ESTIMATING GINI RATIOS WITH VARYING PROPORTIONATE STRATIFIED SAMPLING

by

C. Geoffrey Swenson

May 1978 No. 78-14

ESTIMATING GINI RATIOS WITH VARYING PROPORTIONATE STRATIFIED SAMPLING

C. Geoffrey Swenson*

The purpose of this note is to demonstrate how the Gini ratio can be computed when a stratified sample of a given population has been taken and the sampling proportions vary between strata. Normally, one might expect the sampling proportion to be constant for all strata if a stratified sample is taken. However, there may be circumstances, as outlined below, when it is advisable to vary the proportion taken in each strata. This note will show how resultant data can be adjusted in order to combine the strata and compute the representative Gini ratio for the total population.

The procedures outlined herein were developed out of necessity. The author had collected data on incomes in two villages in South India where the population was composed of farm operators and landless laborers.1/

In order to learn as much as possible about the varying effects of a technological change in rice production and because the total number of farm operators in the two villages was within a reasonable limit for statistical analysis, a complete census of the farm operator group was accomplished. However, in the belief that the variation in income was much smaller between households for the landless laborer group, a 30 percent random sample of landless laborer households was drawn. After discarding non-usable questionnaires, the final sample was as shown below:

	Total number of households	Sample households 2/	Percent
Farm operators	156	145	93
Landless laborers	247	67	27

* Associate of The Agricultural Development Council, Inc., Visiting Fellow with the Department of Agricultural Economics, Cornell University; Ithaca, New York.

1/ See Swenson, Clyde Geoffrey, "The Effect of Increases in Rice Production on Employment and Income Distribution in Thanjavur District, South India," unpublished Ph.D. dissertation, Michigan State University, 1973.

2/ It may be noted that an ex-post analysis of this sampling procedure revealed that the resultant varying sampling proportion was quite appropriate for this population. Using the estimated means and a confidence interval of 10 percent on either side of the mean, the appropriate sample size would have been 146 and 68 for the farm operator and landless laborer household groups, respectively.

The basic equation used for computing the Gini ratio was developed by Morgan3/ and given below for data which has been arrayed in ascending order:

$$G = 1 - \Sigma(B-A)(C+D) \tag{1}$$

where:

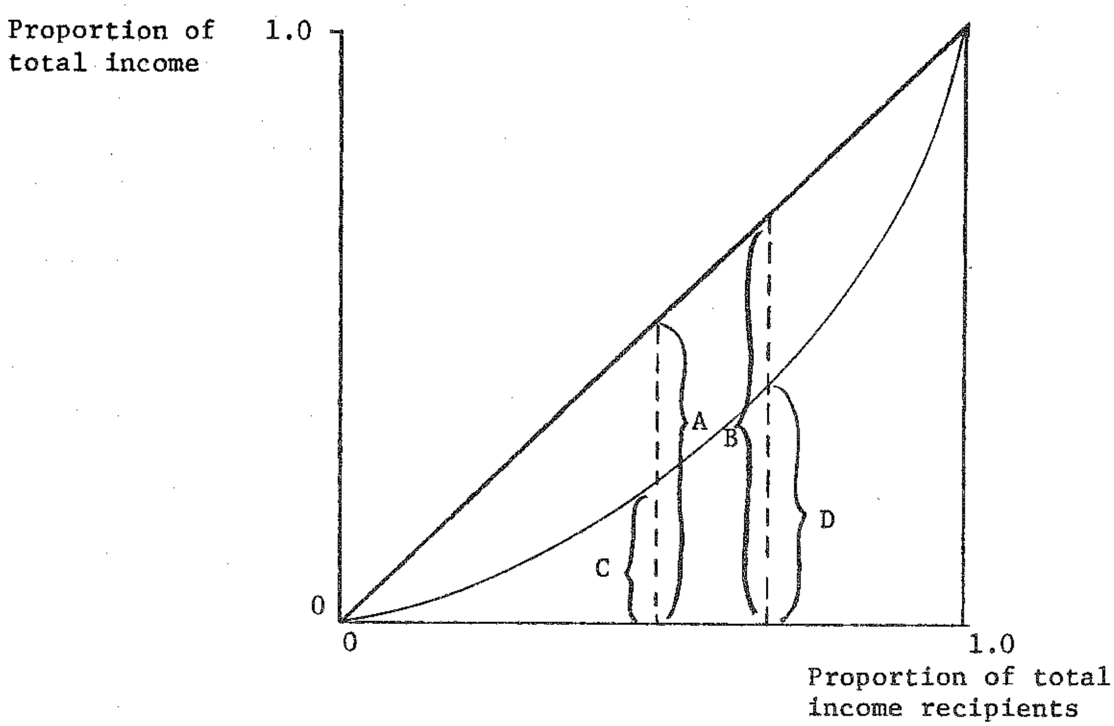
A = fractional proportion of the cumulative recipients of the previous observation to the total number of observations.

B = fractional proportion of the cumulative recipients of the present observation to the total number of observations.

C = fractional proportion of the cumulative income of the previous observations to the total income for all observations.

D = fractional proportion of the cumulative income of the present observation to the total income for all observations.

These measurements can be seen more easily in the figure below depicting the Lorenz curve, from which the Gini ratio is derived.



3/ Morgan, James, "The Anatomy of Income Distribution," The Review of Economics and Statistics, Vol. XLIV, August 1962.

In order to compute the Gini ratio from equation (1), one must first know the total number of observations as well as the total income (or other variable being measured) for all observations to compute the cumulative fractional proportion of income recipients (A and B) and the cumulative fractional proportion of total income (C and D), respectively. Where a simple random sample or a constant proportion stratified sample has been used, the computations are quite straight forward. The quantity (B-A) reduces to a constant which is equal for all observations. The reduced form is shown below:

$$G = 1 - \Sigma K(C+D) \tag{2}$$

The calculating equation may be written as:

$$G = 1 - K(2\Sigma D - 1) \tag{3}$$

or

$$G = 1 - K\left(\frac{2\Sigma\Sigma y_i}{\Sigma y_i} - 1\right) \tag{4}$$

where K = constant fractional proportions for each income recipient, 1/n.

What remains to be computed is the quantity (C+D). This is shown in the following example of computing the Gini ratio for the distribution of income for a sample of 10 observations:

Observations	Income	Cumulative Income
1	5	5
2	10	15
3	20	35
4	35	70
5	40	110
6	63	173
7	98	271
8	125	396
9	314	710
10	571	1281
	1281	3066

$$K = 1/10 = .10$$

$$\Sigma\Sigma y_i = 3066$$

$$\Sigma y_i = 1281$$

$$G = 1 - 1/10\left[\frac{2(3066)}{1281} - 1\right] = .621$$

These procedures need to be adjusted in two ways when a stratified sample is being measured where the sampling proportion varies between strata. Both the number of observations (income recipients) and the total income must be adjusted along with the relevant changes in values for K, C, and D. The general formula for computing the Gini ratio for a stratified sample with varying sampling proportions is given below:

$$G = 1 - \Sigma K_i (C+D) \tag{5}$$

The calculating equation can be written as:

$$G = 1 - \frac{\Sigma K_i (\Sigma y'_{ij} + \Sigma y'_{ij-1})}{Y'} \tag{6}$$

or

$$G = 1 - \frac{\Sigma K_i (y'_{ij} + 2\Sigma y'_{ij-1})}{Y'} \tag{7}$$

where $(y_{i1} < y_{i2} < \ldots < y_{in})$

$$K_i = L_i / N$$

$$L_i = \frac{n'_i}{n_i}$$

$$y'_{ij} = L_i y_{ij}$$

$$Y' = \Sigma L_i y_{ij}$$

K_i = constant fractional proportion for each income recipient j in strata i

L_i = income and recipient adjustment factor for each strata i

N = total population, all strata

n_i = number of observations in sample in strata i

n'_i = total number of observations in population in strata i

y_{ij} = measured income for each income recipient j in strata i

y'_{ij} = adjusted income for each income recipient j in stata i

Y' = total adjusted income, all strata

Using the data from the previous example, we are able to demonstrate the use of the adjusted Gini ratio for a two strata sample. The following information is given:

Observation	Strata 1 (y_{1j})	Strata 2 (y_{2j})	
1		5	
2	10		$X_1=25$
3	20		$x_1= 7$
4		35	$X_2=20$
5		40	$x_2= 3$
6	63		$Y_1=1201$
7	98		$Y_2=80$
8	125		
9	314		
10	571		
	1201	80	

If we use the above information, we are able to compute the appropriate adjustment factors.

$$L_1 = 3.571$$

$$L_2 = 6.667$$

$$K_1 = 0.0794$$

$$K_2 = 0.1482$$

$$Y' = 4822.13$$

The calculations for the adjusted Gini ratio are as follows:

Observation Identification	Observation	Adjusted Income	$K_i(y'_{ij} + 2\Sigma y'_{ij-1})$
	(y_{ij})	(y'_{ij})	
y_{21}	5	33.335	4.939
y_{12}	10	35.710	8.124
y_{13}	20	71.420	16.626
y_{24}	35	233.345	76.193
y_{25}	40	266.680	150.274
y_{16}	63	224.973	119.506
y_{17}	98	349.958	165.130
y_{18}	125	446.375	228.323
y_{19}	314	1121.294	352.726
y_{110}	571	2039.041	603.517
Totals		4822.131	1725.358

$$G = 1 - 1725.358/4822.131$$

$$G = 1 - 0.358 = 0.642$$

The above calculations are rather tedious, time-consuming, and open to error in the several steps required to compute the Gini ratio for large quantities of data. In order to reduce the calculating time and eliminate some of the source of possible error, two programs are offered. One program is for use on a computer (IBM 370/168), and the other program is for a hand calculator (HP-65). Instruction for their use are given.

Program for calculating the adjusted Gini ratio on the HP-65

Key Entry	Code Shown	Description
LBL	23	begin subroutine
A	11	subroutine A
RCL1	3401	recall memory 1
x	71	multiply
STO8	3308	store memory 8
RCL7	3407	recall memory 7
+	61	add
STO7	3307	store memory 7
RCL8	3408	recall memory 8
RCL6	3406	recall memory 6
+	61	add
RCL1	3401	recall memory 1
x	71	multiply
RCL3	3403	recall memory 3
÷	81	divide
RCL5	3405	recall memory 5
+	61	add
STO5	3305	store memory 5
RCL7	3407	recall memory 7
2	02	number 2
x	71	multiply
STO6	3306	store memory 6
RCL4	3404	recall memory 4
RCL1	3401	recall memory 1
+	61	add
STO4	3304	store memory 4

Key Entry	Code Shown	Description
RCL8	3408	recall memory 8
RCL1	3401	recall memory 1
÷	81	divide
RTN	24	end subroutine
LBL	23	begin subroutine
B	12	subroutine B
RCL2	3402	recall memory 2
x	71	multiply
STO8	3308	store memory 8
RCL7	3407	recall memory 7
+	61	add
STO7	3307	store memory 7
RCL8	3408	recall memory 8
RCL6	3406	recall memory 6
+	61	add
RCL2	3402	recall memory 2
x	71	multiply
RCL3	3403	recall memory 3
÷	81	divide
RCL5	3405	recall memory 5
+	61	add
STO5	3305	store memory 5
RCL7	3407	recall memory 7
2	02	number 2
x	71	multiply

Key Entry	Code Shown	Description
ST06	3306	store memory 6
RCL4	3404	recall memory 4
RCL2	3402	recall memory 2
+	61	add
ST04	3304	store memory 4
RCL8	3408	recall memory 8
RCL2	3402	recall memory 2
÷	81	divide
CHS	42	change sign
RTN	24	end of subroutine
LBL	23	begin subroutine
C	13	subroutine C
RCL5	3405	recall memory 5
RCL7	3407	recall memory 7
÷	81	divide
CHS	42	change sign
1	01	number 1
+	61	add
RTN	24	end of subroutine
LBL	23	begin subroutine
D	14	subroutine D
RCL4	3404	recall memory 4
RCL3	3403	recall memory 3
—	51	subtract
RTN	24	end of subroutine

Key Entry	Code Shown	Description
LBL	23	begin subroutine
E	15	subroutine E
0	00	number 0
STO4	3304	store memory 4
STO5	3305	store memory 5
STO6	3306	store memory 6
STO7	3307	store memory 7
STO8	3308	store memory 8
RTN	24	end subroutine

Steps to be followed:

1. Enter program
2. Enter L_1 in memory 1 (STO1 = 3301)
3. Enter L_2 in memory 2 (STO2 = 3302)
4. Enter N in memory 3 (STO3 = 3303)
5. Enter observations arrayed in ascending value ($y_{i1} < y_{i2} < \ldots < y_{in}$)

   Depress A for observations in strata 1

   Depress B for observations in strata 2

   The observation appears after entry of each observation with observations in strata 2 appearing with a negative sign.

6. Compute Gini ratio

   Depress C

7. Check Computation

   Depress D

   This should equal zero to at least the fourth place beyond the decimal point. If this does not equal zero, an error has been made in L_1, L_2, N, or an error in entering the data. If L_1, L_2, and N are correct and depressing D does not equal zero, depress E (erasing memory 4-8) and enter data again.

This program may also be used for computing Gini ratios for a simple random sample or for a stratified sample where the sampling proportion is the same for all strata. In this case, the steps to be followed would be:

1. Enter program
2. Enter L_1 in memory 1
3. Enter N in memory 3
4. Enter all observations by depressing A ($y_1 < y_2 < \dots < y_n$)

Alternatively, the following steps may be followed to obtain equal results:

1. Enter program
2. Enter 1 in memory 1
3. Enter n in memory 3
4. Enter arrayed observations (in ascending order) in A.

Computer program for calculating the adjusted Gini ratio on the IBM 370/168

```
*WATFIV
C
C          PROGRAM TO CALCULATE GINI COEFFICIENT USING STRATIFIED SAMPLES
C          $$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$¢$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
           DIMENSION ARRAY(6),A(6)
           DATA ARRAY/6*0.0/
           SMADJ=0
           YSMT=0
           YSTT=0
           CNT=0
  190      READ(5,200,END=800)IONE,KARD,K1,K2,K3,K4,A1,A2,A3,A4,A5,A6
  200      FORMAT(I1,I2,I1,I2,2I1,1X,6F10.3)
C
C     POP= TOTAL POPULATION IN REGION CONSIDERED
           POP=45
C
C     SM1= SAMPLE SIZE OF STRATA 1
C     SM2= SAMPLE SIZE OF STRATA 2
C     SM3,SM4,SM5 HAVE THE SAME EXPLANATION AS SM1 AND SM2
           SM1=7
           SM2=3
           SM3=0
           SM4=0
           SM5=0
C
C     POP1= POPULATION OF STRATA 1
C     POP2,POP3,POP4,POP5 HAVE THE SAME EXPLANATION AS POP1
           POP1=25
           POP2=20
           POP3=0
           POP4=0
           POP5=0
C
C     Y= FIELD NAME WHERE THE INCOME OBSERVATIONS LAY
           Y=A2
C
C     STR=FIELD NAME WHICH IDENTIFIES WHICH STRATA DATA IS FROM
           STR=IONE
C
C     STR1=NUMBER THAT IDENTIFIES STRATA 1 IN STR
C     STR2,STR3,STR4,STR5 HAVE THE SAME EXPLANATION AS STR1
           STR1=1
           STR2=2
           STR3=0
           STR4=0
           STR5=0
```

```
C
C     SUMMING AND ADJUSTING OBSERVATIONS FOR GINI CALCULATION
          CNT=CNT+1
          IF(CNT.EQ.1) YLAG=0
C
C     CALCULATION OF ADJUSTMENT FACTORS
C     CHECKING TO SEE WHICH STRATA THE DATA IS FROM
          IF(STR.EQ.STR1) CFPYR=(POP1/   SM1)/POP
          IF(STR.EQ.STR2) CFPYR=(POP2/   SM2)/POP
          IF(STR.EQ.STR3) CFPYR=(POP3/   SM3)/POP
          IF(STR.EQ.STR4) CFPYR=(POP4/   SM4)/POP
          IF(STR.EQ.STR5) CFPYR=(POP5/   SM5)/POP
          IF(STR.EQ.STR1) ADJ=(POP1/   SM1)
          IF(STR.EQ.STR2) ADJ=(POP2/   SM2)

          IF(STR.EQ.STR3) ADJ=(POP3/   SM3)
          IF(STR.EQ.STR4) ADJ=(POP4/   SM4)
          IF(STR.EQ.STR5) ADJ=(POP5/   SM5)
          YSMT=YSMT+CFPYR*(2*SMADJ+(ADJ*Y))
          YLAG=Y
          SMADJ=(ADJ*YLAG)+SMADJ
          YSTT=YSTT+(ADJ*Y)
          GO TO 190
  800     IF(CNT.EQ.POP) GO TO 810
C
C            CALCULATION OF  GINI
C
  810     G=1-(YSMT/YSTT)
          PRINT 802
  802     FORMAT('1')
C
          PRINT 900, G
  900     FORMAT(' GINI =',F8.4)
          RETURN
          END

*DATA
```

```
GINI =  0.6422
```