beta(2,5)

Beta distribution, Beta(6,2)

Lognormal distribution, Lognormal(1.5,0.5)

Lognormal distribution, Lognormal(2,0.8)

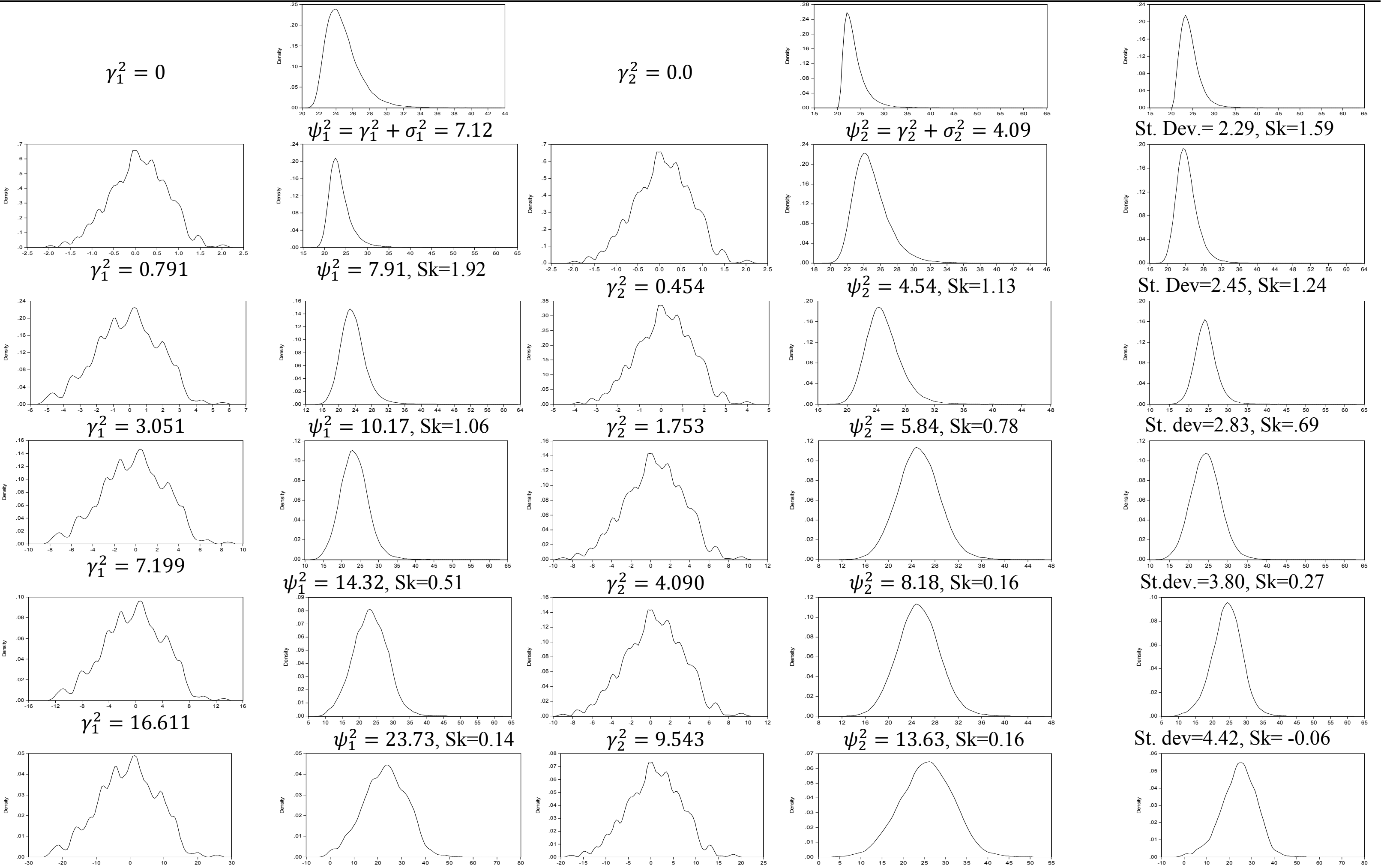
Chi-square distribution, Chi-square (5)

Chi-square distribution, Chi-square (10)

Pareto distribution, Pareto(2)

Pareto distribution, Pareto (5)

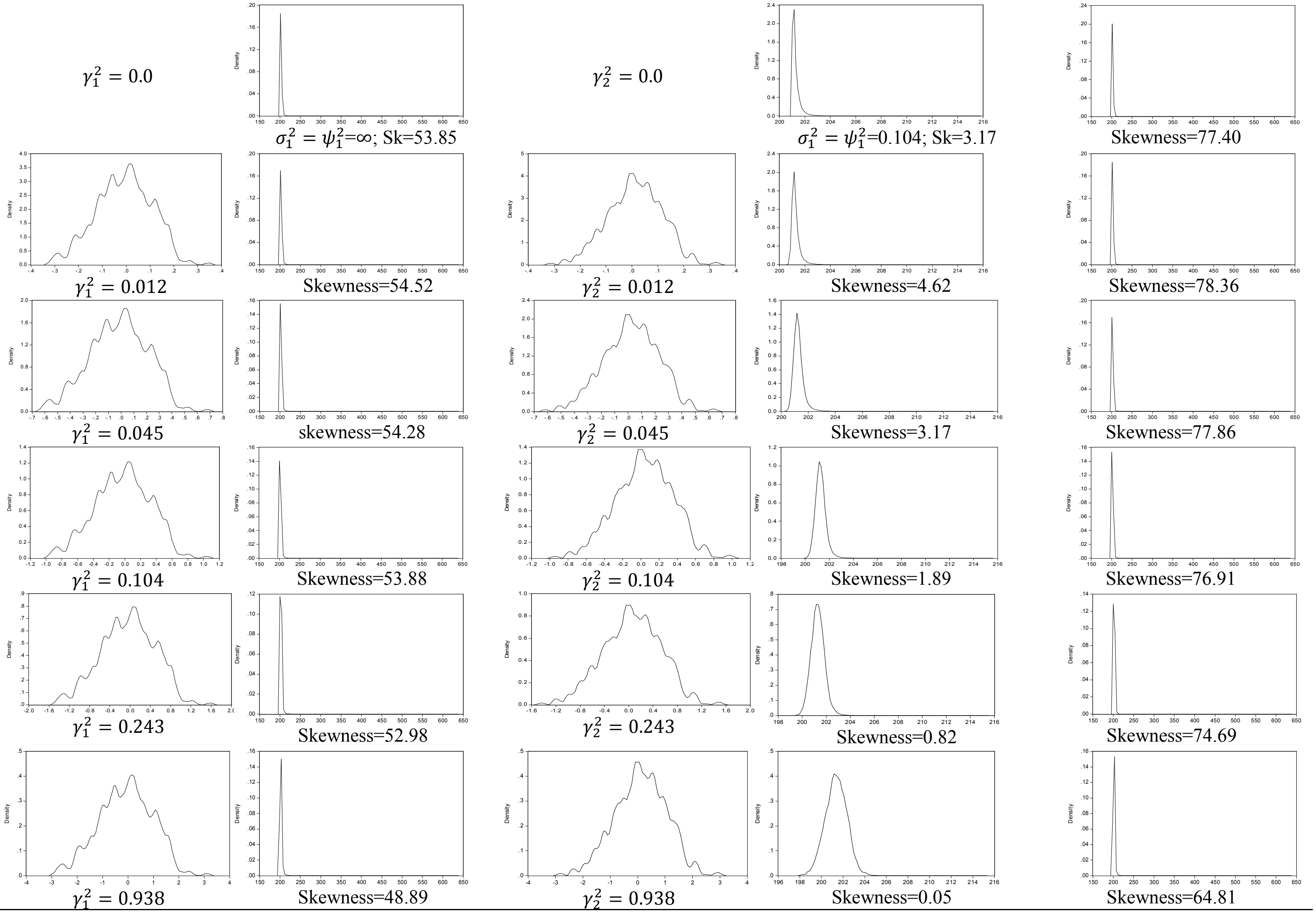
Figure C.2: Changes in distributions of strata with increasing ICC values: lognormal distribution for z_{hci}.



	Stratum 1		Stratum 2		Population
DGP	$\lambda_{1c} \sim N(0,\gamma_1^2))$	$y_{1ci} \sim (\mu_0,\psi_1^2)$	$\lambda_{1c} \sim N(0,\gamma_2^2)$	$y_{2ci} \sim (\mu_0,\psi_2^2)$	
DGP2A $\rho_1 = \rho_2$ $= 0.0$	$\gamma_1^2 = 0$	$\psi_1^2 = \gamma_1^2 + \sigma_1^2 = 7.12$	$\gamma_2^2 = 0.0$	$\psi_2^2 = \gamma_2^2 + \sigma_2^2 = 4.09$	St. Dev.= 2.29, Sk=1.59
DGP2B $\rho_1 = \rho_2$ $= 0.1$	$\gamma_1^2 = 0.791$	$\psi_1^2 = 7.91$, Sk=1.92	$\gamma_2^2 = 0.454$	$\psi_2^2 = 4.54$, Sk=1.13	St. Dev=2.45, Sk=1.24
DGP2C $\rho_1 = \rho_2$ $= 0.3$	$\gamma_1^2 = 3.051$	$\psi_1^2 = 10.17$, Sk=1.06	$\gamma_2^2 = 1.753$	$\psi_2^2 = 5.84$, Sk=0.78	St. dev=2.83, Sk=.69
DGP2D $\rho_1 = \rho_2$ $= 0.5$	$\gamma_1^2 = 7.199$	$\psi_1^2 = 14.32$, Sk=0.51	$\gamma_2^2 = 4.090$	$\psi_2^2 = 8.18$, Sk=0.16	St.dev.=3.80, Sk=0.27
DGP2E $\rho_1 = \rho_2$ $= 0.7$	$\gamma_1^2 = 16.611$	$\psi_1^2 = 23.73$, Sk=0.14	$\gamma_2^2 = 9.543$	$\psi_2^2 = 13.63$, Sk=0.16	St. dev=4.42, Sk= -0.06
DGP2F $\rho_1 = \rho_2$ $= 0.9$	$\gamma_1^2 = 64.071$	$\psi_1^2 = 71.19$, Sk=-0.09	$\gamma_2^2 = 36.810$	$\psi_2^2 = 40.90$, Sk=-0.06	St.dev=7.71, Sk=-0.26

Notes: For $z_{1ci} \sim LN(1.119, 0.602^2)$, $V(z_{1ci}) = \sigma_1^2 = 7.12$ and $z_{2ci} \sim LN(1.5, 0.4^2)$, and $Var(z_{2ci}) = \sigma_2^2 = 4.090$. Variance in stratum h is $\psi_h^2 = \sigma_h^2 + \gamma_h^2$.

Figure C.3: Changes in distributions of strata and population for increasing ICC values: the Pareto distribution for z_{hci}



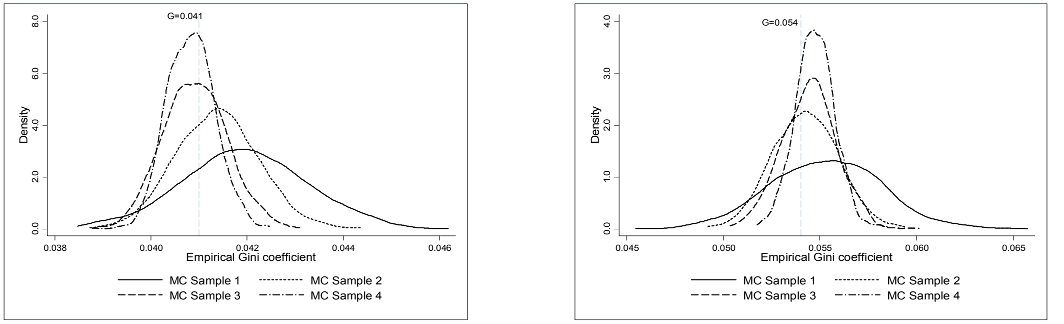
	Stratum 1		Stratum 2		Population
DGP4A $\rho_1 = \rho_2 = 0.0$	$\gamma_1^2 = 0.0$	$\sigma_1^2 = \psi_1^2$=∞; Sk=53.85	$\gamma_2^2 = 0.0$	$\sigma_1^2 = \psi_1^2$=0.104; Sk=3.17	Skewness=77.40
DGP2B $\rho_1 = \rho_2 = 0.1$	$\gamma_1^2 = 0.012$	Skewness=54.52	$\gamma_2^2 = 0.012$	Skewness=4.62	Skewness=78.36
DGP4C $\rho_1 = \rho_2 = 0.3$	$\gamma_1^2 = 0.045$	skewness=54.28	$\gamma_2^2 = 0.045$	Skewness=3.17	Skewness=77.86
DGP4D $\rho_1 = \rho_2 = 0.5$	$\gamma_1^2 = 0.104$	Skewness=53.88	$\gamma_2^2 = 0.104$	Skewness=1.89	Skewness=76.91
DGP4E $\rho_1 = \rho_2 = 0.7$	$\gamma_1^2 = 0.243$	Skewness=52.98	$\gamma_2^2 = 0.243$	Skewness=0.82	Skewness=74.69
DGP4F $\rho_1 = \rho_2 = 0.9$	$\gamma_1^2 = 0.938$	Skewness=48.89	$\gamma_2^2 = 0.938$	Skewness=0.05	Skewness=64.81

Notes: when $z_{1ci} \sim Pareto(2)$, $V(z_{1ci}) = \sigma_1^2 = \infty$ and $z_{2ci} \sim Pareto(5)$, and $Var(z_{2ci}) = \sigma_2^2 = 0.104$. Variance in stratum h is $\psi_h^2 = \sigma_h^2 + \gamma_h^2$.

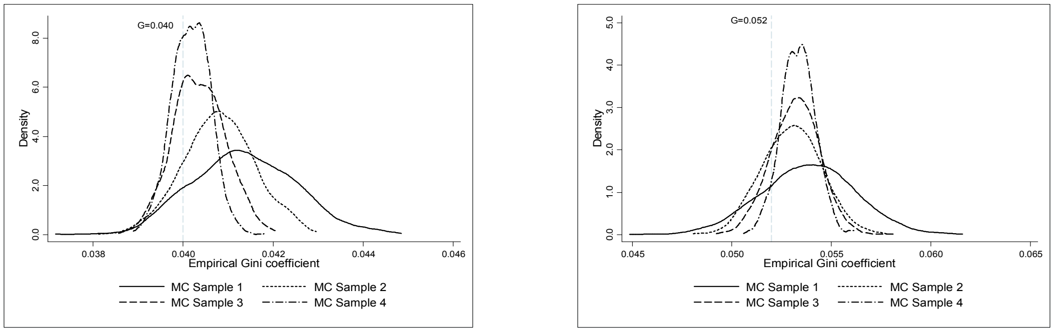
Figure C.4 MC empirical distributions of Gini coefficients for the increasing number of clusters: beta and lognormal distribution for z_{hci}.

Beta Distribution **Lognormal Distribution**

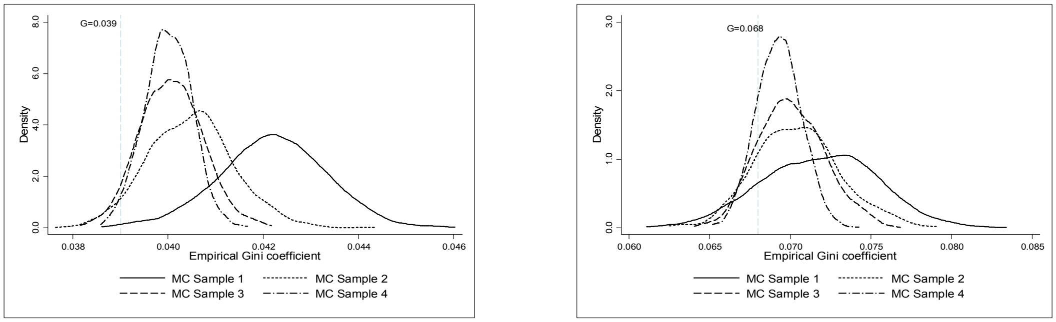
a) DGP1B: Population Gini=0.041 a) DGP2B: Population Gini =0.054



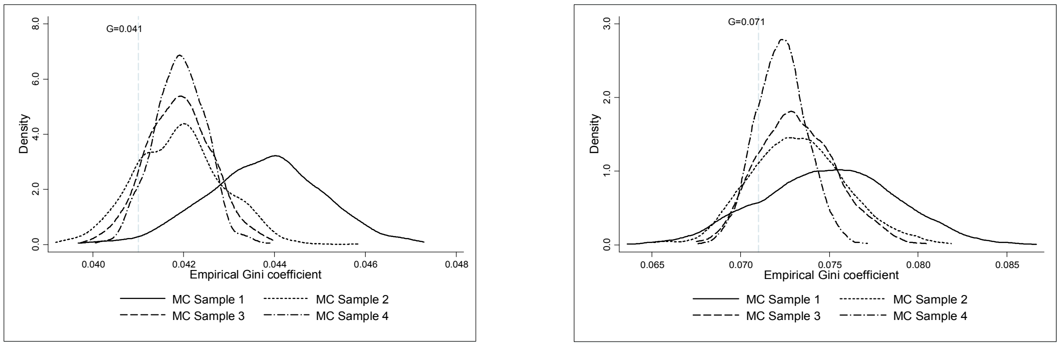
b) DGP1G: Population Gini=0.040 b) DGP2G: Population Gini=0.052



c) DGP1H: Population Gini =0.039 c) DGP2H: Population Gini=0.068



d) DGP1I: Population Gini =0.041 d) DGP2I: Population Gini=0.071

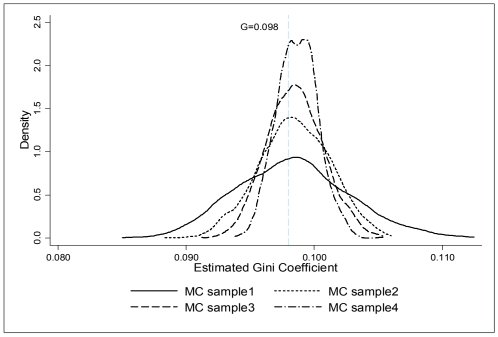


Notes: In first and second rows MC samples are from DGPs with two strata, and MC samples are MCS1: $n_1 = 10$, n_2 =15; MCS2: n_1 =25, n_2=25; MCS3: n_1 =40, , n_2=50; and MCS4: n_1=75, n_2=90. In third and fourth rows MC samples are from DGPs with five strata, and MC samples are MCS1: $n_1 =$ 3, n_2 =5, n_3=6, n_4=8, n_5=9; MCS2: $n_1 = 7$, n_2=9, n_3=12, n_4=15, n_5=20; MCS3: n_1 =12, n_2=20, n_3=24, n_4=32, n_5=36; and MCS4: n_1=25, n_2=35, n_3=45, n_4=64, n_5=70.

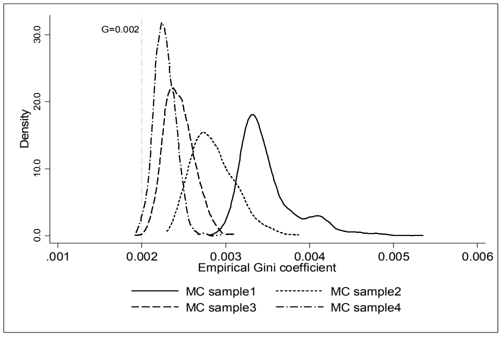
Figure C.5 MC empirical distributions of Gini coefficients for the increasing number of clusters: chi-square and the Pareto distribution for z_{hci}.

Chi-square Distribution	The Pareto Distribution

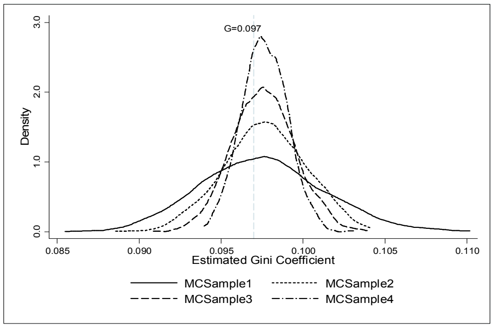
a) DGP3B: Population Gini=0.098



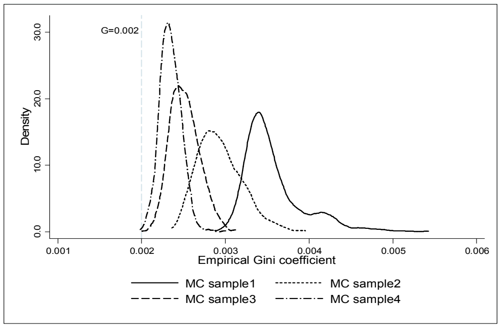
a) DGP4B: Population Gini =0.002



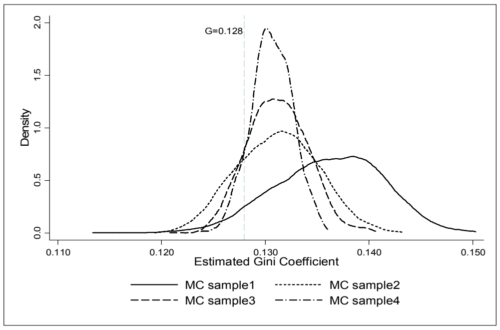
b) DGP3G: Population Gini=0.097



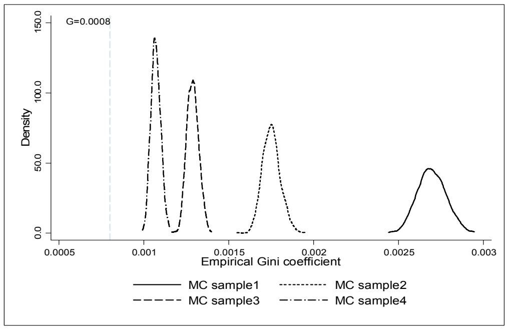
b) DGP4G: Population Gini=0.002



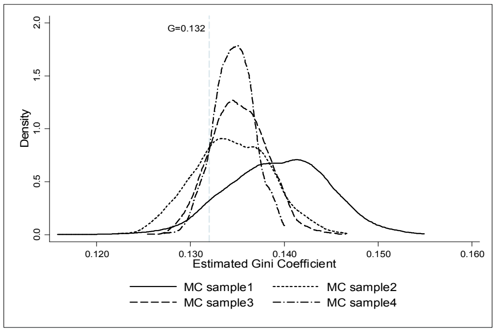
c) DGP3H: Population Gini =0.128



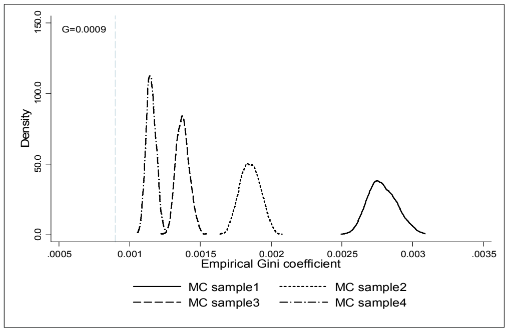
c) DGP4H: Population Gini=0.0008



d) DGP3I: Population Gini =0.132



d) DGP4I: Population Gini=0.0009



Notes: In first and second rows MC samples are from DGPs with two strata, and MC samples are MCS1: $n_1 = 10$, n_2 =15; MCS2: n_1 =25, n_2=25; MCS3: n_1 =40, , n_2=50; and MCS4: n_1=75, n_2=90. In third and fourth rows MC samples are from DGPs with five strata, and MC samples are MCS1: $n_1 =$ 3, n_2 =5, n_3=6, n_4=8, n_5=9; MCS2: $n_1 = 7$, n_2=9, n_3=12, n_4=15, n_5=20; MCS3: n_1 =12, n_2=20, n_3=24, n_4=32, n_5=36; and MCS4: n_1=25, n_2=35, n_3=45, n_4=64, n_5=70.

Table C.1: Empirical coverage probability (CP%), lower (L%) and upper (U%) error rates of nominal 95% CIs for G in proportional increase in the number of clusters: beta distribution for z_{hci}.

DGP	No of sampled clusters	BTAM			BKAM			SBSM			SBPM			WSPM		
		CP	L	U	CP	L	U	CP	L	U	CP	L	U	CP	L	U
DGP1B	n_1=10, n_2=15	75.30	24.60	0.10	78.20	21.70	0.10	78.10	21.70	0.20	92.70	2.10	5.20	95.70	2.00	2.30
	n_1=25, n_2=25	84.50	15.20	0.30	85.10	14.60	0.30	85.70	14.00	0.30	93.30	3.70	3.00	94.90	2.70	2.40
	n_1=40, n_2=50	88.60	11.10	0.30	89.10	10.60	0.30	88.20	11.50	0.30	94.00	3.90	2.10	94.70	4.50	0.80
	n_1=75, n_2=90	87.10	12.50	0.40	87.40	12.20	0.40	86.70	13.20	0.20	92.20	6.80	1.00	91.80	7.70	0.80
DGP1G	n_1=10, n_2=15	67.70	32.10	0.20	71.00	28.80	0.20	72.60	27.20	0.20	94.30	3.50	2.20	95.90	2.70	1.40
	n_1=25, n_2=25	75.80	24.00	0.20	77.50	22.30	0.20	76.30	23.60	0.10	90.40	7.20	2.40	93.20	5.90	0.90
	n_1=40, n_2=50	85.10	14.70	0.20	85.70	14.10	0.20	85.40	14.20	0.40	92.80	4.80	2.40	90.80	7.40	1.80
	n_1=75, n_2=90	89.80	9.80	0.40	90.00	9.60	0.40	89.10	10.70	0.20	93.80	4.10	2.10	94.40	4.90	0.70
DGP1H	n_1=3, n_2=5, n_3=6, n_4=8, n_5=9	22.40	77.60	0.00	27.50	72.50	0.00	29.10	70.90	0.00	71.10	28.80	0.10	73.00	27.00	0.00
	n_1=7, n_2=9, n_3=12, n_4=15, n_5=20	76.20	23.50	0.30	78.90	21.00	0.10	76.90	22.80	0.30	90.30	7.10	2.60	91.00	8.20	0.80
	n_1=12, n_2=20, n_3=24, n_4=32, n_5=36	84.70	14.80	0.50	85.70	13.90	0.40	85.40	14.00	0.60	92.70	5.20	2.10	93.20	5.60	1.20
	n_1=25, n_2=35, n_3=45, n_4=64, n_5=70	79.90	20.00	0.10	81.00	18.90	0.10	80.40	12.20	0.10	87.20	12.20	0.60	87.00	12.60	0.40
DGP1I	n_1=3, n_2=5, n_3=6, n_4=8, n_5=9	45.10	54.90	0.00	51.50	48.50	0.00	54.20	45.80	0.00	84.20	15.00	0.80	84.80	15.00	0.20
	n_1=7, n_2=9, n_3=12, n_4=15, n_5=20	89.30	8.80	1.90	90.30	7.90	1.80	89.50	8.80	1.70	90.10	2.20	7.70	94.20	1.50	4.30
	n_1=12, n_2=20, n_3=24, n_4=32, n_5=36	90.10	9.00	0.90	90.80	8.30	0.90	90.80	8.50	0.70	93.40	3.80	2.80	94.60	3.40	2.00
	n_1=25, n_2=35, n_3=45, n_4=64, n_5=70	85.90	13.80	0.30	86.80	12.90	0.30	86.40	13.30	0.30	90.70	7.90	1.50	91.40	7.90	0.70

Notes: BTAM: SN approximation CI using Bhattacharya's (2007) standard error, BKAM: SN approximation CI using Binder and Kovačević's (1995) standard error, SBSM: SN approximation CI using bootstrap standard error, SBPM: standard MC bootstrap percentile CI and WSPM: warp-speed MC percentile CI.

Table C.2: Empirical coverage probability (CP%), lower (L%) and upper (U%) error rates of nominal 95% CIs for G in proportional increase in the number of clusters: lognormal distribution for z_{hci}.

DGP	No of sampled clusters	BTAM			BKAM			SBSM			SBPM			WSPM		
		CP	L	U	CP	L	U	CP	L	U	CP	L	U	CP	L	U
DGP2B	n_1=10, n_2=15	88.70	7.60	3.70	89.70	6.90	3.40	89.70	9.70	3.60	89.00	3.60	7.40	94.00	3.10	2.90
	n_1=25, n_2=25	93.00	3.10	3.90	93.40	3.00	3.60	93.10	3.20	3.70	88.60	2.00	9.40	90.60	1.50	7.90
	n_1=40, n_2=50	91.70	4.80	3.50	91.90	4.70	3.40	92.80	3.90	3.30	91.50	3.00	5.50	92.80	4.20	3.00
	n_1=75, n_2=90	90.90	8.20	0.90	91.00	8.10	0.90	90.90	8.00	1.00	91.40	7.30	1.30	92.60	6.50	0.90
DGP2G	n_1=10, n_2=15	89.30	8.10	2.60	90.10	7.70	2.60	90.40	7.00	2.60	89.60	2.80	7.60	94.40	1.40	3.90
	n_1=25, n_2=25	93.40	3.60	3.00	93.90	3.30	2.80	93.10	4.00	2.90	89.60	1.90	8.50	90.90	1.70	7.40
	n_1=40, n_2=50	92.20	5.40	2.40	92.80	5.10	2.10	93.00	5.00	2.00	91.60	2.90	5.50	93.40	3.60	3.00
	n_1=75, n_2=90	91.90	7.40	0.70	92.10	7.20	0.70	91.30	7.80	0.90	93.20	5.50	1.30	93.40	5.70	0.90
DGP2H	n_1=3, n_2=5, n_3=6, n_4=8, n_5=9	67.80	31.30	0.90	73.30	26.10	0.60	73.10	26.30	0.60	80.20	17.90	1.90	85.10	14.10	0.80
	n_1=7, n_2=9, n_3=12, n_4=15, n_5=20	81.80	17.50	0.70	83.20	16.10	0.70	82.80	16.50	0.70	87.50	10.90	1.60	89.70	9.40	0.90
	n_1=12, n_2=20, n_3=24, n_4=32, n_5=36	81.40	18.10	0.50	83.10	16.50	0.40	82.30	17.40	0.30	85.70	13.40	0.90	83.90	15.90	0.20
	n_1=25, n_2=35, n_3=45, n_4=64, n_5=70	85.30	14.10	0.60	85.90	13.60	0.50	86.00	13.50	0.50	88.20	11.00	0.80	84.00	15.50	0.50
DGP2I	n_1=3, n_2=5, n_3=6, n_4=8, n_5=9	68.40	31.00	0.60	72.50	26.90	0.60	73.40	26.00	0.60	80.60	18.30	1.10	83.10	16.50	0.40
	n_1=7, n_2=9, n_3=12, n_4=15, n_5=20	79.80	19.30	0.90	82.00	17.10	0.90	81.40	17.70	0.90	86.80	18.30	1.10	89.50	9.60	0.90
	n_1=12, n_2=20, n_3=24, n_4=32, n_5=36	78.70	21.10	0.20	79.50	20.30	0.20	79.60	20.20	0.20	82.30	17.10	0.60	81.00	19.00	0.00
	n_1=25, n_2=35, n_3=45, n_4=64, n_5=70	83.20	16.50	0.30	83.50	16.20	0.30	83.50	16.20	0.30	87.30	12.20	0.50	81.60	18.20	0.20

Notes: BTAM: SN approximation CI using Bhattacharya's (2007) standard error, BKAM: SN approximation CI using Binder and Kovačević's (1995) standard error, SBSM: SN approximation CI using bootstrap standard error, SBPM: standard MC bootstrap percentile CI and WSPM: warp-speed MC percentile CI.

Table C.3: Empirical coverage probability (CP%), lower (L%) and upper (U%) error rates of nominal 95% CIs for G in proportional increase in the number of clusters: Chi-square distribution for z_{hci}.

DGP	No of sampled clusters	BTAM			BKAM			SBSM			SBPM			WSPM		
		CP	L	U	CP	L	U	CP	L	U	CP	L	U	CP	L	U
DGP3B	n_1=10, n_2=15	90.50	5.70	3.80	91.50	5.10	3.40	91.00	5.40	3.60	87.00	3.20	8.90	93.20	3.20	4.20
	n_1=25, n_2=25	92.90	4.60	2.50	93.20	4.50	2.30	92.40	5.10	2.50	92.90	3.50	3.60	94.60	3.10	2.30
	n_1=40, n_2=50	91.20	6.90	1.90	91.60	6.70	1.70	91.30	7.10	1.60	91.00	5.90	3.10	91.70	6.60	1.70
	n_1=75, n_2=90	91.40	8.10	0.50	91.50	8.00	0.50	91.80	7.70	0.50	92.00	7.20	0.80	92.10	7.40	0.50
DGP3G	n_1=10, n_2=15	91.30	5.80	2.90	92.20	5.20	2.60	91.50	5.90	2.60	89.90	2.90	7.20	92.70	2.30	5.00
	n_1=25, n_2=25	91.60	6.40	2.00	91.90	6.10	2.00	91.10	6.60	2.30	91.00	5.10	3.90	94.80	3.20	2.00
	n_1=40, n_2=50	90.40	8.20	1.40	90.60	8.00	1.40	90.70	7.90	1.40	91.30	5.90	2.80	91.30	7.30	1.40
	n_1=75, n_2=90	90.30	9.30	0.40	90.50	9.10	0.40	90.00	9.70	0.30	91.30	7.70	1.00	91.80	7.80	0.40
DGP3H	n_1=3, n_2=5, n_3=6, n_4=8, n_5=9	40.30	59.30	0.40	46.00	53.70	0.30	43.30	56.30	0.40	29.60	70.20	0.20	29.30	70.60	0.10
	n_1=7, n_2=9, n_3=12, n_4=15, n_5=20	78.80	20.90	0.30	81.20	18.50	0.30	79.40	20.30	0.30	72.90	26.70	0.40	77.00	22.90	0.10
	n_1=12, n_2=20, n_3=24, n_4=32, n_5=36	74.80	25.10	0.10	76.30	23.60	0.10	75.30	24.60	0.10	70.30	29.60	0.10	75.50	24.40	0.10
	n_1=25, n_2=35, n_3=45, n_4=64, n_5=70	69.90	30.00	0.10	70.60	29.30	0.10	70.40	29.50	0.10	68.20	31.80	0.00	67.30	32.50	0.20
DGP3I	n_1=3, n_2=5, n_3=6, n_4=8, n_5=9	48.30	51.10	0.60	53.80	45.70	0.50	52.80	46.70	0.50	35.50	64.30	0.20	38.40	64.10	0.20
	n_1=7, n_2=9, n_3=12, n_4=15, n_5=20	82.30	16.80	0.90	84.70	14.60	0.70	83.40	15.70	0.90	77.10	22.20	0.70	83.60	16.00	0.40
	n_1=12, n_2=20, n_3=24, n_4=32, n_5=36	77.20	22.70	0.10	78.40	21.50	0.10	77.70	22.10	0.20	72.30	27.60	0.10	79.30	20.60	0.10
	n_1=25, n_2=35, n_3=45, n_4=64, n_5=70	74.20	25.60	0.20	74.80	25.10	0.10	74.40	25.50	0.10	70.90	29.00	0.10	66.10	33.70	0.20

Notes: BTAM: SN approximation CI using Bhattacharya's (2007) standard error, BKAM: SN approximation CI using Binder and Kovačević's (1995) standard error, SBSM: SN approximation CI using bootstrap standard error, SBPM: standard MC bootstrap percentile CI and WSPM: warp-speed MC percentile CI.

Table C.4: Empirical coverage probability (CP%), lower (L%) and upper (U%) error rates of nominal 95% CIs for G in proportional increase in the number of clusters: Pareto distribution for z_{hci}.

DGP	No of sampled clusters	BTAM			BKAM			SBSM			SBPM			WSPM		
		CP	L	U	CP	L	U	CP	L	U	CP	L	U	CP	L	U
DGP4B	n_1=10, n_2=15	0.00	1.00	0.00	0.00	1.00	0.00	1.20	98.80	0.00	100	0.00	0.00	99.10	0.90	0.00
	n_1=25, n_2=25	0.20	99.80	0.00	0.30	99.70	0.00	2.90	97.10	0.00	96.80	1.20	2.00	97.00	3.00	0.00
	n_1=40, n_2=50	46.50	53.50	0.00	47.70	52.30	0.00	54.50	45.60	0.00	93.20	0.10	6.70	96.70	0.80	2.50
	n_1=75, n_2=90	48.50	51.50	0.00	49.50	50.50	0.00	57.40	42.60	0.00	94.00	1.40	4.60	95.90	2.40	1.70
DGP4G	n_1=10, n_2=15	0.00	100	0.00	0.00	100	0.00	1.40	98.60	0.00	100	0.00	0.00	99.20	0.80	0.00
	n_1=25, n_2=25	0.20	99.80	0.00	0.30	99.70	0.00	2.90	97.10	0.00	97.10	1.10	1.80	96.80	3.20	0.00
	n_1=40, n_2=50	22.60	77.40	0.00	23.40	76.60	0.00	31.10	68.90	0.00	92.60	0.10	7.30	97.80	1.90	0.30
	n_1=75, n_2=90	51.60	48.40	0.00	52.50	47.50	0.00	53.70	46.30	0.00	93.040	1.30	5.30	96.40	1.90	1.70
DGP4H	n_1=3, n_2=5, n_3=6, n_4=8, n_5=9	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00	100	0.00	0.00	100	0.00	0.00
	n_1=7, n_2=9, n_3=12, n_4=15, n_5=20	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00	100	0.00	0.00	100	0.00	0.00
	n_1=12, n_2=20, n_3=24, n_4=32, n_5=36	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00	99.90	0.00	0.10	100	0.00	0.00
	n_1=25, n_2=35, n_3=45, n_4=64, n_5=70	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00	98.70	0.80	0.50	99.40	0.50	0.10
DGP4I	n_1=3, n_2=5, n_3=6, n_4=8,	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00	100	0.00	0.00	100	0.00	0.00
	n_1=7, n_2=9, n_3=12, n_4=15, n_5=20	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00	99.90	0.10	0.00	100	0.00	0.00
	n_1=12, n_2=20, n_3=24, n_4=32, n_5=36	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00	99.90	0.10	0.00	100	0.00	0.00
	n_1=25, n_2=35, n_3=45, n_4=64, n_5=70	0.00	100	0.00	0.00	100	0.00	0.00	100	0.00	95.90	3.40	0.70	97.90	1.80	0.30

Notes: BTAM: SN approximation CI using Bhattacharya's (2007) standard error, BKAM: SN approximation CI using Binder and Kovačević's (1995) standard error, SBSM: SN approximation CI using bootstrap standard error, SBPM: standard MC bootstrap percentile CI and WSPM: warp-speed MC percentile CI.

Appendix D

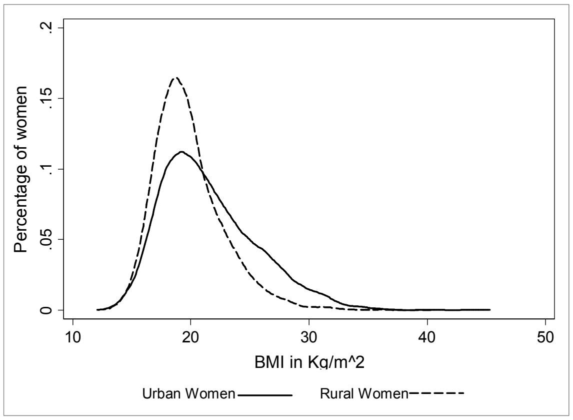
Table D.1 Additional summary statistics and hypothesis tests for BMI: all women and by place of residence (rural; urban) for BDHS 2004, 2007 and 2011 surveys.

Category		Survey Year		
		2004	2007	2011
All women	Number of Observations	11,166	10,836	17,309
	Mean BMI	20.258	20.733	21.449
	Standard Error of Mean	0.059	0.068	0.057
	Standard Deviation of BMI	3.315	3.519	3.797
Urban	Number of Observations	3,816	4,118	6,024
	Mean BMI	21.586	22.290	23.028
	Standard Error of Mean	0.143	0.165	0.121
	Standard Deviation of BMI	4.997	5.411	4.860
Rural	Number of Observations	7,350	6,718	11,285
	Mean BMI	19.871	20.273	20.900
	Standard Error of Mean	0.064	0.073	0.063
	Standard Deviation of BMI	2.723	2.823	3.269
Hypothesis Tests				
$H_0: \mu_{2004} = \mu_{2007} = \mu_{2011}$				
	Wald statistic (χ^2)	213.950		
	p-value	0.000		
$H_0: \mu_{2004}^{\text{urban}} = \mu_{2007}^{\text{urban}} = \mu_{2011}^{\text{urban}}$				
	Wald statistic (χ^2)	59.778		
	p-value	0.000		
$H_0: \mu_{2004}^{\text{rural}} = \mu_{2007}^{\text{rural}} = \mu_{2011}^{\text{rural}}$				
	Wald statistic (χ^2)	133.269		
	p-value	0.000		
$H_0: \mu_{2004}^{\text{urban}} = \mu_{2004}^{\text{rural}}$				
	Wald statistic $\chi^2(1)$	119.36		
	p-value	0.000		
$H_0: \mu_{2007}^{\text{urban}} = \mu_{2007}^{\text{rural}}$				
	Wald statistic $\chi^2(1)$	124.47		
	p-value	0.000		
$H_0: \mu_{2011}^{\text{urban}} = \mu_{2011}^{\text{rural}}$				
	Wald statistic $\chi^2(1)$	244.24		
	p-value	0.000		

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 for the mean.

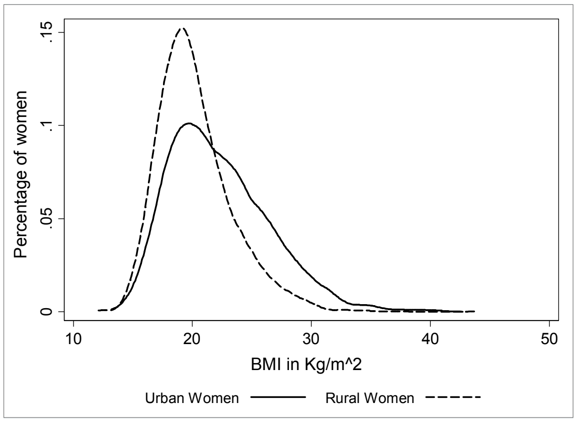
Figure D.1: BMI distributions for Bangladeshi urban and rural women: a) BDHS 2004; b) BDHS 2007; and c) BDHS 2011.

a)



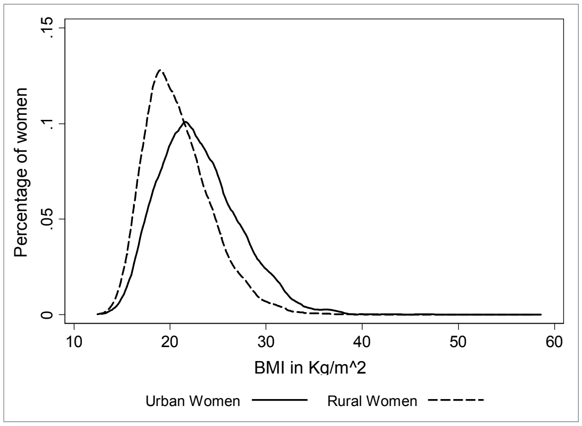
Notes: Normality test statistics for urban women sample: JB 650.914 (0.000), skewness=0.878 and kurtosis=4.000; for rural women sample: JB=2739 (p-value=0.000), skewness=1.062 and kurtosis=5.104

b)



Notes: Normality test results for urban sample: JB=570.930 (p-value=0.000), skewness=0.817, kurtosis=3.822; for rural sample: JB=2133.700 (p-value=0.000), skewness=1.008, kurtosis=4.979

c)



Notes: Normality test results for urban sample: JB=1241.343 (p-value=0.000), skewness=0.776, kurtosis=4.594; for rural sample: JB=4279.950 (p-value=0.000), skewness=0.970, kurtosis=5.310

Table D.2 Summary statistics and hypothesis tests for BMI: by wealth category and BDHS survey years 2004, 2007, and 2011.

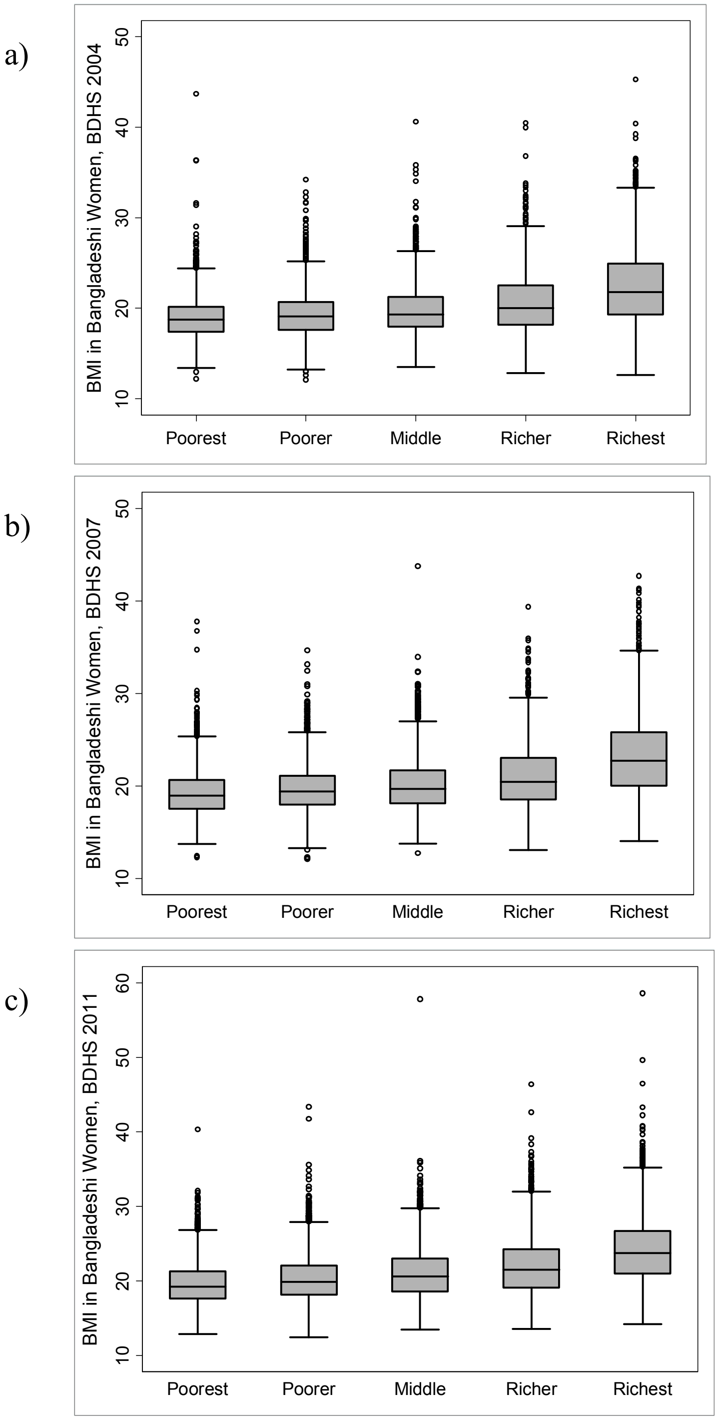
Category		Survey Year		
		2004	2007	2011
Poorest	Number of Observations	1,979	1,751	3,013
	Mean BMI	19.042	19.354	19.676
	Standard Error of Mean	0.061	0.086	0.072
	Standard Deviation of BMI	2.367	2.424	2.812
Poorer	Number of Observations	1,997	1,964	3,235
	Mean BMI	19.395	19.761	20.339
	Standard Error of Mean	0.068	0.083	0.064
	Standard Deviation of BMI	2.465	2.574	2.812
Middle	Number of Observations	2,092	2,070	3,328
	Mean BMI	19.885	20.235	21.022
	Standard Error of Mean	0.082	0.084	0.072
	Standard Deviation of BMI	2.764	2.974	3.252
Richer	Number of Observations	2,230	2,173	3,670
	Mean BMI	20.536	20.994	21.936
	Standard Error of Mean	0.099	0.095	0.075
	Standard Deviation of BMI	3.246	3.337	3.783
Richest	Number of Observations	2,868	2,878	4,063
	Mean BMI	22.379	23.150	23.959
	Standard Error of Mean	0.107	0.131	0.102
	Standard Deviation of BMI	4.509	4.739	4.436

Hypothesis Tests

$H_0: \mu_{2004}^{\text{poorest}} = \mu_{2004}^{\text{poorer}} = \mu_{2004}^{\text{middle}} = \mu_{2004}^{\text{richer}} = \mu_{2004}^{\text{richest}}$	
Wald statistic (χ^2)	826.601
p-value	0.000
$H_0: \mu_{2007}^{\text{poorest}} = \mu_{2007}^{\text{poorer}} = \mu_{2007}^{\text{middle}} = \mu_{2007}^{\text{richer}} = \mu_{2007}^{\text{richest}}$	
Wald statistic (χ^2)	689.159
p-value	0.000
$H_0: \mu_{2011}^{\text{poorest}} = \mu_{2011}^{\text{poorer}} = \mu_{2011}^{\text{middle}} = \mu_{2011}^{\text{richer}} = \mu_{2011}^{\text{richest}}$	
Wald statistic (χ^2)	1440.613
p-value	0.000

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.

Figure D.2: BMI distributions for Bangladeshi women by wealth category: a) BDHS 2004; b) BDHS 2007; and c) 2011



Notes: Outliers are present in all five wealth categories, making the distributions positively skewed. Although there are more obese (with a BMI≥30) in the richest category, such women are present in other category as well.

Table D.3 Summary statistics and hypothesis tests for BMI: by educational attainment and BDHS survey years 2004, 2007, and 2011

Category		Survey Year		
		2004	2007	2011
No education	Number of Observations	4,333	3,466	4,532
	Mean BMI	19.563	19.945	20.723
	Standard Error of Mean	0.055	0.071	0.076
	Standard Deviation of BMI	2.772	3.075	3.448
Primary incomplete	Number of Observations	2,269	2,262	3,133
	Mean BMI	20.051	20.511	20.928
	Standard Error of Mean	0.088	0.095	0.088
	Standard Deviation of BMI	3.215	3.247	3.668
Primary complete	Number of Observations	1,018	950	2,038
	Mean BMI	20.475	20.903	21.190
	Standard Error of Mean	0.112	0.139	0.103
	Standard Deviation of BMI	3.341	3.666	3.668
Secondary incomplete	Number of Observations	2,566	2,608	5,345
	Mean BMI	20.924	21.415	21.705
	Standard Error of Mean	0.092	0.104	0.075
	Standard Deviation of BMI	3.549	3.515	3.831
Secondary or Higher	Number of Observations	980	1,528	2,261
	Mean BMI	22.511	22.573	23.583
	Standard Error of Mean	0.186	0.153	0.122
	Standard Deviation of BMI	4.398	4.420	4.279

Hypothesis Tests

$H_0: \mu_{2004}^{\text{no.edu}} = \mu_{2004}^{\text{prim.inc}} = \mu_{2004}^{\text{prim.com}} = \mu_{2004}^{\text{secon.inc}} = \mu_{2004}^{\text{higher}}$

Wald statistic (χ^2)	354.842
p-value	0.000

$H_0: \mu_{2007}^{\text{no.edu}} = \mu_{2007}^{\text{prim.inc}} = \mu_{2007}^{\text{prim.com}} = \mu_{2007}^{\text{secon.inc}} = \mu_{2011}^{\text{higher}}$

Wald statistic (χ^2)	315.897
p-value	0.000

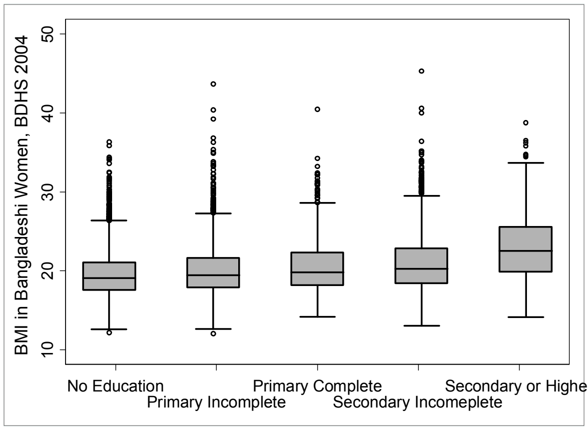
$H_0: \mu_{2011}^{\text{no.edu}} = \mu_{2011}^{\text{prim.inc}} = \mu_{2011}^{\text{prim.com}} = \mu_{2011}^{\text{secon.inc}} = \mu_{2011}^{\text{higher}}$

Wald statistic (χ^2)	448.976
p-value	0.000

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. no.edu, prim.inc, prim.com, secon.inc and higher indicate women with no education, primary incomplete education, primary complete education, secondary income education and secondary compete or higher education levels, respectively.

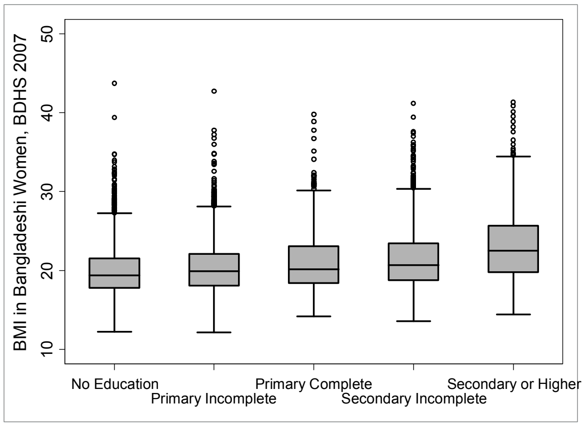
Figure D.3: BMI distributions for Bangladeshi women by educational attainment: a) BDHS 2004; b) BDHS 2007; and c) BDHS 2011

a)



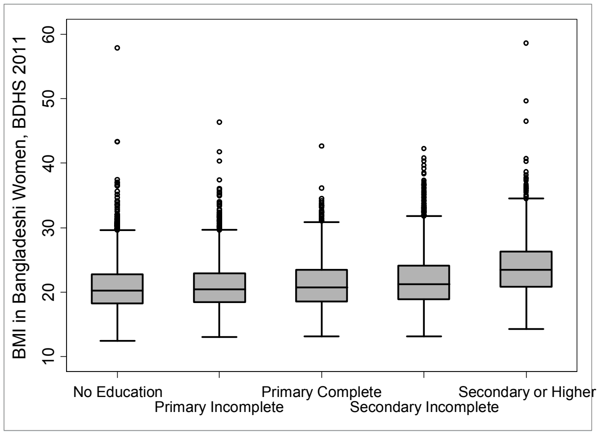
Notes: All five distributions are positively skewed. The "secondary complete" distribution has the largest variability (SD 4.398) followed by the "secondary incomplete" (SD 3.549) and the "no education" distribution has the least variability (SD 2.772).

b)



Notes: BMI distributions are positively skewed and the variability increases with a higher educational attainment.

c)



Notes: Although mean BMIs increase with education level, extreme overweight (BMI≥40) prevalence is similar across education categories.

Table D.4: Estimated intracluster correlation coefficients (ICC) and components for BMI data by stratum for BDHS 2004

Stratum h	Number of clusters in h	Average cluster size, m_h	Within cluster variance, $\hat{\sigma}_h^2$	Between cluster variance, $\hat{\gamma}_h^2$	Intracluster correlation coefficient, $\hat{\rho}_h$
1. Barisal Rural	31	31.63	8.426	0.262	0.030
2. Chittagong Rural	41	30.84	8.815	0.271	0.030
3. Dhaka Rural	48	29.81	8.327	0.846	0.092
4. Khulna Rural	36	30.43	8.229	0.252	0.030
5. Rajshahi Rural	57	30.96	7.226	0.319	0.042
6. Sylhet Rural	26	31.22	8.903	0.785	0.081
7. Barisal Urban: Town	8	27.71	19.658	0.797	0.039
8. Chittagong Urban: Town	15	33.04	11.632	1.991	0.146
9. Dhaka Urban: Town	25	30.63	14.619	2.093	0.125
10. Khulna Urban: Town	13	32.36	12.766	2.850	0.183
11. Rajshahi Urban: Town	15	32.51	11.429	0.802	0.066
12. Sylhet Urban: Town	6	34.81	18.396	0.933	0.048
13. Barisal Urban: Small city	4	29.98	11.672	0.486	0.040
14. Chittagong Urban: Small City	4	31.36	17.880	1.294	0.067
15. Dhaka Urban: Small City	3	28.91	10.272	0.275	0.026
16. Khulna Urban: Small City	3	32.79	16.909	2.622	0.134
17. Rajshahi Urban: Small City	6	31.31	18.403	1.367	0.069
18. Sylhet Urban: Small City	3	32.99	15.603	0.180	0.011
19. Chittagong Urban: City Corp.	4	33.19	20.479	0.342	0.016
20. Dhaka Urban: City Corp.	9	26.71	16.775	0.988	0.056
21. Khulna Urban: City Corp.	2	28.42	17.700	5.390	0.233
22. Rajshahi Urban: City Corp.	2	30.87	9.233	0.000*	0.000

Notes: *The between-cluster variance, $\hat{\gamma}_h^2$, is truncated at zero in STATA when the mean squares within clusters is larger than the mean squares between clusters, or $F_h = \frac{MSB_h}{MSW_h} < 1$.

Table D.5: Estimated intracluster correlation coefficients (ICC) and components for BMI data by stratum for BDHS 2007

Stratum h	Number of clusters in h	Average cluster size, m_h	Within cluster variance, $\hat{\sigma}_h^2$	Between cluster variance, $\hat{\gamma}_h^2$	Intracluster correlation coefficient, $\hat{\rho}_h$
1. Barisal Rural	33	27.80	8.155	0.407	0.048
2. Barisal Urban: Town	10	30.82	15.274	0.985	0.061
3. Barisal Urban: Small City	6	31.17	9.518	0.920	0.088
4. Chittagong Rural	38	30.46	9.583	0.527	0.052
5. Chittagong Urban: City Corp.	12	31.94	17.960	1.854	0.094
6. Chittagong Urban: Town	7	30.95	17.144	0.472	0.027
7. Chittagong Urban: Small City	5	31.50	12.013	2.647	0.181
8. Dhaka Rural	42	29.49	9.051	0.938	0.094
9. Dhaka Urban: City Corp.	26	28.67	16.985	2.189	0.114
10. Dhaka Urban: Town	7	32.26	10.483	4.693	0.309
11. Dhaka Urban: Small City	3	33.66	13.178	0.000*	0.000
12. Khulna Rural	36	29.43	9.534	0.195	0.020
13. Khulna Urban: City Corp.	7	30.41	15.709	1.268	0.075
14. Khulna Urban: Town	9	31.19	11.029	2.484	0.184
15. Khulna Urban: Small City	4	32.41	13.028	4.601	0.260
16. Rajshahi Rural	46	29.54	9.267	0.475	0.049
17. Rajshahi Urban: City Copr.	3	30.62	18.352	3.714	0.168
18. Rajshahi Urban: Town	14	31.66	17.419	2.395	0.121
19. Rajshahi Urban: Small City	6	27.10	13.261	2.904	0.180
20. Sylhet Rural	32	30.74	8.947	0.858	0.088
21. Sylhet Urban: Town	10	31.31	14.969	2.910	0.163
22. Sylhet Urban: Small City	5	30.35	17.359	2.924	0.144

Notes: *The between-cluster variance, $\hat{\gamma}_h^2$, is truncated at zero in STATA when the mean squares within clusters is larger than the mean squares between clusters, or $F_h = \frac{MSB_h}{MSW_h} < 1$.

Table D.6: Estimated intracluster correlation coefficients (ICC) and components for BMI data by stratum for BDHS 2011

Stratum h	Number of clusters in h	Average cluster size, m_h	Within cluster variance, $\hat{\sigma}_h^2$	Between cluster variance, $\hat{\gamma}_h^2$	Intracluster correlation coefficient, $\hat{\rho}_h$
1. Barisal City Corp	7	30.79	16.406	3.549	0.178
2. Chittagong City Corp.	16	30.03	18.810	2.501	0.117
3. Dhaka City Corp.	23	23.51	18.811	0.835	0.043
4. Khulna City Corp.	9	31.05	18.608	0.812	0.042
5. Rajshahi City Corp.	5	28.44	17.926	0.790	0.042
6. Sylhet City Corp.	10	27.66	15.929	3.158	0.165
7. Barisal Other Urban	15	28.16	14.955	1.220	0.075
8. Chittagong Other Urban	17	31.79	13.659	1.770	0.115
9. Dhaka Other Urban	26	28.52	13.207	1.461	0.100
10. Khulna Other Urban	20	31.46	15.996	1.860	0.104
11. Rajshahi Other Urban	23	30.27	14.673	1.338	0.084
12. Rangpur Other Urban	24	28.90	15.894	2.108	0.117
13. Sylhet Other Urban	12	29.81	17.632	2.142	0.108
14. Barisal Rural	50	26.09	10.778	0.411	0.037
15. Chittagong Rural	59	29.99	12.322	1.028	0.077
16. Dhaka Rural	60	28.21	10.066	1.041	0.094
17. Khulna Rural	56	30.16	12.538	0.285	0.022
18. Rajshahi Rural	59	28.59	11.759	0.902	0.071
19. Rangpur Rural	61	28.50	9.313	0.584	0.059
20. Sylhet Rural	48	29.19	11.968	1.340	0.101

Table D.7: Percentages of women in BMI clusters by wealth category: ever-married Bangladeshi women aged 15-49 samples from BDHS 2004, 2007 and 2011 samples

Survey year	Wealth category	Number of Strata	Number of clusters	Average size of a cluster	Minimum number of obs. in a cluster	Maximum number of obs. in a cluster	% of clusters with 1 obs.	% of clusters with 2 obs.	% of clusters with 3 obs.	% of clusters with ≥10 children
2004	Poorest	18	306	6.458	1	24	10.131	8.824	12.092	25.490
	Poorer	18	323	6.146	1	18	6.811	8.359	6.192	16.409
	Middle	21	338	6.186	1	20	6.509	4.142	9.172	18.639
	Richer	22	337	6.617	1	19	6.638	9.792	11.276	22.849
	Richest	22	289	9.924	1	36	13.149	8.651	9.689	37.716
2007	Poorest	20	260	6.700	1	27	12.692	10.769	10.769	28.077
	Poorer	21	297	6.603	1	20	8.081	7.407	8.754	21.212
	Middle	22	320	6.469	1	21	7.188	7.813	9.375	19.375
	Richer	22	323	6.728	1	24	6.502	10.526	7.740	24.149
	Richest	22	262	10.985	1	34	10.687	11.450	10.687	41.221
2011	Poorest	18	484	6.221	1	23	10.950	11.64	11.777	20.661
	Poorer	19	515	6.300	1	18	7.379	9.903	10.485	19.417
	Middle	19	523	6.363	1	18	6.883	8.031	9.560	20.650
	Richer	20	549	6.685	1	21	9.300	8.561	6.193	23.133
	Richest	20	441	9.213	1	31	15.873	10.431	7.029	35.601

Table D.8: Percentages of women in clusters for BMI among ever-married Bangladeshi women aged 15-49 - BDHS 2004, 2007 and 2011 samples: by educational attainment

Survey year	Educational attainment	Number of Strata	Number of clusters	Average size of a cluster	Minimum number of obs. in a cluster	Maximum number of obs. in a cluster	% of clusters with 1 obs.	% of clusters with 2 obs.	% of clusters with 3 obs.	% of clusters with ≥10 children
2004	No education	22	359	12.100	1	29	2.228	1.950	1.114	73.816
	Primary incomplete	22	360	6.303	1	16	3.056	5.556	6.944	19.444
	Primary complete	22	325	3.132	1	11	23.385	19.692	17.846	0.923
	Secondary incomplete	22	358	7.168	1	21	2.793	5.589	5.589	25.970
	Secondary/Higher	22	271	3.616	1	20	28.044	19.557	11.808	13.332
2007	No education	22	358	9.682	1	29	2.514	3.411	6.704	48.324
	Primary incomplete	22	358	6.318	1	18	1.112	7.263	6.145	12.570
	Primary complete	22	322	2.950	1	11	22.981	28.571	16.456	0.621
	Secondary incomplete	22	356	7.326	1	21	2.528	4.775	6.180	24.438
	Secondary/higher	22	311	4.913	1	24	15.434	21.222	13.183	13.183
2011	No education	20	586	7.734	1	25	5.461	4.949	7.508	32.082
	Primary incomplete	20	590	5.310	1	14	6.949	10.339	10.678	6.441
	Primary complete	20	567	3.594	1	11	14.109	20.988	21.869	1.058
	Secondary incomplete	20	594	8.998	1	25	1.684	3.367	4.209	40.741
	Secondary/higher	20	498	4.501	1	22	20.281	17.068	14.458	11.847

Figure D.4: Estimated Gini coefficients by wealth category for BMI: ever-married Bangladeshi women aged 15-49, BDHS 2004, 2007 and 2011

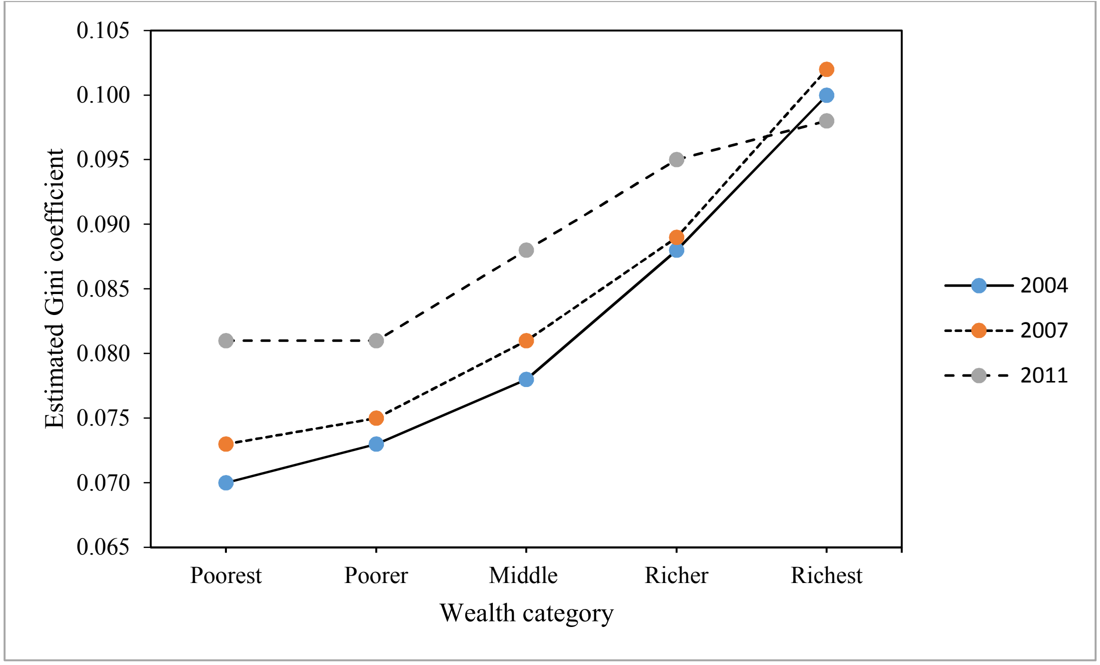


Figure D.5: Estimated Gini coefficients for BMI among ever-married Bangladeshi women aged 15-49 across educational attainment: BDHS 2004, 2007 and 2011 surveys

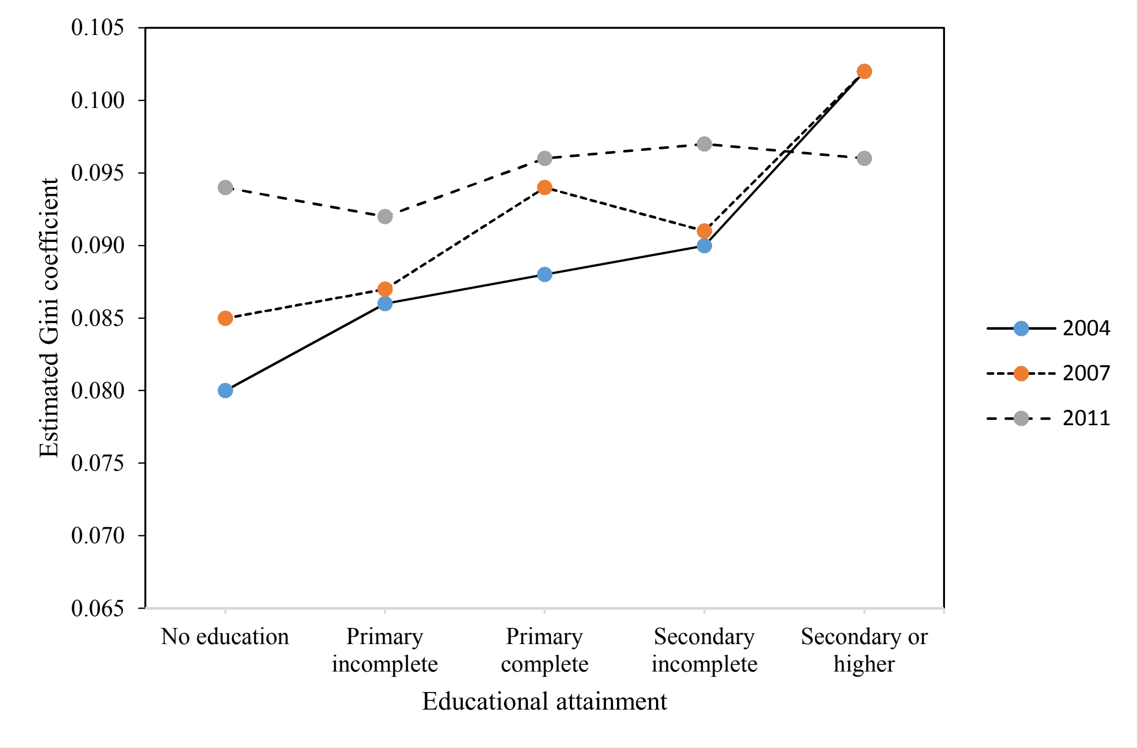


Table D.9: Hypothesis tests for BMI among ever-married Bangladeshi women aged 15-49: by educational attainment, BDHS 2004, 2007 and 2011

Null Hypothesis	Variance Used	Test statistic	Educational attainment				
			No education	Primary incomplete	Primary complete	Secondary incomplete	Secondary or higher
$H_0: G_{2007} = G_{2004}$	$\widehat{Var_{BK}}(\hat{G})$	Wald statistic	8.044	0.176	4.234	0.235	0.000
		p-value	0.005	0.675	0.051	0.628	1.000
	$\widetilde{Var_{BT}}(\hat{G})$	Wald statistic	8.078	0.178	3.341	0.237	0.000
		p-value	0.005	0.673	0.068	0.626	1.000
	$\widetilde{Var_{SRS}}(\hat{G})$	Wald statistic	7.977	0.174	3.151	0.227	0.000
		p-value	0.005	0.680	0.076	0.634	1.000
$H_0: G_{2011} = G_{2007}$	$\widehat{Var_{BK}}(\hat{G})$	Wald statistic	23.629	5.617	0.437	11.222	5.329
		p-value	0.000	0.018	0.509	0.001	0.021
	$\widetilde{Var_{BT}}(\hat{G})$	Wald statistic	23.719	5.650	0.446	11.278	5.380
		p-value	0.000	0.017	0.504	0.001	0.020
	$\widetilde{Var_{SRS}}(\hat{G})$	Wald statistic	23.350	5.538	0.423	10.959	5.334
		p-value	0.000	0.019	0.515	0.001	0.021
$H_0: G_{2011} = G_{2007} = G_{2004}$	$\widehat{Var_{BK}}(\hat{G})$	Wald statistic	78.309	9.126	13.630	19.466	7.177
		p-value	0.000	0.010	0.001	0.000	0.028
	$\widetilde{Var_{BT}}(\hat{G})$	Wald statistic	78.680	9.190	8.776	19.583	7.255
		p-value	0.000	0.010	0.012	0.000	0.027
	$\widetilde{Var_{SRS}}(\hat{G})$	Wald statistic	77.095	8.985	8.455	18.994	7.191
		p-value	0.000	0.011	0.015	0.000	0.027

Table D.10: Percentages of observations in clusters in Hb among Bangladeshi children age 6-59 months from BDHS 2011 surveys: by gender, age, place of residence, wealth and division/region categories

	Number of Strata	Number of clusters	Number of obs.	Average size of a cluster	Minimum number of obs. in a cluster	Maximum number of obs. in a cluster	% of clusters with 1 obs.	% of clusters with 2 obs.	% of clusters with 3 obs.	% of clusters with >3 obs.	% of clusters with ≥10 obs.
All children	20	583	2432	4.172	1	16	10.292	16.981	15.437	57.290	2.715
Gender											
Male	20	503	1244	2.473	1	9	30.020	28.231	20.278	21.471	0
Female	20	495	1188	2.400	1	9	30.909	31.313	18.384	19.394	0
Age Group					1						
6-23 Months	20	425	783	1.842	1	7	48.176	33.882	15.294	5.647	0
24-59 Months	20	556	1649	2.966	1	11	19.065	26.799	19.964	34.173	0.180
Place of Residence											
Rural	7	385	1711	4.444	1	16	9.910	14.286	13.766	62.338	2.597
Urban	13	198	721	3.641	1	9	11.610	22.222	18.687	47.475	0
Wealth Category											
Poorest	17	261	559	2.142	1	7	41.379	27.969	14.559	16.092	0
Poorer	17	255	476	1.867	1	8	50.587	29.020	11.373	9.020	0
middle	18	269	443	1.647	1	9	57.249	28.996	8.550	5.204	0
Richer	20	260	460	1.769	1	6	52.308	28.462	11.923	7.308	0
Richest	20	230	494	2.148	1	14	45.217	26.087	14.348	14.348	0.434
Division/Region											
Barisal	3	71	281	3.958	1	9	9.859	22.535	8.450	59.155	0
Chittagong	3	92	456	4.957	1	16	7.609	7.609	15.217	69.565	3.261
Dhaka	3	103	406	3.942	1	10	38.834	16.505	19.417	51.456	1.941
Khulna	3	82	273	3.329	1	8	13.414	21.951	21.951	42.683	0
Rajshahi	3	82	294	3.585	1	8	13.414	21.951	13.415	51.220	0
Rangpur	2	83	323	3.892	1	11	9.639	18.072	20.482	51.507	1.205
Sylhet	3	70	399	5.700	1	14	1.493	11.940	5.714	78.571	5.714

Table D.11: Estimated intracluster correlation coefficients and components of variance for all children by stratum, BDHS 2011

Stratum	Number of clusters in h	Average Cluster Size, m_h	Within cluster variance, $\hat{\sigma}_h^2$	Between cluster variance, $\hat{\gamma}_h^2$	Intracluster correlation coefficient, $\hat{\rho}_h$
1. Barisal City Corp	7	3.64	1.810	0.294	0.140
2. Chittagong City Corp.	16	3.62	1.641	0.237	0.126
3. Dhaka City Corp.	20	2.22	3.768	0.000*	0.000
4. Khulna City Corp.	8	2.95	1.375	0.000*	0.000
5. Rajshahi City Corp.	5	3.21	3.490	0.000*	0.000
6. Sylhet City Corp.	10	5.08	1.760	0.048	0.027
7. Barisal Other Urban	15	3.75	0.923	0.190	0.171
8. Chittagong Other Urban	18	5.13	1.318	0.094	0.067
9. Dhaka Other Urban	24	3.72	1.315	0.014	0.011
10. Khulna Other Urban	18	2.78	1.487	0.000*	0.000
11. Rajshahi Other Urban	23	3.37	1.688	0.000*	0.000
12. Rangpur Other Urban	23	3.39	1.524	0.008	0.005
13. Sylhet Other Urban	12	4.34	1.626	0.000*	0.000
14. Barisal Rural	49	3.98	1.556	0.000*	0.000
15. Chittagong Rural	59	5.21	1.403	0.179	0.113
16. Dhaka Rural	59	4.57	1.350	0.032	0.023
17. Khulna Rural	56	3.52	1.576	0.030	0.019
18. Rajshahi Rural	54	3.66	1.318	0.087	0.062
19. Rangpur Rural	60	4.05	1.227	0.044	0.035
20. Sylhet Rural	48	6.10	1.639	0.122	0.069

Table D.12 Mean statistics and hypothesis tests for Hb levels: all children and by gender, age groups 6-23 and 24-59 months and place of residence, BDHS 2011

	Mean Hb Level, g/dl			Hypothesis tests $H_0: \mu_{\text{male}} = \mu_{\text{female}}$ $H_1: \mu_{\text{male}} \neq \mu_{\text{female}}$ Wald Test	
	All Children	Male	Female	$\chi^2(1)$	p-value
All children					
Mean	10.794	10.767	10.824	1.030	0.311
Lin. St. error of mean	0.032	0.043	0.042		
Number of Obs.	2432	1244	1188		
Age 6-23 months					
Mean	10.188	10.121	10.259	1.630	0.202
Lin. St. error of mean	0.056	0.075	0.080		
Number of Obs.	783	401	382		
Age 24-59 months					
Mean	11.086	11.081	11.090	0.020	0.876
Lin. St. error of mean	0.035	0.045	0.045		
Number of Obs.	1649	843	806		
Rural Children					
Mean	10.761	10.719	10.804	1.720	0.191
Lin. St. error of mean	0.037	0.050	0.049		
Number of Obs.	1711	866	845		
Urban Children					
Mean	10.923	10.942	10.902	0.130	0.716
Lin. St. error of mean	0.061	0.080	0.085		
Number of Obs.	721	378	343		

Hypothesis tests

$H_0: \mu_{6-23} = \mu_{24-59}$ and $H_1: \mu_{6-23} \neq \mu_{24-59}$

Wald Test: $\chi^2(1)$= 200.500; p-value=0.00

$H_0: \mu_{\text{rural}} = \mu_{\text{urban}}$ and $H_1: \mu_{\text{rural}} \neq \mu_{\text{urban}}$

Wald Test: $\chi^2(1)$=5.130; p-value=0.024

Notes: Lin. St. error (linearized standard error) is the survey design adjusted standard error. Recall, the mean is estimated assuming a stratified sampling design.

Table D.13: Mean level of Hb among Bangladeshi children aged 6-59 months and hypothesis tests, BDHS 2011: by wealth category

	Mean Hb Level, g/dl				
	Poorest	Poorer	Middle	Richer	Richest
All children					
Mean	10.663	10.555	10.874	10.980	10.985
Lin. St. error	0.066	0.061	0.071	0.061	0.073
Number of Obs.	559	476	443	460	494
Age 6-23 months					
Mean	10.035	9.982	10.205	10.381	10.406
Lin. St. error	0.135	0.117	0.129	0.103	0.130
Number of Obs.	153	168	144	155	163
Age 24-59 months					
Mean	10.884	10.900	11.203	11.030	11.282
Lin. St. error	0.074	0.074	0.076	0.070	0.092
Number of Obs.	406	303	299	305	331

Hypothesis Test

All Children

$H_0: \mu_{\text{poorest}} = \mu_{\text{poorer}} = \mu_{\text{middle}} = \mu_{\text{richer}} = \mu_{\text{richest}}$

Wald Test:

$\chi^2(4)$=37.2; p-value=0.000

By Wealth Category

$H_0: \mu_{6-23} = \mu_{24-59}$ and $H_1: \mu_{6-23} \neq \mu_{24-59}$

Wald Test:

	$\chi^2(1)$	p-value
Poorest	35.430	0.000
Poorer	42.240	0.000
Middle	54.670	0.000
Richer	49.920	0.000
Richest	26.080	0.000

Notes: In survey design adjusted mean comparison hypothesis testing, adjusted Wald statistic has F distribution with finite sample. Chi-square statistics are calculated as $\chi^2(\text{df1}) = \text{df1} \times F_{(\text{df1},\text{df2})}$ for asymptotic approximation.

Table D.14: Mean level of Hb among Bangladeshi children aged 6-59 months and hypothesis tests, BDHS 2011: by division or region

	Mean Hb Level, g/dl						
	Barisal	Chittagong	Dhaka	Khulna	Rajshahi	Rangpur	Sylhet
All children							
Mean	10.568	10.789	10.841	10.885	10.884	10.626	10.782
Lin. St. error	0.078	0.084	0.067	0.080	0.069	0.071	0.084
Num. of Obs.	281	456	406	273	294	323	399
Age 6-23 months							
Mean							
Lin. St. error	9.821	10.178	10.185	10.392	10.312	10.056	10.194
Num. of Obs.	0.141	0.105	0.124	0.179	0.142	0.134	0.113
	87	168	130	84	98	95	121
Age 24-59 months							
Mean	10.877	11.155	11.147	11.095	11.165	10.848	11.049
Lin. St. error	0.083	0.081	0.076	0.089	0.094	0.080	0.092
Num. of Obs.	194	288	276	189	196	228	278

Hypothesis test

All Children

H_0: $\mu_{\text{Bari}} = \mu_{\text{Chitt}} = \mu_{\text{Dhak}} = \mu_{\text{Khul}} = \mu_{\text{Raj}} = \mu_{\text{Ran}} = \mu_{\text{Syl}}$
Wald Test:
$\chi^2(6)$=16.740; p-value=0.011

By Division/Region

H_0: $\mu_{6-23} = \mu_{24-59}$ and H_1: $\mu_{6-23} \neq \mu_{24-59}$
Wald Test:

	$\chi^2(1)$	p-value
Barisal	42.160	0.000
Chittagong	83.080	0.000
Dhaka	43.380	0.000
Khulna	14.230	0.000
Rajshahi	21.940	0.000
Rangpur	23.790	0.000
Sylhet	47.610	0.000

Note: In survey design adjusted mean comparison hypothesis testing, adjusted Wald statistic has F distribution with finite sample. Chi-square statistics are calculated as $\chi^2(\text{df1}) = \text{df1} \times F_{(\text{df1,df2})}$ for asymptotic approximation.

Appendix E

Sample Program Codes

Chapter 3

The following EViews code generates population data with two strata and an ICC parameter of zero, for which strata 1 and 2 contain 400 and 500 clusters, respectively, under the Beta distribution. The population is called DGP 1A.

```
'Strata#2, N1=400, N2=500, HH=80~200, CE=0
'creating two strata with 400 and 500 clusters in stratum 1 and 2,
respectively. In each cluster, assumed # of household is between 80-200.

rndseed 1234567
smpl @all

vector (400) cluster_s1
rndint(cluster_s1, 120)
vector cluster_s1=cluster_s1+80

vector (500) cluster_s2
rndint(cluster_s2, 120)
vector cluster_s2=cluster_s2+80

'total # households & mean # of households
for !i =1 to 2
scalar mean_hh_s!i=@mean(cluster_s!i)
scalar sum_hh_s!i=@sum(cluster_s!i)
next

' mean_hh_s1=141; mean_hh_s2=139; sum_hh_s1=56427; sum_hh_s2=69461; total#
HHs=125,888

' create full matrix of id's and values
' col1 =strata; col2=clusterid; col3=hh id; col4=y; col5= sample weight for
n1=10, n2=15; col6= sample weight for n1=20, n2=25; col7=sample weight for
n1=40&n2=50, col8=sample weight for n1=80, n2=100. sample weight is generated
assuming that 30 HH sampled from each cluster, 40 clusters sampled from
stratum 1 and 50 clusters sampled from stratum 2.

matrix(125888,8) full

'clusterid
smpl @all
series clusterid=0
scalar t1=0
for !i=1 to 400
scalar t2=cluster_s1(!i,1)-1+t1
smpl @first+t1 @first+t2
genr clusterid=!i
scalar t1=t2+1
next
```

```
scalar t1=56427
for !i=1 to 500
scalar t2=cluster_s2(!i,1)-1+t1
smpl @first+t1 @first+t2
genr clusterid=!i+400
scalar t1=t2+1
next
smpl 1 125888
stom(clusterid, clusteridm)
matplace(full,clusteridm,1,2)

'hh id
smpl @all
series hhid=0
genr hhid=@trend+1

smpl 1 125888
stom(hhid, hhidm)
matplace(full,hhidm,1,3)

' for HHs in each stratum generate well-being variable and weight
' strata=0 for S1 and strata=1 for S2

smpl @all

series y=0
series w=1
series strata=0
series fdist=0

smpl @first 56427
genr y=2+@rbeta(2,5)

smpl 56428 125888
genr y=2+@rbeta(3,6)
genr strata=1

' weight series based on 30 HH sampled from each cluster, 40 clusters sampled
from stratum 1 and 50 clusters sampled from stratum 2
'sample weights are created for four samples
'sample weight 1 is for when n1=10 and n2=15

smpl @all
series weight1=0

'stratum 1
scalar t1=0
for !i=1 to 400
scalar t2=cluster_s1(!i,1)-1+t1
scalar tt2=cluster_s1(!i,1)
smpl @first+t1 @first+t2
genr weight1=(tt2*400)/(30*10)
scalar t1=t2+1
next

scalar t1=56427
for !i=1 to 500
scalar t2=cluster_s2(!i,1)-1+t1
scalar tt2=cluster_s2(!i,1)
smpl @first+t1 @first+t2
genr weight1=(tt2*500)/(30*15)
scalar t1=t2+1
next
```

```
smpl 1 125888
stom(weight1, weightm1)
matplace(full,weightm1,1,5)


'sample weight 2 is for when n1=20 and n2=25

smpl @all
series weight2=0

'stratum 1
scalar t1=0
for !i=1 to 400
scalar t2=cluster_s1(!i,1)-1+t1
scalar tt2=cluster_s1(!i,1)
smpl @first+t1 @first+t2
genr weight2=(tt2*400)/(30*20)
scalar t1=t2+1
next

'stratum 2
scalar t1=56427
for !i=1 to 500
scalar t2=cluster_s2(!i,1)-1+t1
scalar tt2=cluster_s2(!i,1)
smpl @first+t1 @first+t2
genr weight2=(tt2*500)/(30*25)
scalar t1=t2+1
next

smpl 1 125888
stom(weight2, weightm2)
matplace(full,weightm2,1,6)

'sample weight 3 is for when n1=40 and n2=50
smpl @all
series weight3=0

'stratum 1
scalar t1=0
for !i=1 to 400
scalar t2=cluster_s1(!i,1)-1+t1
scalar tt2=cluster_s1(!i,1)
smpl @first+t1 @first+t2
genr weight3=(tt2*400)/(30*40)
scalar t1=t2+1
next

'stratum 2
scalar t1=56427
for !i=1 to 500
scalar t2=cluster_s2(!i,1)-1+t1
scalar tt2=cluster_s2(!i,1)
smpl @first+t1 @first+t2
genr weight3=(tt2*500)/(30*50)
scalar t1=t2+1
next
smpl 1 125888
stom(weight3, weightm3)
matplace(full,weightm3,1,7)

'sample weight 4 is for when n1=80 and n2=100
```

```
smpl @all
series weight4=0

'stratum 1
scalar t1=0
for !i=1 to 400
scalar t2=cluster_s1(!i,1)-1+t1
scalar tt2=cluster_s1(!i,1)
smpl @first+t1 @first+t2
genr weight4=(tt2*400)/(30*80)
scalar t1=t2+1
next

'stratum 2
scalar t1=56427
for !i=1 to 500
scalar t2=cluster_s2(!i,1)-1+t1
scalar tt2=cluster_s2(!i,1)
smpl @first+t1 @first+t2
genr weight4=(tt2*500)/(30*100)
scalar t1=t2+1
next

smpl 1 125888
stom(weight4, weightm4)
matplace(full,weightm4,1,8)

stom(y, yvec)
stom(strata, strat)

'strata
matplace(full,strat,1,1)

'y series
matplace(full,yvec,1,4)
```

The following Stata code is for selecting a sample with 40 and 50 clusters from strata 1 and 2, respectively, from DGP1A, along with estimating the Gini coefficient, its variances and 95% confidence intervals.

```
  set memory 1g
  set matsize 1000
/* this file uses 1000 reps in Stata 9 - see notes for some explanations */
/* this file codes a MC experiment that examines the bias for the Gini
coefficient estimator, CI coverage when using the linearization variance, and
examines its CI coverage */
/* this code also works out the lower error percentage rate (LP), upper error
percentage rate (UP)*/
/* The columns in the dta file are:  strata (0/1); clusterid (1 to 900: 1-400
for stratum 1; 401-900 for stratum 2); hhid (HH id); weight (this is calculated
assuming 40 clusters are sampled from stratum 1 and 50 from stratum 2; i.e.,
10%) - the weights need normalizing each time a sample is drawn; y - the "well-
being" variables, as described and generated in EViews codes above.

  set more off
  set seed 1234567
  scalar count_ciBT=0
  scalar count_ciBK=0
  scalar count_ciB=0
  scalar count_ciP=0
  scalar count_ciWP=0
  scalar Lcount_B=0
  scalar Ucount_B=0
  scalar Lcount_P=0
  scalar Ucount_P=0
  scalar Lcount_BT=0
  scalar Ucount_BT=0
  scalar Lcount_BK=0
  scalar Ucount_BK=0
  scalar Lcount_WP=0
  scalar Ucount_WP=0
  scalar popgini=0.0381
  scalar bias=0
  scalar nrep=1000
  scalar brep=99
  scalar avg_varBT=0
  scalar avg_varBK=0
  scalar avg_varboot=0
  matrix g_mc=J(nrep,1,0)
  matrix diffw=J(nrep,1,0)
  matrix g_wboot=J(nrep,1,0)

  local i=1
  while `i' <= nrep {
  scalar countmc=`i'
  scalar list countmc

/* uses full dataset, sample 40 clusters from stratum 1 and 50 clusters from
stratum 2. */
  g smpwt=0
  g fdist=0
  g term=0
  g terma=0
  g termb=0
  g uterm=0
  g sumu=0
```

```
  g utermmod=0
  g utermmod_b=0
  bsample, cluster (clusterid) strata(strata) idcluster (idc_mc)
  qui drop if strata==0 & idc_mc>40
  qui drop if strata==1 & idc_mc>50
  qui replace idc_mc=40+idc_mc if strata==1

/* now need to randomly draw 30 households from each cluster - key point is to
have the 30 HH differ across MC reps, as this is what would happen in "real
life - do as follows */
  local k=1
  while `k' <= 90 {
  qui summ hhid if idc_mc==`k'
  scalar max=r(max)
/* pick a random number from r(min) to r(max) */
  matrix hh=matuniform(1,1)
  scalar hhbeg=r(min)+int((r(max)-r(min)+1)*hh[1,1])
/* check can keep right # of households, if not go back 29 households */
  scalar hht=max-29
  if hht<hhbeg {
  scalar hhbeg=hht
  }
  scalar hhend=hhbeg+29
  qui drop if idc_mc==`k' & hhid<hhbeg
  qui drop if idc_mc==`k' & hhid>hhend

/* dataset will now consist of 30 HHs in each cluster in the MC experiment of
40 clusters in stratum 1 and 50 clusters in stratum 2 and the HHs should
reasonably differ across MC reps when the same clusters are drawn */
  local k=`k'+1
  }

  qui total weight
  matrix wt=e(b)
  scalar wt1=wt[1,1]
  qui replace smpwt=weight/wt1
  qui svyset, clear
  qui svyset idc_mc[pweight=smpwt], strata(strata)

/*calculating Gini using formula*/
  qui svy: mean y
  matrix coeff=e(b)
  scalar mu=coeff[1,1]
  sort y
  qui replace fdist=sum(smpwt)
  qui replace term=fdist*y
  qui svy: total term
  matrix coeff=e(b)
  scalar term1=coeff[1,1]
  scalar gini=(2/mu)*term1-1
  scalar bias=bias+(gini-popgini)
  matrix g_mc[`i',1]=gini

/*variance estimation*/
/* Forming various components for variance calculations*/
  qui replace terma=fdist-(gini+1)/2
  qui replace termb=mu-sum(y*smpwt)
  qui replace uterm=(2/mu)*(terma*y+termb-(mu/2)*(gini+1))

/* the following generates B&K se */
  qui svy: total uterm
  matrix temp=e(V)
  scalar var_BK=temp[1,1]
```

```
  scalar se_BK=sqrt(temp[1,1])
  scalar avg_varBK=avg_varBK+var_BK

/*the following generates Bhattacharya's se */
  qui replace utermmod=sqrt(39/40)*uterm if strata==0
  qui replace utermmod=sqrt(49/50)*uterm if strata==1
  qui svy: total utermmod
  matrix temp=e(V)
  scalar var_BT=temp[1,1]
  scalar se_BT=sqrt(temp[1,1])
  scalar avg_varBT=avg_varBT+var_BT

/* Bootstrap Gini using standard method & using n_h*=n_h-1 */
  g fwt=0
  g fwt1=0
  g fwt2=0
  g newsmpwt=0
  g fdist_b=0
  g term_b=0
  g sq_diff=0
  local b=1

/* recall that for the endpoints of the bootstrapped percentile CI to be exact
we need alpha(brep+1)/2 to be an integer */
  matrix g_boot=J(brep,1,0)
  matrix diff=J(brep,1,0)
  while `b'<=brep{
  scalar count_b=`b'
  scalar list count_b
  qui replace fwt=0
  qui replace fwt1=0
  qui replace fwt2=0
  qui replace newsmpwt=0
  qui replace fdist_b=0
  qui replace term_b=0

  bsample 39 if strata==0, weight(fwt1) cluster (idc_mc)
  bsample 49 if strata==1, weight(fwt2) cluster (idc_mc)

  qui replace fwt=fwt1+fwt2
  qui replace newsmpwt=fwt*weight
  qui total newsmpwt
  matrix wt=e(b)
  scalar wt2=wt[1,1]
  qui replace newsmpwt=newsmpwt/wt2
  qui svyset, clear
  qui svyset idc_mc[pweight=newsmpwt], strata(strata)
  qui sort y
  qui replace fdist_b=sum(newsmpwt)
  qui replace term_b=fdist_b*y
  qui svy: mean y
  matrix a=e(b)
  scalar mu_b=a[1,1]
  qui svy: total term_b
  matrix aa=e(b)
  scalar sum_b=aa[1,1]
  scalar g_b=((2/mu_b)*sum_b)-1
  matrix g_boot[`b',1]=g_b
/* use when b=1 for warp-speed MC percentile */
  if `b'==1 {
  matrix g_wboot[`i',1]=g_b
  matrix diffw[`i',1]=g_wboot[`i',1]-g_mc[`i',1]
  }
```

```
/* form diff */
  matrix diff[`b',1]=g_b-gini
  local b=`b'+1
  }
/* form bootstrap se */
  svmat g_boot, names(g_boots)
  qui mean g_boots
  matrix coeff=e(b)
  scalar mean_gboot=coeff[1,1]
  qui replace sq_diff=(g_boots-mean_gboot)^2
  qui replace sq_diff=. if sq_diff==0
  qui total sq_diff
  matrix aa=e(b)
  scalar tt=aa[1,1]
  scalar var_boot=tt/(brep-1)
  scalar se_boot=sqrt(var_boot)
  scalar avg_varboot=avg_varboot+var_boot

/* following is often called bootstrapped normal-approximation confidence
interval */
  scalar LLB=gini-se_boot*(invnormal(0.975))
  scalar ULB=gini+se_boot*(invnormal(0.975))
  if LLB<popgini & ULB>popgini {
  scalar count_ciB=count_ciB+1
  }
/* calculate counts for tail error rates */
  if LLB>popgini {
  scalar Lcount_B=Lcount_B+1
  }
  if ULB<popgini {
  scalar Ucount_B=Ucount_B+1
  }
/* standard percentile CI */
  svmat diff, names(diffb)
  sort diffb
  _pctile diffb, p(2.5,97.5)
  scalar Lsp=r(r1)
  scalar Usp=r(r2)
  scalar LLP=gini-Usp
  scalar ULP=gini-Lsp
  if LLP<popgini & ULP>popgini {
  scalar count_ciP=count_ciP+1
  }
/* calculate counts for tail error rates */
  if LLP>popgini {
  scalar Lcount_P=Lcount_P+1
  }
  if ULP<popgini {
  scalar Ucount_P=Ucount_P+1
  }
/*1. 95% Ordinary CI for Bhattacharya, Binder & Kovacevic and Bootstrap*/
/* often called normal approximation confidence intervals using asymptotic
standard errors */
  scalar LLBT=gini-se_BT*(invnormal(0.975))
  scalar ULBT=gini+se_BT*(invnormal(0.975))
  scalar LLBK=gini-se_BK*(invnormal(0.975))
  scalar ULBK=gini+se_BK*(invnormal(0.975))

/*Calculating coverage probabilities for 95% CI*/
  if LLBT<popgini & ULBT>popgini {
  scalar count_ciBT=count_ciBT+1
  }
  if LLBK<popgini & ULBK>popgini {
```

```
  scalar count_ciBK=count_ciBK+1
  }
/* calculate counts for tail error rates */
  if LLBT>popgini {
  scalar Lcount_BT=Lcount_BT+1
  }
  if ULBT<popgini {
  scalar Ucount_BT=Ucount_BT+1
  }
  if LLBK>popgini {
  scalar Lcount_BK=Lcount_BK+1
  }
  if ULBK<popgini {
  scalar Ucount_BK=Ucount_BK+1
  }
/*dropping everything except scalars & matrices */
  local i=`i'+1
  }

/* form warp-speed bootstrapped CI */
  svmat diffw, names(dw)
  sort dw
  _pctile dw, p(2.5,97.5)
  scalar Ltw=r(r1)
  scalar Utw=r(r2)
  local j=1
  while `j' <= nrep {
  scalar LLWP=g_mc[`j',1]-Utw
  scalar ULWP=g_mc[`j',1]-Ltw
  if LLWP<popgini & ULWP>popgini {
  scalar count_ciWP=count_ciWP+1
  }
  if LLWP>popgini {
  scalar Lcount_WP=Lcount_WP+1
  }
  if ULWP<popgini {
  scalar Ucount_WP=Ucount_WP+1
  }
  local j=`j'+1
  }
/* average variances */
  scalar list avg_varBT avg_varBK avg_varboot
  scalar bias_mc=bias/nrep
  scalar list bias_mc
  scalar count_ciBT=count_ciBT/nrep*100
  scalar count_ciBK=count_ciBK/nrep*100
  scalar count_ciB=count_ciB/nrep*100
  scalar count_ciP=count_ciP/nrep*100
  scalar count_ciWP=count_ciWP/nrep*100
  scalar Lcount_BT=Lcount_BT/nrep*100
  scalar Lcount_BK=Lcount_BK/nrep*100
  scalar Lcount_B=Lcount_B/nrep*100
  scalar Lcount_P=Lcount_P/nrep*100
  scalar Lcount_WP=Lcount_WP/nrep*100
  scalar Ucount_BT=Ucount_BT/nrep*100
  scalar Ucount_BK=Ucount_BK/nrep*100
  scalar Ucount_B=Ucount_B/nrep*100
  scalar Ucount_P=Ucount_P/nrep*100
  scalar Ucount_WP=Ucount_WP/nrep*100

  scalar list count_ciBT count_ciBK count_ciB count_ciP count_ciWP
  scalar list Lcount_BT Lcount_BK Lcount_B Lcount_P Lcount_WP
  scalar list Ucount_BT Ucount_BK Ucount_B Ucount_P Ucount_WP
```

Chapter 4

The following Stata code is for estimating the Gini coefficient, its variances and 95% confidence intervals, for BMI among women (all women) with BDSH2011 survey (see Table 15 and 20 for results)

```
/*Following codes are to estimate Gini coefficient, it variances using Binder
and Kovacevic (1995), Bhattacharya (2007) and bootstrap methods, and to
construct 95% confidence intervals using these variances for all women from
BDHS 2011 data.

The original data file BDIR61FL contained information on 17,842 ever-married
12-49 years old women. After dropping observations with V445 (=BMI) =9998
(flagged cases) and =9999 (missing) and those age 12-14, there are 17,309 women
in the sample, called all women sample. The original data was reported as 4
digits with no decimals, so divided by 100 - called BMI.

The sample weight, V005, is divided by 10^6, and named as smplwt.  Variable
smplwt is normalized, so that the sum is equal, named lswt.

There are 20 strata - V023 and the cluster variable (PSUs) is V021.

The dataset (bmi) has been saved with the following svyset command:

svyset v021 [pweight=lswt], strata(v023)

In addition, as the data contain women from age 15-18, which are not expected
to confirm with those for adult women >=19, below we standardize the BMI for
these women to those of a 19 years old, using the WHO percentiles, and the LMS
transformation. */

  set memory 1g
  set more off

  g bmi_std=bmi
  g zterm=0

/* standardize the BMI for 15-18 year old women to that of 19 year old females
- see, e.g., appendix 1 of Araar et al.2009 and de Onis et al. 2007. Age in
years is v447a, so use age:0 LMS coefficients */

matrix st1=(15,-1.1311,20.2125,0.13904\16,-1.0368,20.7008,0.14070\17,-
0.9423,21.0367,0.14208\18,-0.8462,21.2603,0.1433\19,-0.7496,21.4269,0.14441)

  forvalues j=1(1) 4 {
  scalar t2=`j'
  scalar list t2
  scalar tt2=`j'+14
  scalar list tt2
  quietly replace
  zterm=((bmi_std/st1[`j',3])^(st1[`j',2]))/((st1[`j',4])*(st1[`j',2])) if
  v447a==tt2

  quietly replace
  bmi_std=((zterm*(st1[5,2])*(st1[5,4])+1)^(1/(st1[5,2])))*(st1[5,3]) if
  v447a==tt2
  }
```

```
/* full sample */

sort bmi_std

/* estimate distribution function using normalized weights & run regression*/
g fdist=sum(lswt)
g term=sqrt(lswt*bmi_std)
g newd=(2*fdist-1)*term
regress newd term, noc

/* using svy:regress */
g term2=sqrt(bmi_std)
g newd2=(2*fdist-1)*term2
svy:regress newd2 term2, noc
matrix coef=e(b)
scalar gini=coef[1,1]
svylorenz bmi_std

/* form terms for B&K u* */
g aterm=fdist-(gini+1)/2
svy: mean bmi_std
matrix mu=e(b)
scalar mus=mu[1,1]
g termb=lswt*bmi_std
g bterm=-sum(termb)+mus
g uterm=(2/mus)*(aterm*bmi_std+bterm-(mus/2)*(gini+1))
svy: total uterm
matrix temp=e(V)
scalar se_BK=sqrt(temp[1,1])
scalar var_BK=temp[1,1]
scalar se_BK=sqrt(var_BK)


/* form Bhattacharya's variance for this purpose as modifcation of B&K */
g utermmod=0
qui replace utermmod=sqrt(6/7)*uterm if v023==1
qui replace utermmod=sqrt(15/16)*uterm if v023==2
qui replace utermmod=sqrt(22/23)*uterm if v023==3
qui replace utermmod=sqrt(8/9)*uterm if v023==4
qui replace utermmod=sqrt(4/5)*uterm if v023==5
qui replace utermmod=sqrt(9/10)*uterm if v023==6
qui replace utermmod=sqrt(14/15)*uterm if v023==7
qui replace utermmod=sqrt(16/17)*uterm if v023==8
qui replace utermmod=sqrt(25/26)*uterm if v023==9
qui replace utermmod=sqrt(19/20)*uterm if v023==10
qui replace utermmod=sqrt(22/23)*uterm if v023==11
qui replace utermmod=sqrt(23/24)*uterm if v023==12
qui replace utermmod=sqrt(11/12)*uterm if v023==13
qui replace utermmod=sqrt(49/50)*uterm if v023==14
qui replace utermmod=sqrt(58/59)*uterm if v023==15
qui replace utermmod=sqrt(59/60)*uterm if v023==16
qui replace utermmod=sqrt(55/56)*uterm if v023==17
qui replace utermmod=sqrt(58/59)*uterm if v023==18
qui replace utermmod=sqrt(60/61)*uterm if v023==19
qui replace utermmod=sqrt(47/48)*uterm if v023==20

svy: total utermmod
matrix tmp=e(V)
scalar se_BT=sqrt(tmp[1,1])
scalar var_BT=tmp[1,1]
scalar list var_BT
```

```
//Calculating bootstrap variance
g fwt=0
g fwt1=0
g fwt2=0
g bsmpwt=0
g fdist_b=0
g term_b=0
g sqdiff=0
g gini_b=0
g sum_d=0
g terma_boot=0
g termb_boot=0
g uterm_boot=0

set matsize 800
scalar brep=799

matrix g_boot=J(brep,1,0)
matrix t_boot=J(brep,1,0)

local i=1
while `i'<=brep{
scalar count_b=`i'
scalar list count_b
qui bsample, weight(fwt) cluster(v021) strata(v023)
qui replace bsmpwt=fwt*weight
qui total bsmpwt
matrix wt=e(b)
scalar wt1=wt[1,1]
qui replace bsmpwt=bsmpwt/wt1

qui svyset, clear
qui svyset v021[pweight=bsmpwt], strata(v023)

qui sort bmi_std
qui replace fdist_b=sum(bsmpwt)
qui replace term_b=fdist_b*bmi_std

qui svy: mean bmi_std
matrix a=e(b)
scalar mu_b=a[1,1]

qui svy: total term_b
matrix aa=e(b)
scalar sum_b=aa[1,1]
scalar gini_boot=((2/mu_b)*sum_b)-1
matrix g_boot[`i',1]=gini_boot
replace terma_boot=fdist_b-(gini_boot+1)/2
replace termb_boot=mu_b-sum(bmi_std*bsmpwt)
replace uterm_boot=(2/mu_b)*(terma_boot*bmi_std+termb_boot-
(mu_b/2)*(gini_boot+1))

quie svy:total uterm_boot
matrix temp1=e(V)
scalar se_BK_boot=sqrt(temp1[1,1])

scalar t=(gini_boot- 0.097)/se_BK_boot
scalar list t
matrix t_boot[`i',1]=t
replace fwt=0
replace bsmpwt=0
local i=`i'+1
}
```

```
svmat g_boot, names(gini_b)

total gini_b1
matrix coeff=e(b)
scalar sum_gboot=coeff[1,1]
scalar mean_ginib=sum_gboot/brep
replace sqdiff=(gini_b1-mean_ginib)^2
total sqdiff
matrix aa=e(b)
scalar tt=aa[1,1]
scalar var_boot=tt/(brep-1)

scalar se_boot=sqrt(var_boot)
svmat t_boot, names(percentile)
sort percentile
_pctile percentile, p(2.5, 97.5)

scalar L_per=r(r1)
scalar U_per=r(r2)

scalar list  se_BK se_BT se_boot
scalar list var_BK var_BT var_boot

//95% Confidence interval

//Using Binder and Kovacevic’s variance

scalar LL_bk= .097-se_BK*(invnormal(0.975))
scalar UL_bk= .097+se_BK*(invnormal(0.975))

//Using Bhattacharya’s variance
scalar LL_bt= .097-se_BT*(invnormal(0.975))
scalar UL_bt=  .097+se_BT*(invnormal(0.975))

//Boot percentile

sort gini_b1
_pctile gini_b1, p(2.5, 97.5)
scalar LL_pt=r(r1)
scalar UL_pt=r(r2)

//Boot t-percentile
scalar Lp= 0.097-se_BK*U_per
scalar Up= .097-se_BK*L_per

scalar list var_BK var_BT var_boot
scalar list LL_bk UL_bk LL_bt UL_bt LL_pt UL_pt Lp Up
```

The following EViews code is for decomposing Bhattacharya's variance into the SRS variance, cluster effect and stratum effect (see Table 15 for estimates)

```
'2011 Bangadesh data - there are 17,309 observations, variables v021, v023,
v025, uterm & lswt are transferred onto EViews.

  smpl @all

  sort v021

' recall that ustar=uterm * lswt

  scalar sum=0
  scalar srsv=0

' the following calculates the cluster effect + srsvar as it forms the sum for
each cluster over all k
' to get the cluster effect just subtract off the srsvar, as this is the
sum(sum of the diagonal elements of the tmat matrix)

  for !i=1 to 600
  smpl @all if v021=!i
  stom(ustar, ustar_mat)
  matrix tmat=ustar_mat*@transpose(ustar_mat)
  scalar srsv=srsv+@trace(tmat)
  vector vmat=@vec(tmat)
  matrix dmat=@makediagonal(vmat)
  matrix ssum=@trace(dmat)
  scalar sum=sum+ssum(1,1)
  delete ustar_mat tmat vmat dmat
  next

  scalar cluster = sum-srsv

'third term, stratification effect. Number of clusters in each stratum are
stored in the vector cn.

  scalar stratavar=0

  for !i=1 to 20

  smpl @all if v023=!i
  scalar sum_!i=@sum(ustar)
  scalar stratavar=stratavar+((sum_!i)^2)/cn(!i,1)
  nextscalar var_bhatt=srsv+cluster-stratavar
```